Advances in Meteorology

Computational Algorithms for Climatological and Hydrological Applications

Lead Guest Editor: Upaka Rathnayake
Guest Editors: Priyantha Ranjan Sarukkalige and Sangam Shrestha

Computational Algorithms for Climatological and Hydrological Applications

Advances in Meteorology

Computational Algorithms for Climatological and Hydrological Applications

Lead Guest Editor: Upaka Rathnayake
Guest Editors: Priyantha Ranjan Sarukkalige and Sangam Shrestha

Contents

Hindawi
Advances in Meteorology
Volume 2023, Article ID 9458678, 18 pages
https://doi.org/10.1155/2023/9458678

Research Article

Study on the Impact of Future Climate Change on Extreme Meteorological and Hydrological Elements in the Upper Reaches of the Minjiang River

Ting Chen,[1,2,3] **Yao Ye**,[2] **Kebi Yang**,[3] **Xu Zhang**,[3,4,5] **and Tianqi Ao**[3]

[1]*College of Resources and Environment, Chengdu University of Information Technology, Chengdu, Sichuan 610225, China*
[2]*Institute of Plateau Meteorology, China Meteorological Administration (CMA), Heavy Rain and Drought Flood Disaster in Plateau and Basin Key Laboratory of Sichuan Province, Chengdu 610072, China*
[3]*State Key Laboratory of Hydraulics and Mountain River Engineering, College of Water Resources and Hydropower, Sichuan University, Chengdu 610065, China*
[4]*State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Nanjing Hydraulic Research Institute, Nanjing 210029, China*
[5]*Research Center for Climate Change, Ministry of Water Resources, Nanjing 210029, China*

Correspondence should be addressed to Ting Chen; cting43@foxmail.com

Received 18 May 2022; Revised 9 November 2022; Accepted 17 January 2023; Published 9 February 2023

Academic Editor: Upaka Rathnayake

Global warming increases global average precipitation and evaporation, causing extreme climate and hydrological events to occur frequently. Future changes in temperature, precipitation, and runoff from 2021 to 2050 in the upper reaches of the Minjiang River were analyzed using a distributed hydrological model, the SWAT (Soil and Water Assessment Tool), under a future climate scenario. Simultaneously, future variation characteristics of extreme climate hydrological elements in the upper reaches of the Minjiang River were analyzed using extreme climate and runoff indicators. The research shows that the frequency and intensity of the extreme temperature warming index will increase, while those of the extreme temperature cooling index will increase and then weaken in the upper reaches of the Minjiang River under a future climate scenario. The duration of precipitation, the intensity of continuous heavy precipitation, and the frequency of heavy precipitation will increase, whereas the intensity of short-term heavy precipitation and the frequency of heavy precipitation will decrease. However, spatial distribution of flood in the upper reaches is different, and thus flood risk in the upstream source area will still tend to increase. Particular attention should be given to the increase in autumn flood risk in the upper reaches of the Minjiang River.

1. Introduction

Global warming will aggravate the global hydrological cycle and increase global average precipitation and evaporation [1]. Simultaneously, precipitation variability may change, exerting direct effects on evaporation, runoff, and soil humidity. The extreme hydrological events, such as floods and droughts, increase the risk of water disasters, which have become major challenges to human survival [2]. Under the background of climate warming, extreme climate and hydrological events occur frequently in China [3–10]. The change in extreme climate events and the impact of change in hydrological and water resources caused by extreme climate events on human life and production have attracted increasing attention [8, 11]. Research on the changing trend, occurrence mechanism, response, and prediction of hydrological extreme events in river basins under the background of climate change is important to understand. It is significant for scientifically understanding the spatiotemporal evolution law of the land-water cycle under the background of global climate change. It has an important application value in basin flood control planning and design, large-scale hydropower development planning and operation management, regional disaster prevention and

reduction, ecological environment protection, and sustainable economic and social development.

The study of the impact of climate change on water resources focuses on the analysis of the water cycle and water resources trends under climate change [12]. Trends in water resources under climate change first need to provide more accurate climate change scenario data for hydrological simulation. Climate change scenarios are based on several scientific assumptions and use a long series of historical information to establish a continuous and consistent predictive description of future climate including temperature, precipitation, and other elements. Currently, the GCM (Global Climate Model) output method for developing future climate change scenario design is most widely used, but the basin-scale hydrological models do not match the large grid data of GCMs. So, downscaling of future climate change scenarios is needed [13, 14]. The main types of downscaling methods are statistical downscaling, dynamical downscaling, and combined statistical-dynamical downscaling. Statistical downscaling has been increasingly used to predict future climate scenarios such as precipitation and temperature due to the lower operational resource requirements and higher accuracy [15]. Statistical downscaling refers to the use of historical observations to establish statistical relationships between regional observations and large-scale climate elements and to verify such statistical relationships using other regional observations. The frequently used statistical downscaling models are ASD (automated statistical downscaling) and SDSM (statistical downscaling model), among which SDSM has less bias and is convenient to use and is one of the most widely used statistical downscaling models [16]. Gulacha et al. [17] assessed that the model and the observed showed a good fit in the Wami-Ruvu River Basin of Tanzania, and the SDSM's R^2 values between raw and model for temperature ranged from 0.42 to 0.98. Tukimat et al. [18] found that the SDSM successfully provided long-term climate pattern at the gauged stations with an R value close to 1.0 in Kuantan River Basin of Malaysia. Dehghan et al. [19] assessed that there is a good agreement between the simulated and the observed precipitation, and R^2 for precipitation is greater than 0.63.

Research shows that the impact of climate change on the hydrological cycle is regional [20–30]. Pandey et al. [20] assessed that there is a high consensus for increase in temperature but higher uncertainty with respect to precipitations in Mahakali of Nepal. Under the projected changes, the average annual streamflow was simulated to increase gradually from the near to far future under both RCPs (Representative Concentration Pathways). Bajracharya et al. [21] found that the extreme projection of a RCP 8.5 scenario shows that the average annual temperature of the basin is expected to increase by more than 4°C in the Kaligandaki Basin of Nepal. Likewise, the average annual precipitation in the basin is projected to increase by as much as 26% during the late century under a RCP 8.5 scenario. The synergetic effect of an increase in temperature and precipitation shows the aggravated effect on the discharge and water yield with an increase of more than 50% at the outlet of the basin. Fonseca and Santos [22] simulated the potential effects of climate change on the hydrology in the Tâmega River Basin, Northern Portugal, experiencing a Mediterranean climate. The annual precipitation over the Tâmega River Basin exhibits weakly decreasing trends across the entire future period. On the other hand, temperatures show consistently warming trends throughout the basin for the future period, with a mean warming rate of 0.03°C per year. As a result, the mean annual flow rate decreased at all hydrometric stations by about $0.25\,m^3s^{-1}$ per year, with increased flow rates in winter when compared to the historical period but significantly lower flow rates in summer. Meaurio et al. [23] assessed the climate change impact on river discharge in the Bay of Biscay, Spain. It was found that trends for extreme flows show an increase in the duration of low flows.

In China, Yu et al. [24] pointed out that while the precipitation in Northern China is decreasing, the change in precipitation exhibits the form of "waterlogging in the south and drought in the north." The probability of flood disaster in the Yangtze River Basin is considerably higher than that in other regions of China. Since the 20th century, more than 20 floods have occurred in the Yangtze River Basin. Among these, the floods of 1905, 1931, 1954, 1988, 2010, and 2017 were the most severe [24]. Wang et al. [25] reported that the concentrated distribution and high frequency of large-area rainstorms and floods occur in the monsoon region of Eastern China. Meanwhile, in Western China's arid and semiarid regions, disastrous floods are mostly caused by short-term local rainstorms, and small and medium-sized rivers can form high peak flow, causing serious disasters to local regions.

The Minjiang River Basin is a first-class tributary in the upper reaches of the Yangtze River. It is located to Southwest China and is in the southeast edge of the Qinghai-Tibet Plateau. Under the background of the large terrain of the Qinghai-Tibet Plateau and the comprehensive influences of the large, medium, and small complex terrain, the valley is steep, hydropower energy is abundant, and weather and climate disasters, such as rainstorms and mountain torrents, occur frequently. This area is sensitive to climate change and extreme hydrometeorological events. Previous studies on the Minjiang River Basin focused on the response of this basin to climate change. For example, Liang et al. [26] found that the temporal variation characteristic of the upper reaches of the Minjiang River is as follows: annual average precipitation exhibits a downward trend due to the reduction in summer precipitation. Meanwhile, the spatial distribution characteristic is that high-altitude areas exhibit an increasing trend, whereas low-altitude areas present a decreasing trend. The annual runoff in the upper reaches of the Minjiang River shows a significant downward trend from 1937 to 2018, and it may continue to demonstrate a downward trend in the future. Huang et al. [27] found that the average temperature in the upper reaches of the Minjiang River shows an upward trend, whereas precipitation and annual runoff exhibit a downward trend. The trend of the average temperature has an evident positive correlation with precipitation, particularly in spring and autumn. The decrease in water inflow during spring and the extension of the duration of the low-

flow season during autumn will considerably impact irrigation and urban water supply. Huang et al. [28] demonstrated that climate change scenario analysis combined with the Soil and Water Assessment Tool (SWAT) hydrological model can effectively simulate the effect of climate change on runoff. The influence of precipitation change on runoff is greater than that of temperature change. The impact of temperature change on runoff is more evident in dry years than in wet years. Chen et al. [29] showed that the overall temperature in the Minjiang River Basin exhibited a trend of decreasing the number of extreme cold days while increasing the number of extreme warm days. In terms of spatial distribution, the high value of the extreme cold event index in the basin was mostly recorded in the upstream. The spatial distribution of extreme precipitation indicators in the Minjiang River Basin is highly uneven, as manifested in the high value of extreme precipitation indicators mostly appearing in the middle and lower reaches of the basin. From the analysis of the changing trend, the average characteristics of the Minjiang River Basin range from short-term to sustained extreme precipitation [30].

In summary, research on the change in extreme climate events in the upper reaches of the Minjiang River mostly used historical data to analyze temperature and precipitation extreme events. Meanwhile, there is less research on future changes in extreme events, particularly hydrological extreme events. In the current study, the hydrological model (i.e., SWAT) is used to simulate runoff change in the upper reaches of the Minjiang River under a future climate scenario. Simultaneously, the change characteristics of extreme climate hydrological elements in the upper reaches of the Minjiang River under a future climate scenario are analyzed using extreme climate and runoff indices. The study analyzes the future trends of hydrometeorological element extremes under climate scenarios, providing useful support for theoretical studies of potential drought and flood threats in the upper reaches of the Minjiang River, as well as a reference for future water conservancy project design in the upper reaches of the Minjiang River.

2. Materials and Methods

2.1. Study Area and Data Sources. The upper reaches of the Minjiang River Basin are located between 102° 59′–104° 14′ E and 26° 33′–33° 16′ N and have a drainage area of about 23,000 km^2 (Figure 1). The upper reaches of the Minjiang River are sensitive to climate change and frequent natural disasters due to the high intensity of water resource development, the reduction of forest coverage, the degradation of ecological functions, and serious soil and water losses [29, 30].

The Digital Elevation Model (DEM) data in this study were obtained from the geospatial data cloud (https://www.gscloud.cn/search) with a resolution of 1 km (Figure 2(a)). Land use data were obtained from the International Geosphere-Biosphere Programme (IGBP), which uses the United States Geological Survey (USGS) method to classify land use into 17 categories. The land use types of the Minjiang River were mainly classified into 7 categories (Figure 2(b)). Soil data are from the Harmonized World Soil Database (HWSD) at 1 km resolution and can be downloaded from the Food and Agriculture Organization of the United Nations. SPAW was used to calculate the parameters required for the SWAT soil database, which contains 21 soil types in the assessment area of this study (Figure 2(c) and Table 1).

3. Methodology

This study analyzes the impact of future climate change on extreme hydrometeorological elements in the upper reaches of the Minjiang River. The statistical downscaling model (SDSM) is selected for future climate scenario analysis, and the SWAT model is used for hydrological simulation.

3.1. SDSM. The SDSM model scales down by establishing statistical relationships between large-scale climate predictors and observations.

When simulating precipitation, the large-scale climate forecaster is used to first simulate the probability of precipitation on a particular day and then the amount of precipitation on that rainy day:

$$W_i = a_0 + \sum_{j=0}^{n} a_j x_j + a_{i-1} W_{i-1}, \tag{1}$$

where W_i and W_{i-1} represent the precipitation probability of days i and $i-1$, respectively, and x_j is the j-th predictor and regression coefficient.

The occurrence of precipitation is determined by a random number r $(0 \le r \le 1)$ that follows a uniform distribution. If $r \le W_i$, then precipitation will occur on that day. When precipitation occurs on a certain day, a multiple exponential regression function will be used to simulate precipitation on that day:

$$R_i = \exp\left(\beta_0 + \sum_{j=1}^{n} \beta_j x_j + \varepsilon_i\right), \tag{2}$$

where R_i is precipitation on the i-th day, β_0 and β_j are the regression coefficients, x_j is the prediction factor on the j-th day, and ε_i is the error.

The major steps in using SDSM include quality control, downscaling prediction factor screening, model correction, weather generator, and model evaluation. The original meteorological observation data collected from meteorological stations may be missing. Therefore, quality control should be implemented to identify missing data, outliers, and suspicious incomplete data, improving the quality of model output [6, 31].

Referring to Wilby [31], the prediction factors of temperature and precipitation (Table 2) and rainfall station forecast factors are selected.

The temperature and precipitation variables under three climate scenarios in Coupled Model Intercomparison Project Phase 5 (CMIP5) by SDSM are downscaled as shown in Table 3. CMIP5 has designed "Representative Concentration Pathways" (RCPs) for the emissions of four types of

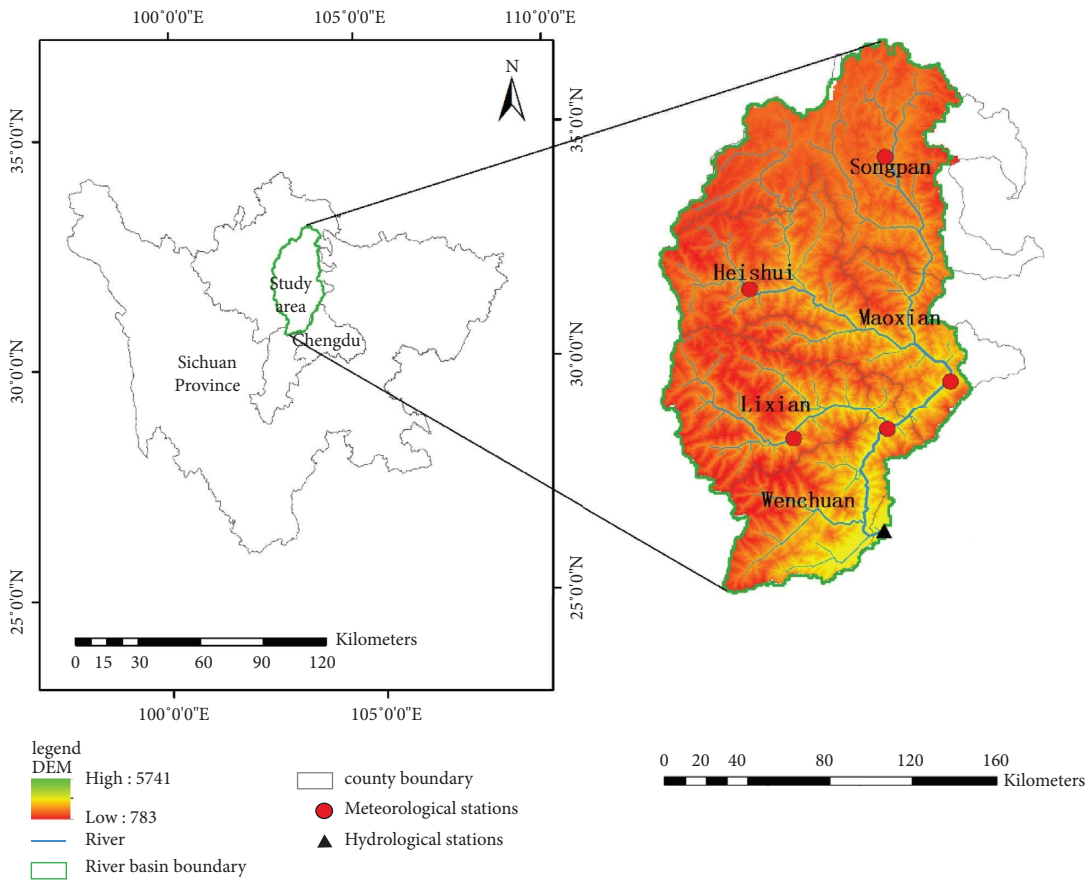


Figure 1: The drainage system in the upper reaches of the Minjiang River.

greenhouse gases and aerosols [32–35], namely, RCP2.6, RCP4.5, RCP6.0, and RCP8.5. Each scenario includes a set of greenhouse gas, aerosol, and other gas emissions.

3.2. SWAT Model. The SWAT model is a semidistributed watershed hydrological model developed by the United States Department of Agriculture-Agricultural Research Service (USDA-ARS). The watershed delineation tool in SWAT delineates the whole study basin into several subbasins in accordance with the characteristics of topographic factors and river network distribution. On this basis, hydrological response units are divided in accordance with the land use type, soil type, and slope area threshold of the basin. Runoff is calculated separately. Finally, the total runoff of an outlet section is obtained through river confluence routing.

3.3. SUFI-2. SWAT-CUP makes the calibrating procedure easy to users[8, 36]. Sequential Uncertainty FItting ver. 2 (SUFI-2) is selected to calibrate the parameters in this study.

In SUFI-2, parameter uncertainty accounts for all sources of uncertainties such as uncertainty in driving variables (e.g., rainfall), conceptual model, parameters, and measured data. The degree to which all uncertainties are accounted for is quantified by a measure referred to as the P factor, which is the percentage of measured data bracketed by the 95% prediction uncertainty (95PPU). As all the processes and model inputs such as rainfall and temperature distributions are correctly manifested in the model output (which is measured with some error)—the degree to which we cannot account for the measurements—the model is in error and is hence uncertain in its prediction. Therefore, the percentage of data captured (bracketed) by the prediction uncertainty is a good measure to assess the strength of our uncertainty analysis. The 95PPU is calculated at the 2.5% and 97.5% levels of the cumulative distribution of an output variable obtained through Latin hypercube sampling, disallowing 5% of the very bad simulations. As all forms of uncertainties are reflected in the measured variables (e.g., discharge), the parameter uncertainties generating the 95PPU account for all uncertainties. Reducing the total uncertainty into its various components is highly interesting but quite difficult to do, and to the best of the authors' knowledge, no reliable procedure yet exists.

3.4. SWAT Model Development. In this study, the 1969–1980 is the calibration period and 1981–1987 is the verification period. The Nash–Sutcliffe model efficiency coefficient (NSE) and the coefficient of determination (R^2) are selected as indices for evaluating the simulation effect of daily runoff in the upper reaches of the Minjiang River. Simultaneously, the simulation effect of the SWAT model on extreme high and low flows is evaluated using the correlation coefficient.

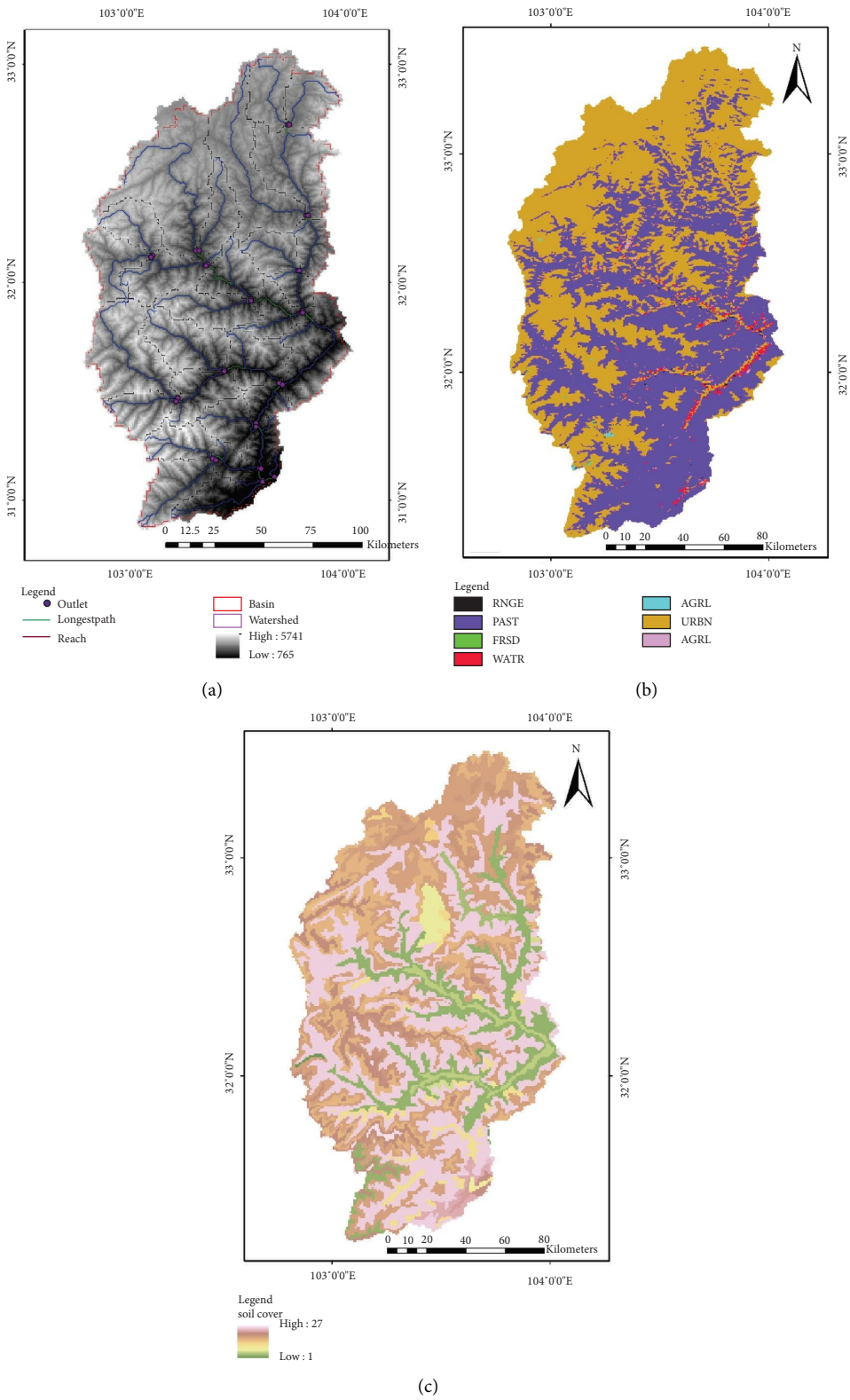


Figure 2: (a) DEM, (b) land use, and (c) soil cover maps of the Minjiang River.

3.4.1. Extreme Climate and Flow Indicators. The extreme temperature and precipitation indicators [37] recommended by the World Meteorological Organization are selected in this study. These indicators can reflect the intensity, frequency, and duration of extreme temperature and precipitation elements [38]. A detailed introduction of these indicators is provided in Table 4. This study also selects extreme runoff indicators to reflect changes in extreme

Table 1: Soil types of the Minjiang River.

No	Name
1	Water bodies
2	Rock outcrops
3	Haplic Luvisols
4	Eutric Cambisols
5	Luvisols
6	Dystric Cambisols
7	Alisols
8	Lithic Leptosols
9	Leptosols
10	Mollic Leptosols
11	Gelic Cambisols
12	Gelic Leptosols
13	Mollic Gleysols1
14	Mollic Gleysols
15	Haplic Phaeozems
16	Calcaric Regosols
17	Eutric Regosols
18	Dystric Regosols
19	Leptosols
20	Chromic Luvisols
21	Haplic Luvisols
22	Haplic Greyzems
23	Calcaric Cambisols
24	Calcic Luvisols
25	Calcaric Cambisols
26	Dystric Podzoluvisols
27	Haplic Luvisols

hydrological factors. Extreme runoff indicators are classified into absolute and relative indicators. Absolute indicators refer to the maximum daily flood volume. Relative indicators refer to the 1% (Q1), 5% (Q5), 10% (Q10), 99% (Q99), 95% (Q95), and 90% (Q90) quantile runoff. Q1, Q5, and Q10 represent low-flow extreme runoff, while Q99, Q95, and Q90 represent extreme high-flow runoff [39]. Simultaneously, this study calculates the hydrological frequency of extreme high and low flows under a future climate scenario and the historical period, in which the annual maximum daily flow can represent extreme high flow while the annual minimum monthly flow can represent extreme low flow [38].

4. Results and Discussion

4.1. Simulation Results of SDSM. The results show that the spatial correlation between large-scale variables simulated and measured using the second-generation Canadian Earth System Model (CANESM2) is 0.76 for all stations in Southwest China, where the upper reaches of the Minjiang River are located. CANESM2 can simulate more accurate large-scale variables, such as the upper atmospheric field and atmospheric circulation, with a resolution of $2.5^{\circ} \times 2.5^{\circ}$; however, it lacks regional climate information.

The large-scale and low-resolution variables of CANESM2 simulation are reduced to a regional scale by SDSM. The analysis shows that the correlation coefficient between the simulated and measured daily maximum and minimum temperatures is about 0.9, while the correlation coefficient between the simulated and measured daily precipitation is between 0.4 and 0.5 (Table 5), indicating that analyzing extreme meteorological elements is feasible by using temperature and precipitation data after SDSM downscaling.

The average absolute error between the simulation and actual measurements during different seasons in the verification period is calculated (Table 6). The average absolute error of the highest and lowest temperatures exhibits minimal seasonal variation, and the daily average absolute error of precipitation in summer is higher than that in other seasons. In general, using SDSM for downscaling research in the upper reaches of the Minjiang River is feasible.

4.2. Simulation Results of SWAT. The SWAT model was used to simulate the daily runoff in the upper Minjiang River, and the optimal values of the model parameters are shown in Table 7. The results showed that R^2 was 0.87 and the NSE was 0.86 in the calibration period, R^2 was 0.79, and the NSE was 0.77 in the validation period, indicating that it is feasible to simulate the runoff of the upper Minjiang River using the SWAT model (Figure 3). From the simulation results, it can be seen that the SWAT model is able to simulate the seasonal distribution characteristics of runoff. SWAT simulates runoff more accurately in the flood season, and the model simulates slightly lower runoff in the dry season, which may be related to the greater contribution of high flows in the flood season to the simulation error evaluation [7, 8].

The simulated effects of extreme high flows and extreme low flows in the upper Min River from 1969 to 1987 simulated by SWAT were calculated, where the annual maximum daily flows were used to represent extreme high flows, and the annual minimum monthly flows were used to represent extreme low flows. The simulation results show that the correlation coefficient between the simulated and measured annual maximum 1-day (AM) flow is 0.75, and the correlation coefficient between the simulated and measured minimum monthly flow (IM) is 0.82, indicating that the SWAT model for the upper Minjiang River can be used to carry out studies of extreme runoff.

4.3. Variation Characteristics of Temperature, Precipitation, and Runoff under Future Scenarios. As shown in Figure 4, the monthly average temperature rises by 2°C to 3°C. Temperature increase is most evident from June to August in summer and December to January in winter. Comparing the three scenarios, temperature rise is most apparent in the RCP8.5 scenario. Precipitation decreases from April to May in spring compared with the historical period, and it increases in other months compared with the historical period, particularly from July to August in summer, with an increase of 27–40 mm. The increase in runoff is most evident from November to April of the following year, while runoff from June to July is less than that in the historical period, i.e., more than 10% less than that in the historical period. From June to July in summer, temperature and precipitation increase, whereas runoff decreases, indicating that the rising range of precipitation cannot compensate for the impact of the rising evapotranspiration caused by increasing temperature. From April to May in spring, temperature and runoff increase,

Table 2: Prediction factors of temperature and precipitation in the study area.

Station	Maximum temperature	Minimum temperature	Precipitation
Maoxian	mslp; p1_u; p8_u; p850; s500; s850; temp	mslp; p1_u; p1_z; p5_z; p500; p850; s850; shum; temp	mslp; p1_v; p5zh; p1zh; p5_v; p5_z; p5zh; p8_v; prcp; s500; shum; temp
Songpan	mslp; p1_v; p5_u; p500; p8_v; p850; prcp; s850; shum; temp	mslp; p1_up1_v; p1zhp5_v; s500; s850; shum; temp;	mslp; p1_v; p1zh; p5_v; p500; p8_z; p850; p8zh; prcp; temp
Wenchuan	mslp; p1zh; p500; p8_z; s850; temp	mslp; p1_z; p500; p8zh; s850; shum; temp	p1_u; p1_v; p1zh; p5_u; p5_z; p500; p5zh; p8_u; p8zh; s500; shum
Lixian	mslp; p1_v; p850; s500; s850; temp	mslp; p1_z; p500; s850; shum; temp	mslp; p1_f; p1_u; p1zh; p5_z; p500; p5zh; p8_u; p8zh; prcp; shum; temp
Heishui	mslp; p1_v; p5_u; p5_z; p500; p5zh; prcp; s850; s500; shum; temp	mslp; p1zh; p5_v; p500; s850; shum; temp	p1_v; p1zh; p5_u; p5_z; p500; p8_ugl; p8zh; shum; temp

Table 3: Typical concentration path of future climate change recommended in the IPCC fifth assessment report.

Climate scenario	Radiation forcing in 2010–2100	The emissions of greenhouse
RCP2.6	It reaches the forced peak of 3 W/m^2 in the medium term and then decreases to 2.6 W/m^{2}by 2100	Relief
RCP4.5	Increases to 4.5 W/m^2	Moderate
RCP8.5	Increases to 8.5 W/m^2	Aggravate

whereas precipitation decreases, indicating that the impact of temperature increase on the snow-melting process is greater than that on evapotranspiration in spring. The increase in melting snow amount increases runoff. The decrease in runoff from June to July in summer will affect agricultural irrigation water and may lead to agricultural production reduction. It may also affect downstream water diversion and the ecological environment.

5. Analysis of the Variation Characteristics of Extreme Temperature Index

5.1. Relative Indices (TX10p, TN10p, TX90p, and TN90p). Table 8 provides the comparison of extreme temperature index values in the upper reaches of the Minjiang River between the historical period and future climate scenarios. Tx10p and Tn10p mostly increase, particularly in the RCP8.5 scenario. The increase in Tn90p and Tx90p is more evident in the future climate scenario, particularly in the RCP2.6 scenario. These findings show that the relative indicators in the upper reaches of the Minjiang River are largely increasing in the future climate scenario, in which the increase of warm indicators is more evident. Meanwhile, the increase of cold indicators is more apparent in the RCP8.5 scenario, and the increase of warm indicators is more noticeable in RCP2.6.

5.2. Extreme Indices (TX_n, TN_n, TX_x, and TN_x). TX_n and TN_n in the upper reaches of the basin will rise under the future climate scenario, particularly in RCP8.5. TX_x and TN_x will rise in the future climate scenario. The rise of TX_x is most evident in the RCP8.5 scenario, while the rise of TN_x is most apparent in the RCP2.6 scenario. These findings show that among the extreme value indices in the upper reaches of the Minjiang River, the cold index decreases, whereas the warm index increases, and the change is more evident in the RCP8.5 scenario.

6. Analysis of the Variation Characteristics of Extreme Precipitation

6.1. Sustainability Index: Crop Water Demand (CWD). Table 9 presents the comparison of extreme precipitation index values in the upper reaches of the Minjiang River in the historical period and future climate scenario. Compared with the historical period, CWD increased in the future climate scenario, with RCP4.5 increasing most significantly in the Maoxian station and RCP8.5 increasing most significantly in the other stations. These findings show that continuous precipitation in the upper reaches of the Minjiang River will further increase under the future climate scenario.

6.2. Threshold Indices (R10 mm and R25 mm). Under the future climate scenario, R25 mm largely decreases, particularly in the Songpan, Wenchuan, and Lixian stations. R10 mm is largely rising, particularly in the RCP8.5 scenario. This result shows that under the future climate scenario, the frequency of strong precipitation in the upper reaches of the Minjiang River will increase, while the frequency of heavy precipitation will decrease.

6.3. Absolute Value Indices (RX1DAY and RX5DAY). Under the future climate scenario, RX1DAY decreases, particularly in RCP2.6 and RCP4.5, whereas RX5DAY increases, particularly in RCP8.5. This result shows that the intensity of continuous heavy precipitation in the upper reaches of the Minjiang River increases, whereas the intensity of short-term heavy precipitation decreases.

6.4. Analysis of the Variation Characteristics of the Extreme Runoff Index Value. Table 10 provides the change rate of extreme runoff in the upper reaches of the Minjiang River relative to the historical period under the future climate scenario. Under the future climate scenario, AM in the upper reaches of the Minjiang River exhibits a downward trend, i.e., a reduction of 32%–42%. Among these, the decrease in Heishui station is the most evident, followed by that in Zipingpu station. When different future climate scenarios are compared, RCP4.5 presents the most noticeable downward trend, followed by RCP2.6. Q10, Q5, and Q1, which represent the extreme runoff of low flow and exhibit increasing trends, indicating that the risk of drought in the upper reaches of the Minjiang River will be weakened under different discharge scenarios in the future.

Q90 and Q95 of Heishui and Shaba stations show an increasing trend, whereas Q99 shows a decreasing trend. Q90 of Zipingpu station presents a weak upward trend (i.e., less than 5%), and Q95 and Q99 exhibit a downward trend. These results show that the overall flood risk in the upper reaches of the Minjiang River is weakened. However, the spatial distribution in the upper reaches is different, and flood risk in the upstream source area demonstrates an increasing trend.

Figure 5 shows the changes in Q90, Q95, and Q99 at Zipingpu station from May to October under the historical period and future climate scenario. The overall characteristics of Q90, Q95, and Q99 that reflect the peak runoff are as follows: they decrease from June to July, decrease weakly in May, and increase in other months. Combined with the runoff wet season in the upper reaches of the Minjiang River from May to October, extreme flood events from June to July

Table 4: Definitions of indices.

	Index	Index title	Index definition	Unit
Extreme temperature indices	TX10p	Cold days	Annual total days when maximum daily temperature is less than the 10th percentile value in 1961–2016	Days
	TN10p	Cold nights	Annual total days when minimum daily temperature is less than the 10th percentile value in 1961–2016	Days
	TX90p	Warm days	Annual total days when maximum daily temperature is more than the 10th percentile value in 1961–2016	Days
	TN90p	Warm nights	Annual total days when minimum daily temperature is more than the 10th percentile value in 1961–2016	Days
	TX_n	Minimum Tmax	The minimum of daily maximum temperature in each month	°C
	TN_n	Minimum Tmin	The minimum of daily minimum temperature in each month	°C
	TX_x	Maximum Tmax	The minimum of daily maximum temperature in each month	°C
	TN_x	Maximum Tmin	The maximum of daily minimum temperature in each month	°C
Extreme precipitation indices	CWD	Consecutive wet days	Maximum number of consecutive days with RR ≥ 1 mm	Days
	R10 mm	Number of heavy precipitation days	Annual count of days when RR ≥ 10 mm	days
	R25 mm	Number of very heavy precipitation days	Annual count of days when RR ≥ 25 mm	Days
	RX1DAY	Maximum 1-day precipitation	Annual maximum 1-day precipitation	mm
	RX5DAY	Maximum 5-day precipitation	Annual maximum 5-day precipitation	mm
Extreme flow indices	AM	Maximum 1-day flow	Annual maximum 1-day flow	m^3/s
	IM	Minimum monthly flow	Annual minimum monthly flow	m^3/s
	Q90	90% quantile runoff	10% of the daily average runoff series in the year exceed this value	m^3/s
	Q95	95% quantile runoff	5% of the daily average runoff series in the year exceed this value	m^3/s
	Q99	99% quantile runoff	1% of the daily average runoff series in the year exceeds this value	m^3/s
	Q10	10% quantile runoff	More than 90% of the annual average daily runoff of the series	m^3/s
	Q5	5% quantile runoff	95% of the daily average runoff series in the year exceed this value	m^3/s
	Q1	1% quantile runoff	99% of the daily average runoff series in the year exceed this value	m^3/s

Table 5: Correlation coefficients of measured and simulated daily maximum and minimum temperature and precipitation during the verification period.

Station	Maximum temperature	Minimum temperature	Precipitation
Maoxian	0.9***	0.9***	0.4**
Songpan	0.9***	0.8***	0.4**
Wenchuan	0.9***	0.9***	0.5**
Lixian	0.9***	0.9***	0.5**
Heishui	0.9***	0.9***	0.5**

**indicates passing the 95% significance test; ***indicates passing the 99% significance test.

Table 6: Mean absolute error of variables during the validation period.

Variables	Seasons			
	Spring	Summer	Autumn	Winter
Maximum temperature (°C)	1.4	1.4	1.4	1.4
Minimum temperature (°C)	1.4	1.3	1.4	1.5
Precipitation (mm)	1.3	2.2	1.3	0.4

Table 7: Sensitivity ranks and optimal values of the model parameters.

Parameter name	Brief description	Optimal value	T-stat	*P* value
v__GW_DELAY.gw	Groundwater delay (days)	13.06	34.63	0
v__ALPHA_BF.gw	Baseflow alpha factor (days)	0.63	−7.46	0
v__TLAPS.sub	Temperature lapse rate (°C/km)	18.31	1.53	0.13
r__SOL_K(1).sol_1	Saturated hydraulic conductivity	0.72	1.23	0.22
v__NPERCO.bsn	Nitrogen percolation coefficient	9.72	1.2	0.23
v__CH_N2.rte	Manning's n value for main channel	−5.34	−1.16	0.25
v__SPEXP.bsn	Exp. re-entrainment parameter for channel sediment routing	1.45	−0.96	0.34
v__GW_REVAP.gw	Groundwater "revap" coefficient	0.02	−1.04	0.3
v__REVAPMN.gw_1	Threshold water depth in the shallow aquifer for "revap" (mm)	2.72	0.69	0.49
v__EPCO.hru	Plant uptake compensation factor	1.27	0.66	0.51
v__PHOSKD.bsn	Phosphorus soil partitioning coefficient	147.57	0.44	0.66
r__SOL_ALB(1).sol	Moist soil albedo	0.22	0.42	0.68
v__CH_K2.rte	Channel effective hydraulic conductivity (mm/hr)	−489.23	−0.42	0.68
v__ESCO.hru	Soil evaporation compensation factor	1.09	−0.31	0.76
a__GWQMN.gw	Threshold depth of the water in the shallow aquifer required for return flow to occur	910.29	−0.29	0.77
v__SMFMN.bsn	Melt factor for snow on December 21 (mm H_2O/°C-day)	11.66	0.25	0.8
r__CN2.mgt	SCS streamflow curve number	−0.58	−0.21	0.84
v__SMFMX.bsn	Melt factor for snow on June 21 (mm H_2O/°C-day)	−0.88	0.13	0.9
r__SOL_AWC(1).sol	Available water capacity (mm H_2O/mm soil)	−3.88	−0.04	0.97
v__TIMP.bsn	Snow pack temperature lag factor	0.76	0.03	0.98
v__SURLAG.bsn	Surface runoff lag time (days)	719.8	0.01	0.99

r_ means an existing parameter value is multiplied by (1þ given value); v_ means the existing parameter value is to be replaced by given value; the detailed description of the parameters was introduced as technical report [36, 40].

will be reduced under the future climate scenario, whereas the extreme flood risk from August to October will increase. Particular attention should be given to the increase in autumn flood risk in the upper reaches of the Minjiang River.

6.5. Hydrological Frequency Analysis of Extreme Runoff under Future Climate Scenarios. Daily runoff under different future climate scenarios is simulated using the SWAT model. Annual maximum daily runoff and annual minimum monthly runoff are selected for hydrological frequency analysis. Annual maximum daily runoff can reflect the risk of flood, while annual minimum monthly runoff can reflect the risk of drought.

Hydrological frequency is analyzed using annual maximum daily runoff under the future climate scenario and the historical period. As indicated in Figure 6 and Table 11, the Monte Carlo simulation [28] shows that the optimal frequency distribution line of RCP2.6 and RCP4.5 in the historical period is generalized logic distribution (GLO), while that of RCP8.5 is Wakeby distribution.

By using the optimal frequency distribution, the annual maximum daily runoff under different discharge scenarios is calculated to be reduced compared with that of the historical period. The annual maximum daily runoff with a 20-year return period under RCP2.6, RCP4.5, and RCP8.5 discharge scenarios is reduced by 47.3%, 45.9%, and 43.3%, respectively, compared with

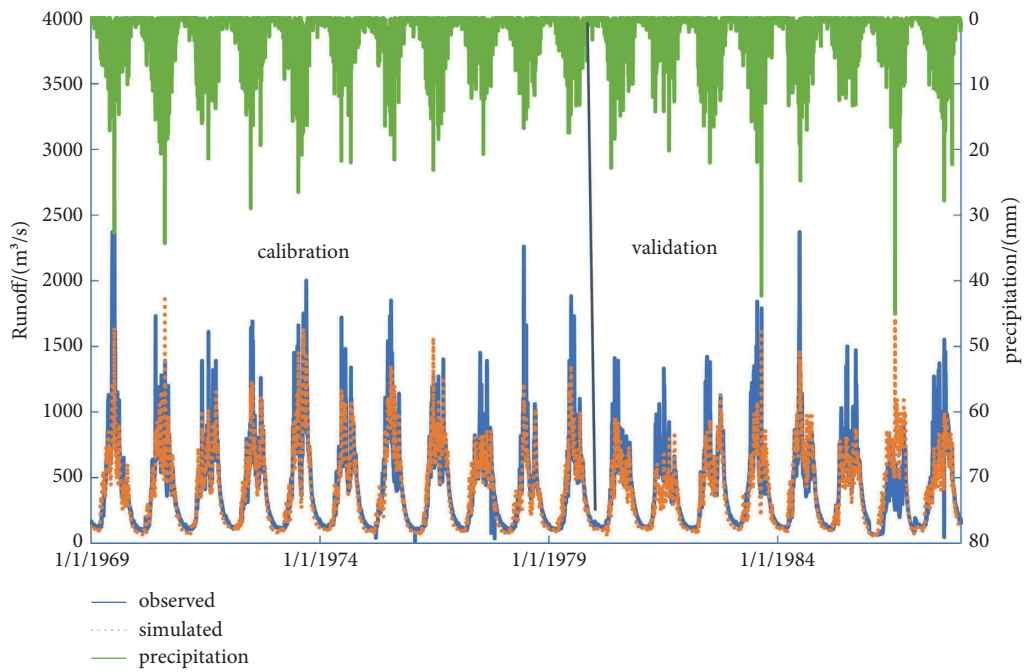


Figure 3: Daily runoff hydrograph of the Zipingpu hydrometric station during calibration and verification period.

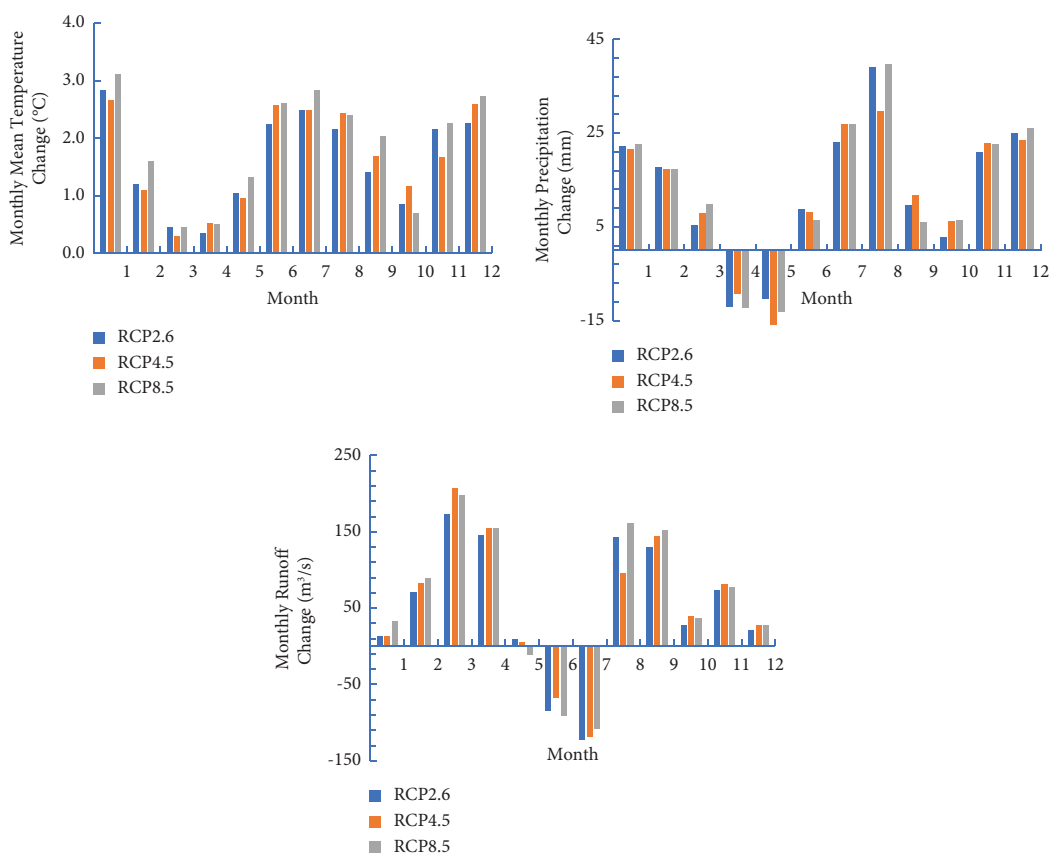


Figure 4: Changes in monthly average temperature, precipitation, and runoff compared with the base year under RCP2.6, RCP4.5, and RCP8.5 from 2021 to 2050.

Table 8: Comparison of extreme temperature index values in the upper reaches of the Minjiang River in the historical period and future climate scenario from 2021 to 2050 (unit: °C or days).

Index classification	Index abbreviation	Maoxian				Songpan				Wenchuan				Lixian				Heishui			
		Historical	RCP2.6	RCP 4.5	RCP 8.5	Historical	RCP 2.6	RCP 4.5	RCP 8.5	Historical	RCP 2.6	RCP 4.5	RCP 8.5	Historical	RCP 2.6	RCP 4.5	RCP 8.5	Historical	RCP 2.6	RCP 4.5	RCP 8.5
Relative index	TX10p	36.0	65.6	72.8	88.8	35.9	26.5	33.8	42.0	35.5	60.9	70.4	78.5	35.5	60.3	68.3	75.1	36.0	26.1	32.9	40.2
	TN10p	35.3	35.5	40.4	46.5	36.1	34.8	40.6	45.4	35.5	35.9	39.0	44.1	36.0	35.5	40.8	44.0	36.2	35.5	40.2	43.9
	TX90p	35.2	85.1	43.9	24.2	36.5	59.0	48.4	29.3	36.1	58.4	54.3	36.1	36.0	59.0	53.9	39.9	35.8	60.3	54.5	39.7
	TN90p	35.3	73.5	121	107	35.1	79.7	75.0	64.4	36.3	101	97.8	84.7	36.0	99.5	95.5	84.5	36.0	75.2	72.8	58.5
Extremal index	ID0	4.5	2.2	2.0	1.5	3.8	1.6	1.9	1.8	1.0	0.0	0.0	1.0	2.9	1.6	1.5	1.5	2.5	1.0	1.0	1.0
	FD0	75.0	49.0	52.5	50.2	163	153	152	151	34.2	9.3	8.5	6.8	77.5	61.9	62.9	57.7	125	109	108	105
	SU25	47.9	65.6	68.6	68.7	17.0	49.2	51.7	52.6	99.9	133	135	136	77.8	109	109	110	53.0	87.9	91.4	88.5
	TR20	4.5	51.3	53.8	57.8	0.0	3.5	3.9	4.6	17.8	50.3	51.2	53.2	3.2	18.8	19.3	22.7	0.0	3.4	3.4	4.2

Table 9: Comparison of extreme precipitation index values in the upper reaches of the Minjiang River in the historical period and future climate scenario from 2021 to 2050 (unit: mm or days).

Index classification	Index abbreviation	Maoxian				Songpan				Wenchuan				Lixian				Heishui			
		Historical	RCP 2.6	RCP 4.5	RCP 8.5	Historical	RCP 2.6	RCP 4.5	RCP 8.5	Historical	RCP 2.6	RCP 4.5	RCP 8.5	Historical	RCP 2.6	RCP 4.5	RCP 8.5	Historical	RCP 2.6	RCP 4.5	RCP 8.5
Sustainability index	CWD	4.1	5.6	5.7	5.5	6.1	7.5	7.6	7.8	5.1	8.3	8.3	9.4	5.5	7.4	7.4	7.7	7.0	6.2	6.3	6.3
Threshold index	R10 mm	9.1	23.7	23.7	24.9	18.0	30.7	31.5	34.6	2.5	19.1	20.5	23.1	11.1	18.1	18.0	19.7	23.5	32.8	32.2	33.7
	R25 mm	1.1	0.1	0.0	0.0	1.5	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.8	0.0	0.0	0.0	2.6	0.3	0.3	0.3
Absolute value index	RX1DAY	33.0	19.9	19.9	20.4	29.6	19.2	19.3	19.6	25.1	17.6	17.2	17.5	27.0	16.7	16.5	17.1	33.7	23.4	23.5	23.6
	RX5DAY	41.1	48.7	48.9	50.6	52.9	51.5	53.1	55.8	47.8	46.7	47.6	49.7	42.3	44.0	43.1	45.0	61.8	61.9	62.6	63.6

Table 10: Variation of extreme runoff in the upper reaches of the Minjiang River from 2021 to 2050 (%).

	Heishui			Shaba			Zipingpu		
	RCP2.6	RCP4.5	RCP8.5	RCP2.6	RCP4.5	RCP8.5	RCP2.6	RCP4.5	RCP8.5
AM	−41.1	−41.8	−38.6	−34.3	−35.1	−32.5	−36.5	−36.7	−34.8
Q90	19.5	17.5	23.3	27.1	24	30.2	2.5	0.6	4.2
Q95	5.2	3.3	8.3	10.6	9.5	13.2	−9.1	−9.3	−6.3
Q99	−17.7	−18.7	−15.6	−13.7	−14.4	−11.5	−24	−23.7	−21.2
Q10	102.4	121.8	130.5	71.1	82.6	93.8	46	48	55.4
Q5	75.2	89.8	99.5	52.7	56.6	70.3	29.7	26	39.5
Q1	51.1	60.9	67.9	29.3	33.1	42.2	16.4	14	26.2

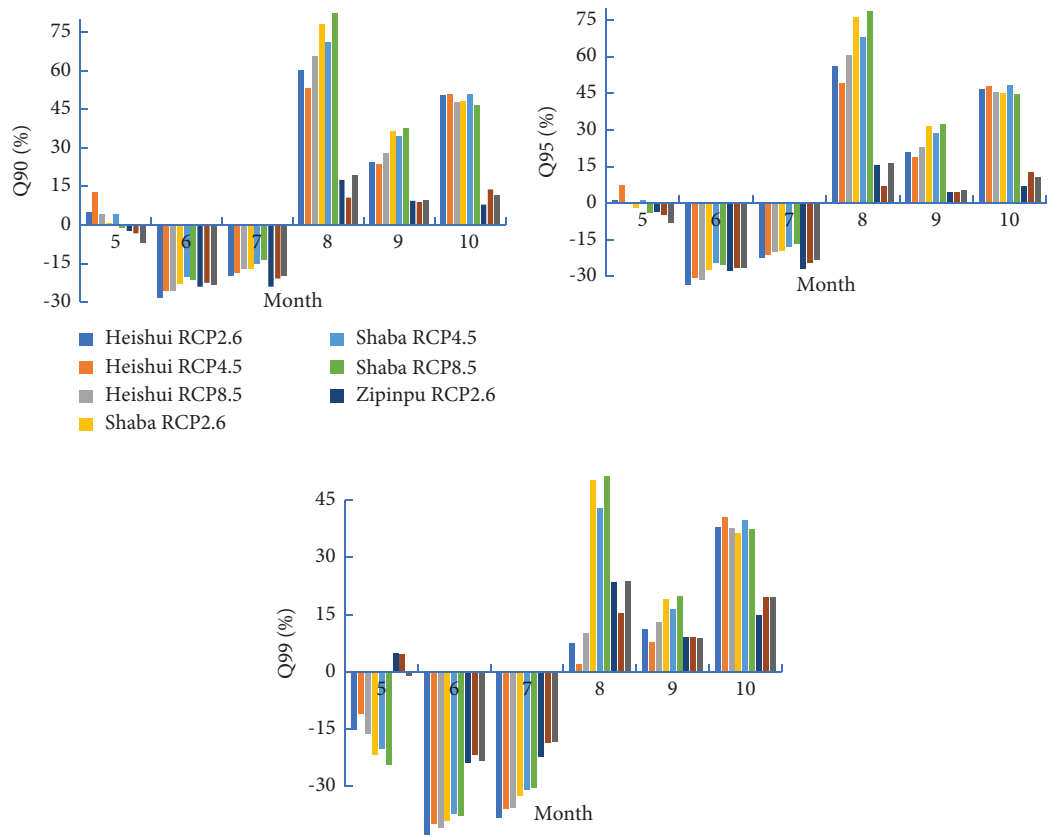


Figure 5: Variation of quantile runoff from May to October in the upper reaches from 2021 to 2050 (%).

that of the historical period. Meanwhile, the annual maximum daily runoff with a 100-year return period is reduced by 54.8%, 52.2%, and 50.6%, respectively, compared with that of the historical period. This result shows that the risk of flood under different discharge scenarios is lower than that in the historical period.

Hydrological frequency is analyzed using annual minimum monthly runoff under the future climate scenario and the historical period. As indicated in Figure 7 and Table 12, Monte Carlo simulation [28] is used to determine that the optimal frequency distribution line of RCP2.6 and RCP8.5 in the historical period is generalized logistic distribution (GLO), while that of RCP4.5 is generalized extreme value distribution (GEV).

Under different discharge scenarios, the annual minimum monthly runoff increases compared to historical period. Under RCP2.6, RCP4.5, and RCP8.5 discharge scenarios, the annual minimum monthly runoff with a 20-year return period increases by 30.9%, 28.8%, and 45.2%, respectively, compared with that of the historical period. Meanwhile, the annual minimum monthly runoff with a 100-year return period increases by 26.6%, 14.1%, and 47.0%, respectively, compared with that of the historical period. Therefore, the risk of drought in the upper reaches of the Minjiang River is reduced under different emission scenarios.

7. Results and Discussion

7.1. Discussion. The research shows that precipitation from April to May in spring in the upper reaches of the Minjiang River is lower than that in the historical period, and the duration of the low-flow season is prolonged due to the reduction of water inflow in spring, which will affect

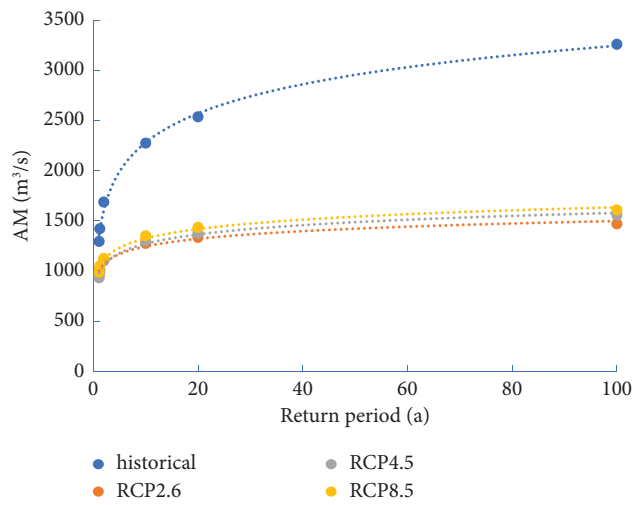


Figure 6: Annual maximum daily runoff with different return periods.

Table 11: Return period of annual maximum daily runoff.

	Optimal distribution	Return period		
		10 years	20 years	100 years
Historical	GLO	2275.2	2537.0	3259.0
RCP2.6	GLO	1276.8	1335.9	1471.9
RCP4.5	GLO	1298.1	1373.7	1559.3
RCP8.5	Wakeby	1352.5	1437.3	1610.4

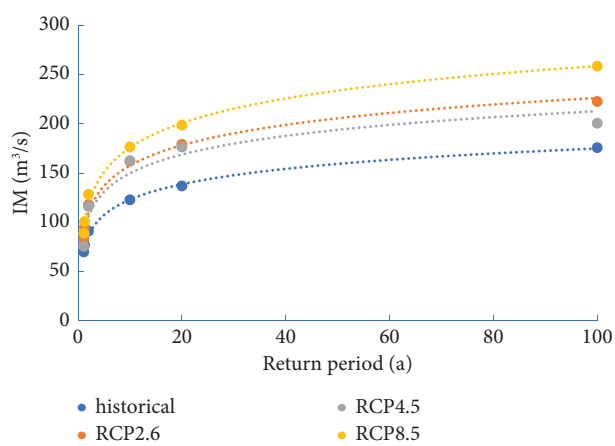


Figure 7: Annual minimum daily runoff with different return periods.

Table 12: Annual minimum monthly runoff under three discharge scenarios and the historical period (m^3/s).

	Optimal distribution	Return period		
		10 years	20 years	100 years
Historical	GLO	122.6	136.7	175.7
RCP2.6	GLO	161.8	179.1	222.5
RCP4.5	GEV	162.5	176.2	200.4
RCP8.5	GLO	176.3	198.6	258.3

irrigation and urban water supply. In the future, temperature and precipitation will increase, whereas runoff will decrease from June to July in summer. Meanwhile, precipitation will decrease and runoff will increase from April to May in spring. Several previous studies have shown that runoff change exhibits a significant positive correlation with precipitation, and the effect of temperature on runoff is more evident in dry years than in wet years [27]. This finding indicates that the effects of temperature on snow melting in spring and evapotranspiration in summer will further increase in the upper reaches of the Minjiang River under future climate change.

Under the future climate scenario, the cold index frequency of extreme temperature increases and intensity decreases. In accordance with the fact that extreme cold events in the historical period mostly occur in the upper reaches [19], extreme cold events in the upper reaches of the Minjiang River will further increase in the future. The intensity of continuous heavy precipitation under the future climate scenario will increase, and this finding is consistent with the overall change trend of the Minjiang River Basin in the historical period [20]. It indicates that continuous extreme precipitation events in the upper reaches of the Minjiang River will further increase in the future.

The research shows that the occurrence of flood events is typically related to the occurrence of short-term extreme heavy precipitation, and it is similar to the results of previous studies [7]. In the future scenario, the intensity and frequency of short-term heavy precipitation decrease, reducing the risk of extreme flood. The research shows that drought risk is related to precipitation during the dry season. In the future climate scenario, an increase in precipitation during the dry season in the upper reaches of the Minjiang River reduces the risk of drought in that area.

At present, research on the impact of climate change on industries is mostly based on climate and hydrological models, including some uncertainties [8, 11], such as the simplification of physical processes by climate and hydrological models, the introduction of parametric processes by climate models, and future climate scenario assumptions. In consideration of the aforementioned characteristics, the current study selects three scenarios (low, medium, and high) in the typical concentration path recommended in the fifth assessment report of the Intergovernmental Panel on Climate Change (IPCC) that can reflect the degree of radiation force under different emission reduction strategies in the future. In the next step, multiple climate models will be considered for uncertainty analysis.

7.2. Results. Under the future climate scenario in the upper reaches of the Minjiang River, the frequency and intensity of the extreme temperature warming index will increase while those of the cold index will increase and weaken.

Under the future climate scenario in the upper reaches of the Minjiang River, the duration of precipitation, the intensity of continuous heavy precipitation, and the frequency of strong precipitation increase, whereas the intensity of short-term heavy precipitation and the frequency of heavy precipitation decrease.

Monte Carlo simulation is used to determine that for the RCP2.6 and RCP4.5 climate scenarios under the historical period, the optimal frequency distribution line of annual maximum daily flow is GLO, while that for the RCP8.5 climate scenario is Wakeby distribution. By using the optimal frequency distribution line, the annual maximum daily flow under different climate scenarios is calculated to be lower than that in the historical period, indicating that the risk of flood in the upper reaches of the Minjiang River under different climate scenarios is lower than that in the historical period.

Monte Carlo simulation is used to determine that for the RCP2.6 and RCP8.5 climate scenarios under the historical period, the optimal frequency distribution line of annual minimum monthly flow is GLO, while that for RCP4.5 is GEV. By using the optimal distribution line, the annual minimum monthly flow under different climate scenarios is determined to increase compared with that in the historical period, indicating that the risk of drought in the upper reaches of the Minjiang River under different climate scenarios is reduced.

The indicators that represent low flow runoff exhibit an increasing trend, indicating that drought risk in the upper reaches of the Minjiang River is reduced. The indicators that represent high water runoff mostly demonstrate an increasing trend at Heishui and Shaba stations, while those at Zipingpu station present a decreasing trend, indicating that the overall flood risk in the upper reaches of the Minjiang River is reduced. Meanwhile, spatial distribution in the upper reaches is different, and thus flood risk in the upstream source area still exhibits an increasing trend.

The distribution characteristics of the indicators that represent high water runoff in the year show that extreme runoff in the upper reaches of the Minjiang River will present an upward trend from August to October in the future. Particular attention should be given to the increase in autumn flood risk in the upper reaches of the Minjiang River.

Data Availability

The data used in this paper were provided by the State Meteorological Administration of China. Relevant station data are available at https://data.cma.cn/.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was financially supported by Chengdu University of Information Engineering Scientific Research Fund Grant Results (no. KYTZ202129), Heavy Rain and Drought-Flood Disasters in Plateau and Basin Key Laboratory of Sichuan Province (no. SZKT202201), the Regional

Innovation Cooperation Program from the Science & Technology Department of Sichuan Province (no. 2020YFQ0013), and the Science and Technology Projects in Tibet Autonomous Region (no. XZ202101ZY0007G).

References

[1] Y. Chen, W. Moufouma-Okia, V. Masson-Delmotte, P. Zhai, and A. Pirani, "Recent progress and emerging topics on weather and climate extremes since the fifth assessment report of the intergovernmental panel on climate change," *Annual Review of Environment and Resources*, vol. 43, no. 1, pp. 35–59, 2018.

[2] J. Sillmann, T. Thorarinsdottir, N. Keenlyside et al., "Understanding, modeling and predicting weather and climate extremes: challenges and opportunities," *Weather and Climate Extremes*, vol. 18, pp. 65–74, 2017.

[3] Z. Ji and S. Kang, "Evaluation of extreme climate events using a regional climate model for China," *International Journal of Climatology*, vol. 35, no. 6, pp. 888–902, 2014.

[4] H. Wang, K. Fan, J. Sun et al., "A review of seasonal climate prediction research in China," *Advances in Atmospheric Sciences*, vol. 32, no. 2, pp. 149–168, 2014.

[5] P. Qin, H. Xu, M. Liu et al., "Climate change impacts on three gorges reservoir impoundment and hydropower generation," *Journal of Hydrology*, vol. 580, Article ID 123922, 2020.

[6] M. M. Gulacha and D. M. M. Mulungu, "Generation of climate change scenarios for precipitation and temperature at local scales using SDSM in Wami-Ruvu River Basin Tanzania," *Physics and Chemistry of the Earth*, vol. 100, pp. 62–72, 2017.

[7] K. Mohammed, A. S. Islam, G. T. Islam, L. Alfieri, S. K. Bala, and M. J. U. Khan, "Extreme flows and water availability of the Brahmaputra River under 1.5 and 2 °C global warming scenarios," *Climatic Change*, vol. 145, no. 1-2, pp. 159–175, 2017.

[8] J. Chen, C. Gao, X. Zeng et al., "Assessing changes of river discharge under global warming of 1.5 °C and 2 °C in the upper reaches of the Yangtze River Basin: approach by using multiple- GCMs and hydrological models," *Quaternary International*, vol. 453, pp. 63–73, 2017.

[9] B. Su, J. Huang, X. Zeng, C. Gao, and T. Jiang, "Impacts of climate change on streamflow in the upper Yangtze River basin," *Climatic Change*, vol. 141, no. 3, pp. 533–546, 2017.

[10] L. L. Liu, H. M. Xu, Y. Wang, and T. Jiang, "Impacts of 1.5 and 2°C global warming on water availability and extreme hydrological events in Yiluo and Beijiang River catchments in China," *Climatic Change*, vol. 145, no. 1-2, pp. 145–158, 2017.

[11] T. Iizumi, H. Takikawa, Y. Hirabayashi, N. Hanasaki, and M. Nishimori, "Contributions of different bias-correction methods and reference meteorological forcing data sets to uncertainty in projected temperature and precipitation extremes," *Journal of Geophysical Research: Atmospheres*, vol. 122, no. 15, pp. 7800–7819, 2017.

[12] P. E. Carvajal, G. Anandarajah, Y. Mulugetta, and O. Dessens, "Assessing uncertainty of climate change impacts on long-term hydropower generation using the CMIP5 ensemble: the case of Ecuador," *Climatic Change*, vol. 144, no. 4, pp. 611–624, 2017.

[13] E. G. Bekele and H. V. Knapp, "Watershed modeling to assessing impacts of potential climate change on water supply availability," *Water Resources Management*, vol. 24, no. 13, pp. 3299–3320, 2010.

[14] H. Zhang, G. H. Huang, D. Wang, and X. Zhang, "Uncertainty assessment of climate change impacts on the hydrology of small prairie wetlands," *Journal of Hydrology*, vol. 396, no. 1-2, pp. 94–103, 2011.

[15] R. L. Wilby and T. M. L. Wigley, "Downscaling general circulation model output: a review of methods and limitations," *Progress in Physical Geography: Earth and Environment*, vol. 21, no. 4, pp. 530–548, 1997.

[16] H. J. Fowler, S. Blenkinsop, and C. Tebaldi, "Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling," *International Journal of Climatology*, vol. 27, no. 12, pp. 1547–1578, 2007.

[17] M. M. Gulacha and D. M. M. Mulungu, "Generation of climate change scenarios for precipitation and temperature at local scales using SDSM in Wami-Ruvu River Basin Tanzania," *Physics and Chemistry of the Earth, Parts A/B/C*, vol. 100, pp. 62–72, 2017.

[18] N. N. A. Tukimat, N. A. Ahmad Syukri, and M. A. Malek, "Projection the long-term ungauged rainfall using integrated statistical downscaling model and geographic information system (SDSM-GIS) model," *Heliyon*, vol. 5, no. 9, Article ID e02456, 2019.

[19] S. Dehghan, N. Salehnia, N. Sayari, and B. Bakhtiari, "Prediction of meteorological drought in arid and semi-arid regions using PDSI and SDSM: a case study in Fars Province, Iran," *Journal of Arid Land*, vol. 12, no. 2, pp. 318–330, 2020.

[20] V. P. Pandey, S. Dhaubanjar, L. Bharati, and B. R. Thapa, "Hydrological response of Chamelia watershed in Mahakali basin to climate change," *Science of the Total Environment*, vol. 650, pp. 365–383, 2019.

[21] A. R. Bajracharya, S. R. Bajracharya, A. B. Shrestha, and S. B Maharjan, "Climate change impact assessment on the hydrological regime of the Kaligandaki Basin, Nepal," *Nepal. Sci," Total Environ*, vol. 625, pp. 837–848, 2018.

[22] A. R. Fonseca and J. A. Santos, "Predicting hydrologic flows under climate change: the Tâmega Basin as an analog for the Mediterranean region," *Science of the Total Environment*, vol. 668, pp. 1013–1024, 2019.

[23] M. Meaurio, A. Zabaleta, L. Boithias et al., "Assessing the hydrological response from an ensemble of CMIP5 climate projections in the transition zone of the Atlantic region (Bay of Biscay)," *Journal of Hydrology*, vol. 548, pp. 46–62, 2017.

[24] R. Yu, B. Wang, and T. Zhou, "Tropospheric cooling and summer monsoon weakening trend over East Asia," *Geophysical Research Letters*, vol. 31, no. 22, 2004.

[25] J. Wang, H. Yao, and T. Guan, "Analysis of skewness coefficient of rainstorm and precipitation," *Advances in Water Science*, vol. 3, pp. 365–370, 2006.

[26] S. Liang, W. Wang, and D. Zhang, "Characteristics of annual and seasonal precipitation variation in the upstream of minjiang river, southwestern China," *Advances in Meteorology*, vol. 2018, Article ID 1362708, 14 pages, 2018.

[27] X. Huang, J. Zhao, W. Li, and H. Jiang, "Impact of climatic change on streamflow in the upper reaches of the minjiang river, China," *Hydrological Sciences Journal*, vol. 59, no. 1, pp. 154–164, 2014.

[28] S. Huang, W. Zang, M. Xu et al., "Study on runoff simulation of the upstream of minjiang river under future climate change scenarios," *Natural Hazards*, vol. 75, no. S2, pp. 139–154, 2015.

[29] T. Chen, T. Q. Ao, X. Zhang, X. Li, and K. Yang, "Climate change characteristics of extreme temperature in the minjiang River Basin," *Advances in Meteorology*, vol. 201915 pages, Article ID 1935719, 2019.

[30] T. Chen, T. Q. Ao, and X. Zhang, "Spatial and temporal variation and probability characteristics of extreme precipitation events in the Min River Basin from 1961 to 2016," *Applied Ecology and Environmental Research*, vol. 17, no. 5, pp. 11375–11394, 2019.

[31] R. L. Wilby, "SDSM-a decision support tool for the assessment of regional climate change impacts," *Environmental Modelling & Software*, vol. 17, pp. 147–159, 2002.

[32] J. Rogelj, M. Meinshausen, and R. Knutti, "Global warming under old and new scenarios using IPCC climate sensitivity range estimates," *Nature Climate Change*, vol. 2, no. 4, pp. 248–253, 2012.

[33] K. Riahi, S. Rao, V. Krey et al., "Rcp 8.5—a scenario of comparatively high greenhouse gas emissions," *Climatic Change*, vol. 109, no. 1-2, pp. 33–57, 2011.

[34] A. M. Thomson, K. V. Calvin, S. J. Smith et al., "RCP4.5: a pathway for stabilization of radiative forcing by 2100," *Climatic Change*, vol. 109, no. 1–2, pp. 77–94, 2011.

[35] D. P. Van Vuuren, E. Stehfest, M. G. J. den Elzen et al., "RCP2.6: exploring the possibility to keep global mean temperature increase below 2°C," *Climatic Change*, vol. 109, no. 1–2, pp. 95–116, 2011.

[36] S. L. Neitsch, J. G. Arnold, and J. R. Kiniry, "Soil and water assessment tool theoretical documentation version," TR-406, Technical Report, Texas Water Resources Institute, College Station, TX, USA, 2009.

[37] L. V. Alexander, X. Zhang, T. C. Peterson et al., "Global observed changes in daily climate extremes of temperature and precipitation," *Journal of Geophysical Research*, vol. 111, no. D5, Article ID D05109, 2006.

[38] T. Yang, X. Chen, and H. W. Yang, "Regional flood frequency analysis in pearl river delta region based on l-moments approach," *Journal of Hohai University(Natural Sciences)*, vol. 37, no. 6, pp. 615–619, 2009.

[39] M. Y. Zhou, F. Yuan, L. M. Zhang, and X. Y. Yao, "Projection of future climate change impacts on extreme runoff in the upper reaches of ganjiang River Basin," *Water Resources and Power*, vol. 38, no. 1, pp. 5–8, 2020.

[40] F. Olivera, M. Valenzuela, R. Srinivasan et al., "ArcGIS-SWAT: a geodata model and GIS interface for SWAT," *Journal of the American Water Resources Association*, vol. 42, no. 2, pp. 295–309, 2006.

Hindawi
Advances in Meteorology
Volume 2022, Article ID 5346647, 20 pages
https://doi.org/10.1155/2022/5346647

Research Article

The Influence of Data Length on the Performance of Artificial Intelligence Models in Predicting Air Pollution

Mohamed Khalid AlOmar,[1] **Faidhalrahman Khaleel**,[1] **Abdulwahab Abdulrazaaq AlSaadi**,[2] **Mohammed Majeed Hameed**,[1,3] **Mohammed Abdulhakim AlSaadi**,[4] and **Nadhir Al-Ansari**[5]

[1]*Department of Civil Engineering, Al-Maarif University College, Ramadi, Iraq*
[2]*Department of Computer Engineering Technics, Al-Maarif University College, Ramadi, Iraq*
[3]*Department of Civil Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia*
[4]*Natural and Medical Sciences Research Center, University of Nizwa, Nizwa, Sultanate of Oman*
[5]*Civil Engineering Department, Environmental and Natural Resources Engineering, Lulea University of Technology, Lulea 97187, Sweden*

Correspondence should be addressed to Mohamed Khalid AlOmar; mohd.alomar@yahoo.com and Nadhir Al-Ansari; nadhir.alansari@ltu.se

Received 4 March 2022; Revised 8 August 2022; Accepted 22 August 2022; Published 30 September 2022

Academic Editor: Upaka Rathnayake

Air pollution is one of humanity's most critical environmental issues and is considered contentious in several countries worldwide. As a result, accurate prediction is critical in human health management and government decision-making for environmental management. In this study, three artificial intelligence (AI) approaches, namely group method of data handling neural network (GMDHNN), extreme learning machine (ELM), and gradient boosting regression (GBR) tree, are used to predict the hourly concentration of $PM_{2.5}$ over a Dorset station located in Canada. The investigation has been performed to quantify the effect of data length on the AI modeling performance. Accordingly, nine different ratios (50/50, 55/45, 60/40, 65/35, 70/30, 75/25, 80/20, 85/15, and 90/10) are employed to split the data into training and testing datasets for assessing the performance of applied models. The results showed that the data division significantly impacted the model's capacity, and the 60/40 ratio was found more suitable for developing predictive models. Furthermore, the results showed that the ELM model provides more precise predictions of $PM_{2.5}$ concentrations than the other models. Also, a vital feature of the ELM model is its ability to adapt to the potential changes in training and testing data ratio. To summarize, the results reported in this study demonstrated an efficient method for selecting the optimal dataset ratios and the best AI model to predict properly which would be helpful in the design of an accurate model for solving different environmental issues.

1. Introduction

1.1. Background. The impacts of urbanization and industrialization have resulted in increased air pollution, considered one of our time's most pressing public health challenges [1]. Pollution can occur both indoors and outdoors [2, 3]. They are both equally dangerous, despite their differing sources. The main difference between indoor and outdoor pollution is that indoor pollution may be addressed using air filters and odour absorbers, whereas there are no effective means for monitoring and detecting air pollution, which, in turn, can then be prevented [2]. Several studies suggest that by 2050, 70% of the global population will reside in urban areas [4]. Therefore, an effective method of monitoring and predicting air pollution, particularly fine particulate matter (FPM), is critical [5–7]. $PM_{2.5}$ is known as

atmospheric particulate matter, and its equivalent diameter ($\leq 2.5\mu m$) enables them to be suspended in the atmosphere for an extended period. Furthermore, the chemical composition of $PM_{2.5}$ generally consists of carbon, nitrate compounds, sulfur, heavy metals, and other substances such as sea salt and sand [8], leading to various respiratory diseases, nervous system damage, cancer, cardiovascular diseases, etc. [9–12]. Furthermore, air pollution becomes more severe as the $PM_{2.5}$ concentration increases. Additionally, worldwide about 3.15 million mature deaths each year are caused by exposure to high concentrations of $PM_{2.5}$. Overall, outdoor pollution causes 3.3 million deaths yearly [13]. Consequently, accurate predictions of $PM_{2.5}$ concentrations are critical for enhancing the public health system and developing an early warning system that predicts pollutant levels. Besides, the advanced warning system can significantly help people, especially those suffering from chronic diseases, to avoid exposure to air pollutants at peak times when pollution reaches high levels that affect their health.

1.2. Previous Works. In the last decades, several studies have been conducted to predict $PM_{2.5}$. Moreover, these studies can roughly be categorized into conventional (deterministic and statistical) and artificial intelligence approaches (AI). The deterministic approach is based on weather research and predicting (WRP) and community multiscale air quality (CMAQ) models [14]. Additionally, calculations based on the deterministic model can account for abrupt changes in weather phenomena that cause the diffusion of atmospheric particles and perform well over extended periods [15]. Besides, the deterministic approaches rely on numerical simulation to obtain large-scale results. However, these models are time-consuming because they require many computational resources, limiting their comprehensive implementation [16]. On the other hand, statistical models such as nonlinear regression [17], classification and regression trees linear model-Kalman filter analog combination [18], autoregressive integrated moving average [19], exponential smoothing with drift model [20], and combination model [21] are more efficient as well as quicker and easier than deterministic models [22]. However, the performances of the statistical models are relatively poor since the characteristics of $PM_{2.5}$ are volatile for different samples [23, 24]. Zhang et al. [19] used an autoregressive integrated moving average (ARIMA) to evaluate and predict the trend of $PM_{2.5}$ concentrations. However, the result showed that the ARIMA model was outdated, which reduced the model's accuracy. Thus, the model enabled only to predict the trend. Furthermore, several factors influence the complicated formation of $PM_{2.5}$ such as meteorological factors (e.g., wind speed, humidity), population, and road network. The relation between these factors is highly nonlinear and complicated, making them almost impossible to be captured using conventional methods [25, 26].

Machine learning (ML) has made tremendous progress in recent years in solving numerous engineering in general [27–32] and $PM_{2.5}$ concentration in particular [33–42]. ML combines data science, statistics, and computing in an interdisciplinary fashion. Furthermore, regarding $PM_{2.5}$ concentration prediction, ML methods have been shown to perform better than traditional statistical models since they can handle nonlinear relationships and interactions between variables [43, 44]. Furthermore, ML methods are valuable tools for tracking pollution baseline and have been proven to identify pollution hotspots accurately. Moreover, many variables from air quality and metrological data can be analyzed using these techniques to enhance the understanding of their patterns and predict weather phenomena such as haze, air pollution, and visibility. Shang and He [45] combined random forest (RF) and ensemble neural network to predict the hourly $PM_{2.5}$ concentrations. The proposed modeling method performed well. Furthermore, Wang et al. [46] used RF, multiple linear regression (MLR), and an ensemble model that combines RF and MLR to forecast the indoor hourly $PM_{2.5}$ concentrations. Different metrological and air quality parameters are considered to develop the proposed models. The results showed that the ensemble model provided better accuracy than the stand-alone models. Additionally, the results showed that the RF model has significant potency in predicting $PM_{2.5}$ concentrations. Murillo et al. [47] proposed three machine learning models, namely artificial neural network (ANN), support vector regression (SVR), and a hybrid model that combines the SVR model with a particle swarm optimization algorithm (PSO) to predict one day in advance of $PM_{2.5}$ concentrations. The models were developed using various air quality and metrological parameters. The result showed that the hybrid model showed better performance in predicting $PM_{2.5}$ concentrations compared to the other models. The hybrid models can find more efficient solutions than traditional ones [48]. In other words, the researchers usually incorporate the bio-inspirited algorithm with the classical models to enhance these models' capability and hence achieve excellent predictive results [49, 50]. Furthermore, these algorithms more frequently are given particular roles, such as optimizing the hyperparameters of the model, which are very difficult to compute via traditional methods.

Moisan et al. [51] compared the performance of three machine learning models, namely dynamic multiple equations (DME), seasonal ARIMA model with exogenous variables (SARIMAX), and ANN, in predicting the hourly $PM_{2.5}$ concentrations. For model development, different historical pollution and metrological parameters were considered inputs for the proposed models. The results showed that the DME model performed better than the other models during the severe episodes. For more examples, Table 1 shows a brief of the applied AL approaches in the $PM_{2.5}$ concentration predictions. Based on the reviewed papers in that table, the researchers did not pay considerable attention to the data division through training the AI and statistical models. Few ratios (70/30, 80/20, and 90/10) were employed to split the data into training and testing datasets to assess applied models' performance. Moreover, the proper data division into the training and testing datasets can significantly influence the model efficiency. In other words, increasing the length of the training dataset would make the model overfit the data.

Table 1: Examples of different AI approach in $PM_{2.5}$ prediction.

Ref.	Used models	Data division	Best model	Statistical matrices
[1]	Random forest, XGBoost, and deep learning	70% for training; 30% for testing	XGBoost	R, MAE, RMSE
[43]	Random forest	10-fold cross-validation	RF	R^2, RPE
[52]	SVM, ANN	80% for training; 20% for testing	ANN	R, MSE
[53]	EEMD-GRNN ANFIS MLR MLP	90% for training; 10% for testing	ANFIS	R, R^2, MAE, RMSE
[54]	GTWR LR ANN ANIFS GRNN	—	GRNN.	R, MAE, RMSE
[55]	ANN MLR	10-fold cross-validation	ANN	R, MAE, MAPE, RMSE, IA
[56]	EEMD-GRNN GRNN MLR PCR ARIMA	90% for training; 10% for testing	EEMD- GRNN	MAE, MAPE, RMSE, IA
[57]	MLP	70% for training; 15% for validation; 15% for testing	MLP	MAE, RMSE, IA
[58]	XGBoost includes NELRM	10-folds cross-validation	XGBoost.	R^2, RMSE, MAPE
[59]	RF model	10-folds cross-validation	RF	R^2, RMSPE, MPE

Nevertheless, insufficient data for training the model may significantly impact prediction accuracy, dramatically lowering the chance of receiving valid estimates.

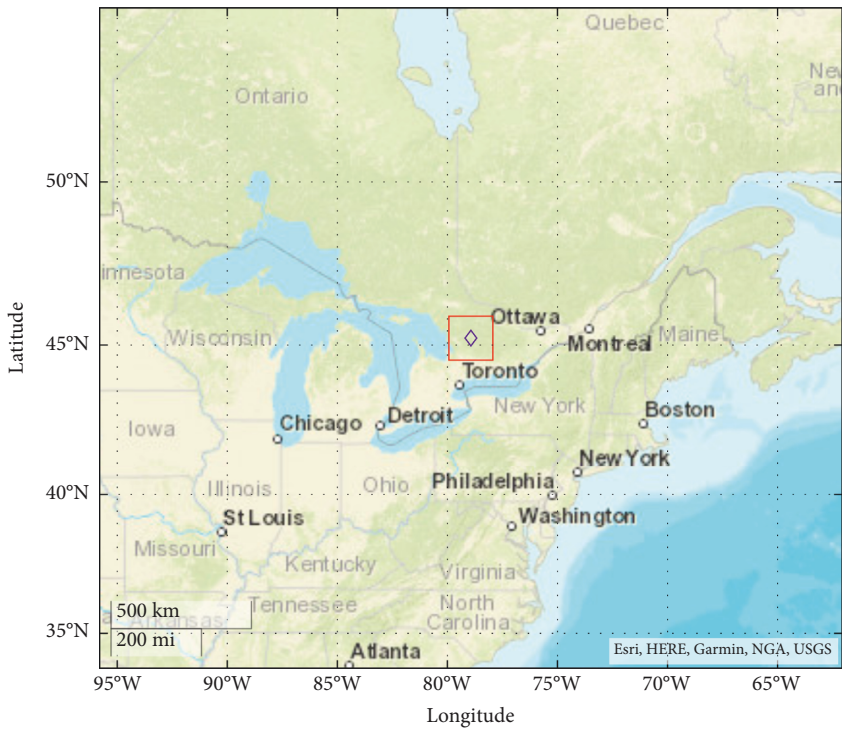
1.3. Research Motivation. Owning to the accurate prediction of $PM_{2.5}$ is very important is very important for the mangers to be alert, establish a robust system for early warning, and minimize adverse health effects and associated costs; this study investigates the influence of data partitioning on the models' efficiency. To the best of authors' knowledge, the investigation of selecting the best training and testing data ratio is not conducted yet. The reported approach in Table 1 explored that the researchers preferred to predict air pollution using ANN-based models. However, new versions of ANN modeling approaches such as ELM were not applied for air pollution forecasting. In addition, models such as ELM, GMDHNN, and GBR, despite their wide popularity in solving complex engineering problems [27, 29, 60–62], were not used in previous works to predict the concentration of $PM_{2.5}$. Therefore, these modeling approaches and their capacity have been explored in more detail.

2. Methodology

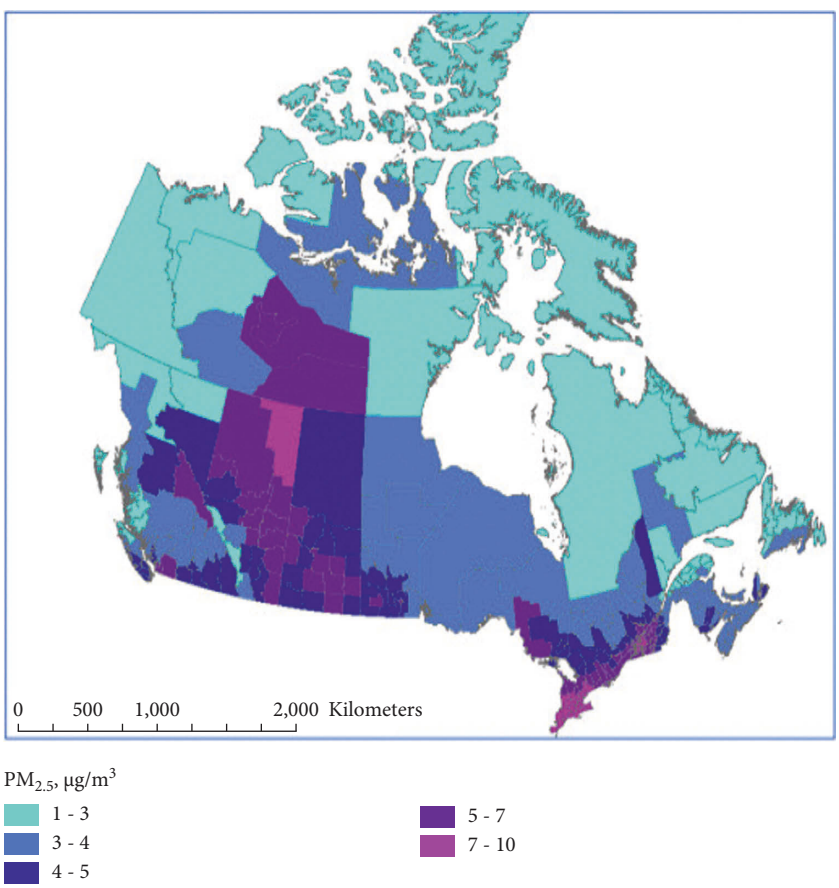
2.1. Case Study and Data Collection. In this study, the hourly $PM_{2.5}$ concentration data from Dorset station from January 01, 2011 to December 31, 2020 are collected. Dorset station is located in Ontario city with a latitude of $45^{\circ}13'27''N$ and longitude of $-78^{\circ}55'59''W$, Canada. The location of the study area and the location of the studied station and the distribution of $PM_{2.5}$ pollution over Canada are provided in Figures 1(a) and 1(b), respectively. More information about the studied station and the statistical description of the data is presented in Table 2. Furthermore, Figure 2 shows Dorset ambient air monitoring station.

2.2. Data Cleaning. Pollutant data such as $PM_{2.5}$ are usually measured using several equipment or sensors. Despite this, sensors are susceptible to hardware issues like power failures, maintenance, and unstable network equipment and hence lead to producing missing measures, zero values, negative values, null values, or others that exceed the normal range. Consequently, the accuracy of model predictions may be affected if data containing defects are directly used as input.

In this study, the percentage of missing data of $PM_{2.5}$ is significantly low (1.78%). To compensate for the missing values, the linear interpolation of neighboring and piecewise cubic spline interpolation methods is used in this study before making predictions. However, the piecewise cubic spline interpolation approach provided unrealistic and negative values, making it unreliable in compensating for missing air pollution data values. In addition, the results of this study are consistent with other studies regarding the unreliability of the piecewise cubic spline interpolation approach in compensating for the missing values [64]. As a result, linear interpolation is more efficient in replacing the missing $PM_{2.5}$ values.



(a)



(b)

Figure 1: (a) The location of the Dorset station used in the study. (b) $PM_{2.5}$ concentrations over Canada [63].

Table 2: The statistical description of the used data.

Parameter	Description
Station name	Dorset (49010)
Latitude	45.22428
Longitude	−78.9329
Location	1026 Bellwood Acres Rd., Ontario, Canada
Elevation	318 meters
Station type	Rural
Height of air intake	3 meters
Pollutants measured	$PM_{2.5}$
Unit	Micrograms per cubic meter
Measuring scale	Hourly
Used data	Jan-01 2011 to Dec-31 2020
Statistical Description	
Mean	4.7332
Standard deviation	3.9876
Maximum	65
Minimum	0

Figure 2: Dorset Ambient AirMonitoring. Source (http://www.airqualityontario.com/history/station.php?stationid=49010).

Besides, this method is chosen because the range of missing values is small, making it easy to recover the hour's conditions from the data. The adopted approach formula can be described as follows:

$$y(h+n) = y(h) + \frac{n}{g+1}\left(y(h+g+1) - y(h)\right), \tag{1}$$

where $y(\cdot)$ is the time-series target, g is the time-series duration, and n is the prediction n^{th} item of missing value, where $g-1 \geq n \geq 1$. Moreover, h corresponds to the previous normal data before the range of missing points.

2.3. Extreme Learning Machine. The ELM is considered a new robust and simple learning algorithm designed by Huang et al. [65] for a single hidden layer feedforward neural network (SLFNN). Unlike the gradient algorithms, the ELM learning speed is significantly faster at the same time, providing better generalization since it does not have the complexities of local minima, learning rate, and epochs, which is considered a considerable drawback for the other models. Furthermore, the ELM model is user-friendly, easy to comprehend, and provides minimum training errors with few norm weights [66, 67]. The ELM network consists of input, hidden, and output layers. In the input layer, the data are provided to the ELM network. Among the three layers, the hidden layer is considered the most fundamental layer since the ELM computations are carried out in it, as well as it serves as a bridge between the input and output layers in which the ELM results are organized. Given R samples of a trained dataset, the mathematical expression of *SLFNN*'s output function with U hidden nodes and activation function G(x) is shown as follows:

$$f_U(x_j) = \sum_{r=1}^{U} B_r G_r(\alpha_r \cdot x_r + \beta_r), \quad j = 1, 2, \ldots, R. \tag{2}$$

The input weights α and biases β are assigned randomly for the hidden nodes, while the output weights B are calculated analytically. The equation above can be compacted in the form as follows:

$$\mathbf{Z}B = T, \tag{3}$$

where $\mathbf{Z}$ is the output matrix,

$$\begin{aligned}
&\mathbf{Z}(\alpha_1, \ldots, \alpha_U, \beta_1, \ldots, \beta_U, x_1, \ldots, x_U) \\
&= \begin{bmatrix} G_r(\alpha_1 \cdot x_1 + \beta_1) & \ldots & G_U(\alpha_U \cdot x_1 + \beta_U) \\ \vdots & & \vdots \\ G_U(\alpha_U \cdot x_U + \beta_1) & \ldots & G_U(\alpha_U \cdot x_1 + \beta_U) \end{bmatrix}_{N \times U}, \\
&B = \begin{bmatrix} B_1^T \\ \vdots \\ B_m^T \end{bmatrix}, \\
&T = \begin{bmatrix} t_1^T \\ \vdots \\ t_U^T \end{bmatrix},
\end{aligned} \tag{4}$$

where $(\cdot)^T$ refers to the transpose of the matrix. Figure 3 shows the main structure of ELM.

2.4. Group Method of Data Handling. Ivakhnenko first proposed the group method of data handling (GMDH) approach as a polynomial neural network to capture the complex relationship between the input and output in a nonlinear system [68]. Since having prior knowledge of the model is inconceivable in the mathematical model, the GMDH neural network (GMDH − NN) is utilized to overcome this issue [27]. As a result, in the GMDH − NN model, the simulation of complex systems can be carried out without needing any prior specialized knowledge. The primary notion of the GMDH − NN model is to establish an analytical function within the feedforward network (FNN), which can be achieved by utilizing the coefficients from a quadratic node transfer function derived through the regression approach. A standard GMDH − NN formula can be expressed as follows:

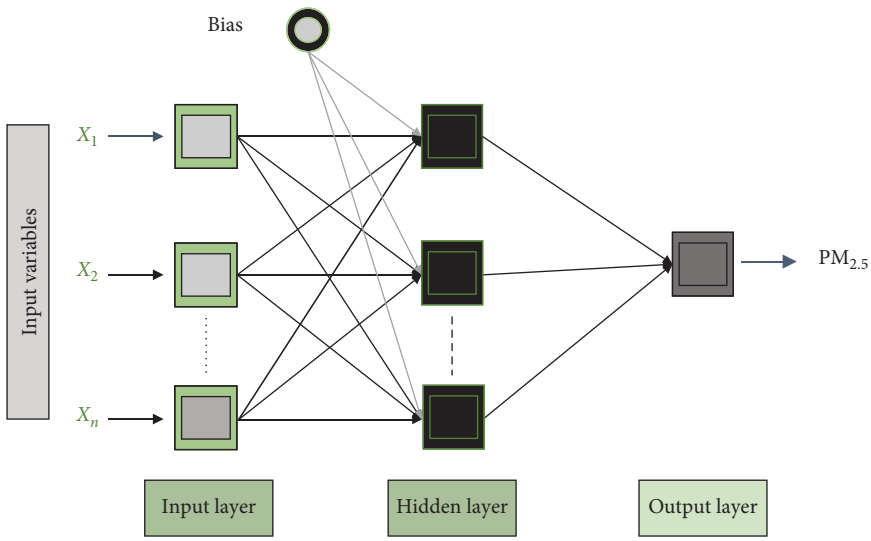


Figure 3: The structure of the ELM model.

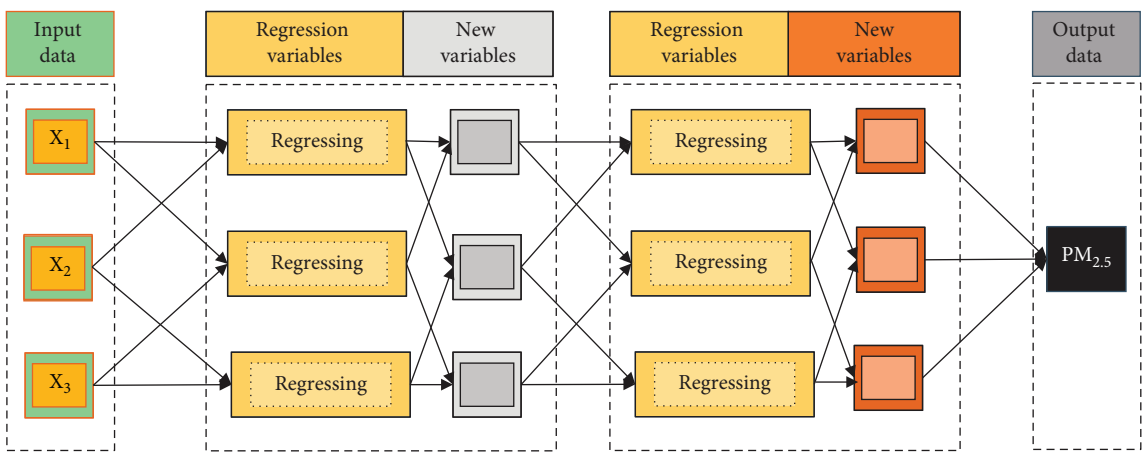


Figure 4: The structure of the GMDHNN model.

$$\gamma = r_o + r_1 x_i + r_2 x_j + r_3 x_i^2 + r_4 x_j^2 + \beta x_i x_j, \tag{5}$$

where γ is the output, and x_i and x_j present the model's inputs. The r_o, r_1, r_3, r_4, and β refer to the polynomial coefficients, which can be obtained through the training dataset. Each layer involves a set of input processing components known as nods, and the outcome of each layer is utilized as new input over the following layer. In order to optimize the weights, the least squares (LS) are adopted to acquire the minimum residual between the actual and the predicted values. Figure 4 shows the structure of the GMDH – NN model.

2.5. Gradient Boosting Regression Trees. Gradient boosting regression tree (GBR) combines the advantages of the boosting approach and decision trees to overcome classification and regression problems. The general notion of GBR is the combination (through boosting approach) of a series of decision trees known as weak learners to obtain an ensemble with multiple decision trees (strong learners), which in turn will increase the accuracy and the performance of the model. The boosting approach involves adding extra trees to the sequence without altering the model parameters that have already been added to minimize the loss function $\cup(\cdot)$ for the model. In other words, the training samples' weights are modified in accordance with the last iteration, and the weights are increased for the observations that are hard to predict while lessened for those well-handled. Assuming $f(x)$ is the approximation function, and x is set of predictors, and utilizing N additive functions, the ensemble tree model can be illustrated as follows [69, 70]:

$$f(x) = \sum_{r=1}^{N} f_r(x) = \sum_{r=1}^{N} \rho_r g(x; \alpha_r), \tag{6}$$

where α_r and ρ_r represent the end nods' mean and the given weights in the rth regression tree, respectively. $g(.)$ represents the basis function's additive expansion. Using the forward "stage – wise" approach, the parameters α_r and ρ_r are optimized. The estimate function can be illustrated

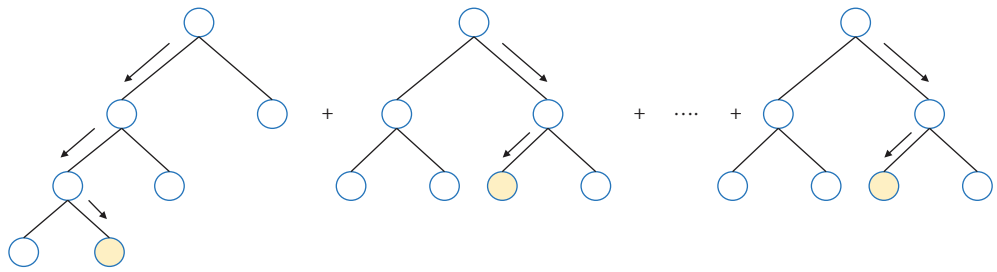


Figure 5: The structure of the GBRT.

through (7) after i^{th} number of iterations, and the optimal ρ_i is obtained using (8).

$$f(x) = f_{i-1}(x) + \rho_i g(x; \alpha_i), \tag{7}$$

$$\begin{aligned} \rho_i &= \arg\min \sum_{l=1}^{L} \cup (y_l, f_i(x_l)) \\ &= \arg\min \sum_{l=1}^{L} (y_l - f_i(x_l))^2, \end{aligned} \tag{8}$$

where $\cup(\cdot)$ is the loss function, L represents the number of observations, x_l represents the predictors set for a given l^{th} observations, and y_l represent the response variable for a given l^{th} observations. Figure 5 shows the structure of GBR model.

2.6. Model Development Performance Evaluation. Three artificial intelligence models, namely extreme learning machine (ELM), group method of data handling neural network (GMDHNN), and gradient boosting regression tree (GBR), are used to predict the hourly concentration of $PM_{2.5}$ over a Dorset station located in Canada. Before training the AI models, it is crucial to replace the missing dataset and determine the proper input-lagged vectors. Notably, the missing records have been replaced using two methods, as shown in the previous section of this study. Furthermore, the autocorrelation function (ACF) and partial autocorrelation function (PACF) are used. Autocorrelation and partial autocorrelation functions are fundamental tools in the analysis of linear time series. The ACF measures the correlation between values and the series' current value at various time points. More specifically, it indicates how similar the observations are considering their time lag. The PACF measures the correlation between values at various time points and the series' current value by partially removing the effects of the intermediate values. According to Figure 6, three input combinations can be used. The possible input groups can be shown below and can be used to predict one hour ahead of $PM_{2.5}$.

$$\begin{aligned} M1 &= f(\mathrm{PM2.5}_{t-1}), \\ M2 &= f(\mathrm{PM2.5}_{t-1}, \mathrm{PM2.5}_{t-2}), \\ M3 &= f(\mathrm{PM2.5}_{t-1}, \mathrm{PM2.5}_{t-2}, \mathrm{PM2.5}_{t-3}). \end{aligned} \tag{9}$$

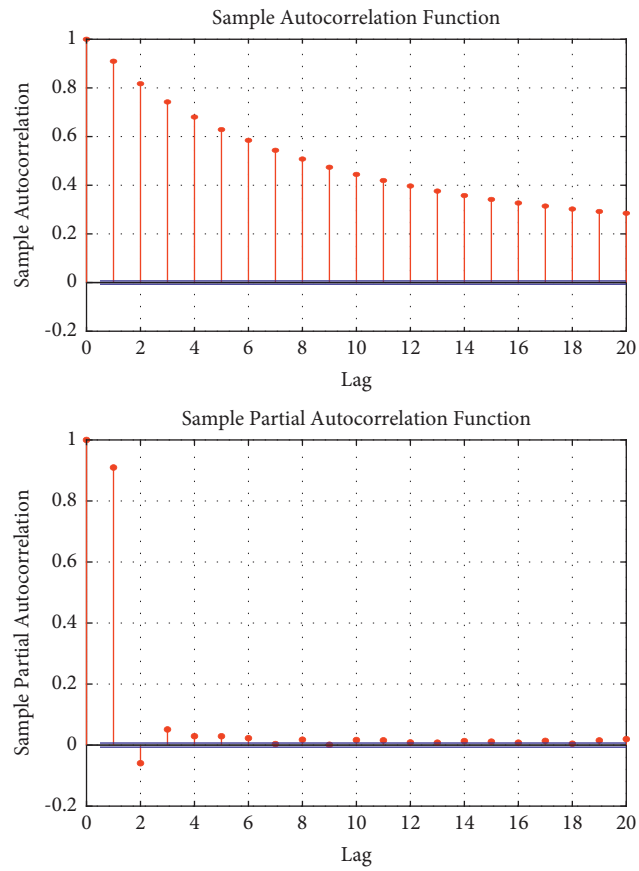


Figure 6: Autocorrelation and partial autocorrelation functions for observed hourly $PM_{2.5}$ time series.

Based on the view of ACF, many possible variables can be used as inputs, and, however, there can be found that these variables were significantly correlated with each other. Thus, PACF was used to select the most significant inputs.

After selecting the input groups, it is essential to determine the possible training/testing ratios. The dataset's length considerably affects the AI models' performances. This study employed nine different ratios (see Figure 7) to split the data into training and testing datasets to assess the applied model's performance. It is worth mentioning that the hyperparameters of the applied models were selected using the trial-and-error method because there is no straightforward approach to compute these critical

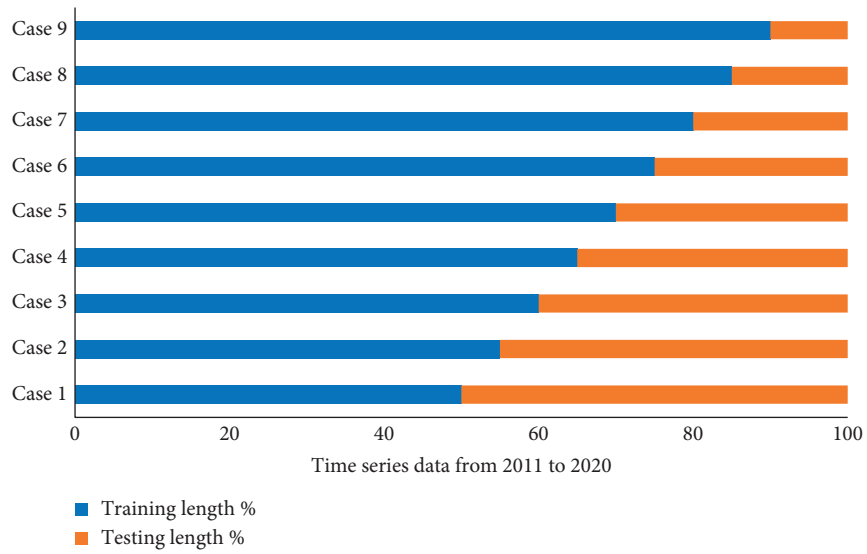


Figure 7: The nine different ratios that were utilized to split the data into training and testing datasets.

parameters, which have considerable effects on the estimation accuracy. Figure 8 shows the main processes that are used in this research. The block diagram in Figure 8(a) shows the seven essential steps related to the research methodology, whereas more details on the models' development are given in Figure 8(b).

2.7. Performance Evaluation. To assess the performance of the proposed models in $PM_{2.5}$ prediction, different statistical matrices are employed as shown below [71–73]:

(i) Mean absolute error (MAE)

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|\mathrm{PM}^{i,m} - \mathrm{PM}^{i,p}\right|. \tag{10}$$

(ii) Root mean square error (RMSE)

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\mathrm{PM}^{i,m} - \mathrm{PM}^{i,p}\right)^2}. \tag{11}$$

(iii) Correlation coefficient (R)

$$R = \frac{\sum_{i=1}^{n}\left(\mathrm{PM}^{i,m} - \overline{\mathrm{PM}}^{i,m}\right)\left(\mathrm{PM}^{i,p} - \overline{\mathrm{PM}}^{i,p}\right)}{\sqrt{\sum_{i=1}^{n}\left(\mathrm{PM}^{i,m} - \overline{\mathrm{PM}}^{i,m}\right)^2\sum_{i=1}^{n}\left(\mathrm{PM}^{i,p} - \overline{\mathrm{PM}}^{i,p}\right)^2}}. \tag{12}$$

(iv) Willmot index (WI)

$$\mathrm{WI} = 1 - \frac{\sum_{r=1}^{n}\left(\mathrm{PM}^{i,p} - \overline{\mathrm{PM}}^{i,m}\right)^2}{\sum_{i=1}^{n}\left(\left|\mathrm{PM}^{i,p} - \overline{\mathrm{PM}}^{i,m}\right| + \left|\mathrm{PM}^{i,m} - \overline{\mathrm{PM}}^{i,m}\right|\right)^2}. \tag{13}$$

(v) Nash-Sutcliffe efficiency (NSE)

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{n}\left|\mathrm{PM}^{i,m} - \mathrm{PM}^{i,p}\right|}{\sum_{i=1}^{n}\left|\mathrm{PM}^{i,p} - \mu\right|}, \tag{14}$$

where $\overline{\mathrm{PM}}^{i,m}$ and $\overline{\mathrm{PM}}^{i,p}$ are the average of the measured and the predicted $PM_{2.5}$ values, respectively; $\mathrm{PM}^{i,m} - \mathrm{PM}^{i,p}$ represent the measured and predicted values of $PM_{2.5}$ for n number of total observations; and μ is the mean deviation of the measured value.

3. Results and Discussion

This section discusses the performance of the proposed models in forecasting the hourly concentration of $PM_{2.5}$ over a long period (from 01/01/2011 to 31/12/2020). Besides, three input combinations and nine data length scenarios have been used to train and validate the models (ELM, GMDHNN, and GBR). The performance of the models through the training phase is given in Table 3. According to the results of the training phase, the ELM provides the most efficient predictions with the lowest forecasting errors (MAE ≈ 0.9710 to 1.1099; RMSE ≈ 1.6088 to 1.8329). Nevertheless, the general performance of the GBR is unsatisfactory compared to other models, providing higher errors (MAE ≈ 3.7064 to 7.5851; RMSE ≈ 4.8536 to 9.7894). The third model (GMDHNN) yields a satisfactory prediction capacity (MAE ≈ 0.9898 to 1.1187; RMSE ≈ 1.6495 to 1.8372) than the GBR model, but its performance is still lower than the ELM model through the training phase. The statistical parameters (i.e., RMSE and MAE) prove that the ELM has an outstanding capability, providing excellent estimates despite the considerable change in the input variables and length of the training dataset. On the other hand, the GBR model shows poor performance and an inability to deal with the extensive dataset. A further remarkable observation that can be deduced from the reported

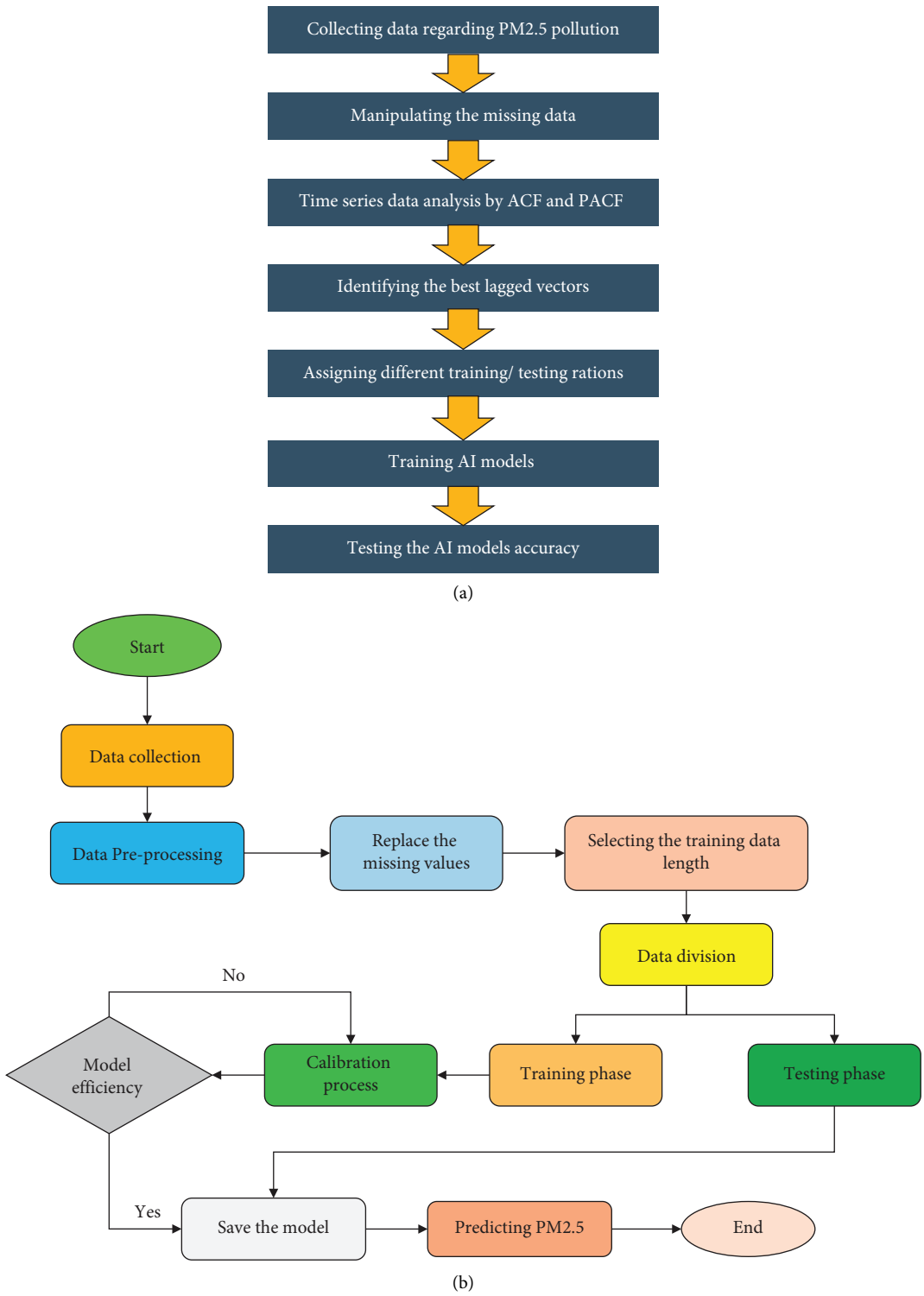


Figure 8: (a) Block diagram of the process of the study. (b) The flowchart of the process of developing the proposed models.

results is that the performances of the machine learning model (ELM and GMDHNN) got reduced when the training dataset was at 50% of the total observations. For the case of 50% of data being used for the training, both models illustrate the difficulty of estimating the $PM_{2.5}$ using a single input parameter. After evaluating the forecasted errors, it is

essential to analyze how the estimated observations of $PM_{2.5}$ are correlated with their corresponding values. in this regards, many performance metrics are performed, namely, the Willmott index (WI), correlation coefficient (R), and Nash-Sutcliffe efficiency (NSE), as presented in Table 3. Overall, the results demonstrate that the ELM model can provide higher accurate estimates in all cases than other models. In other words, the similarity between observed and predicted values by the ELM approach is promising. The R, WI, and NSE criteria for all cases range from 0.906 to 0.9171, 0.9489 to 0.9553, and 0.942 to 0.938, respectively. Similarly, the GMDHNN model yields a good prediction but is slightly lower than the ELM models. On the other hand, the GBR model cannot mimic the fluctuation of the $PM_{2.5}$ concentrations over time, providing poor estimates with WI ranging from 0.6228 to 0.8109 and NSE ranging from 0.798 to 0.547.

As the ELM approach presents excellent performance through the training phase, it is essential to validate this model using testing data points. Several studies emphasized that comparable models can be evaluated more effectively through the testing phase [28, 74]. The reason is that the model in the training phase would be trained in the presence of input points and their corresponding values. On the contrary, the applied model receives only input vectors in the testing phase. Table 4 shows the performance evaluation of the proposed models through the testing phase, and it can be seen that the ELM model outperforms the other proposed models. In other words, the ELM provides estimations that are significantly similar to their actual ones ($R \approx 0.9001$ to 0.9297; WI ≈ 0.9461 to 0.9573; NSE $\approx$ 0.9371 to 0.9281) with lower forecasted error forecasted errors (RMSE ≈ 1.4049 to 1.5327; MAE ≈ 0.9001 to 0.9207) compared to GBR and GMDHNN models. The results also show that the best second model is GMDHNN, but its efficiency in dealing with fluctuated data like $PM_{2.5}$ is not accurate as the ELM model. However, the GBR model faces problems in capturing the dynamics of $PM_{2.5}$ over the time period.

3.1. The Effect of Data Length on the Predictive Models' Performances. This part of the study shows how the input variables and length of the testing dataset affect the applied models' capacity to predict $PM_{2.5}$. In general, AI models require sufficient records and enough input vectors to provide more accurate estimations. In this regard, this study offers forty-five different scenarios of input parameters and the dataset's length as shown in Table 4. The results through the testing phase are considered to analyze the models' performances. The results showed that the ELM is more flexible with data size changes and lagged inputs than other models according to the statistical parameters such as R, RMSE, WI, and MAE. Moreover, the ELM requires only two input vectors when the testing data size ranges from 50% to 25% and adapts very well to the increasing changes in the data length. Also, the results demonstrate that if the testing data are reduced below 25%, the model requires more input vectors. As the length of testing data decreases (i.e., 20%, 15%, and 10%), the training data employed in the model increases, and thus, the training algorithm requires more inputs to complete the training and calibration processes efficiently and elaborately. Accordingly, the proposed model has high flexibility in the changes concerning the length of data and the number of the used inputs. According to the reviewed results obtained from the ELM model, it can be said that this model can provide more accurate results when the testing data size ranges from 40% to 45% of the entire dataset.

The other comparable models, such as the GBR model, do not have a reasonable or deducible pattern in dealing with cases where there is a change in the percentage of training data and the number of inputs. On the contrary, the last model (GMDHNN) tends to have a single pattern that can be deduced by evaluating its performance through statistical coefficients. This model needs, in most cases, the largest possible number of inputs, and therefore it does not show any flexibility for small and large changes that occur in the volume of data used.

For further assessment, the 95% uncertainty criterion (U_{95}) is a very effective tool for selecting the most effective and reliable model [28]. Taking into consideration the RMSE and SD (standard deviation of the computed errors), the mathematical expression of the U_{95} is as follows [75]:

$$U_{95} = 1.96\sqrt{\mathrm{RMSE}^2 + \mathrm{SD}^2}. \tag{15}$$

For different splits and input lags of the testing dataset, Figure 9 shows the evaluation of proposed models using U_{95}. The results demonstrate that the ELM provides the smallest value of U_{95} compared to other models. Furthermore, Figure 9 are consistent with the findings from the statistical parameters, which indicate that the effectiveness and accuracy of the ELM model reach their maximum when the training data represent from 40% to 45% of the total data. However, the GBR has recorded the highest value of uncertainty (U_{95}) followed by GMDHNN models.

It is essential to check the reliability analysis (RA) of the comparable models. This type of analysis is considered very effective in evaluating the consistency and performance of the models. This novel statistical metric provides essential information that can be used in determining whether the suggested models achieve the minimum requirement of allowable level of accuracy. Therefore, the RA is quite helpful in deciding and nominating the best model for air quality prediction. The formula below shows the mathematical expressions for calculating the RA [76]:

$$\mathrm{RA} = \frac{100\%}{n}\sum_{i=1}^{n} S_i. \tag{16}$$

In (16), the n is the total number of $PM_{2.5}$ samples, and S_i is the equivalent factor for each sample and its value is either 1 or 0. Furthermore, the S_i depends mainly on the percentage relative error (RE), which is mathematically derived in the following equation.

Table 3: The performance of the proposed models through the training phase.

	MAE ($\mu g/m^3$)								
Length of data %	GBR-M1	ELM-M1	GMDHNN-M1	GBR-M2	ELM-M2	GMDHNN-M2	GBR-M3	ELM-M3	GMDHNN-M3
50	7.2799	1.1099	1.1187	7.4488	1.0889	1.1134	6.8659	1.0828	1.1131
55	3.7064	1.0937	1.1008	7.2083	1.0738	1.0966	7.5851	1.0643	1.0936
60	7.5298	1.0721	1.0772	7.5252	1.0525	1.0729	5.9251	1.0480	1.0711
65	7.4281	1.0501	1.0556	4.2015	1.0289	1.0515	7.4239	1.0260	1.0484
70	6.6336	1.0317	1.0365	6.9824	1.0133	1.0333	4.9280	1.0063	1.0293
75	7.5004	1.0206	1.0258	4.2329	1.0040	1.0213	6.8401	1.0085	1.0178
80	7.4874	1.0071	1.0117	7.4851	0.9895	1.0074	6.0359	0.9852	1.0038
85	6.2630	0.9919	0.9954	7.3971	0.9818	0.9985	7.3912	0.9767	0.9936
90	7.3670	0.9916	0.9955	7.3636	0.9760	0.9910	7.3607	**0.9710**	0.9898
Min	3.7064	0.9916	0.9954	4.2015	0.9760	0.9910	4.9280	0.9710	0.9898
Max	7.5298	1.1099	1.1187	7.5252	1.0889	1.1134	7.5851	1.0828	1.1131
R									
50	0.9063	0.9062	0.9057	0.9111	0.9106	0.9071	0.9100	0.9137	0.9076
55	0.9051	0.9063	0.9060	0.9092	0.9111	0.9074	0.9108	0.9145	0.9084
60	0.9072	0.9067	0.9064	0.9145	0.9111	0.9077	0.9083	0.9133	0.9087
65	0.9082	0.9076	0.9073	0.9004	0.9131	0.9087	0.9161	0.9143	0.9096
70	0.9084	0.9085	0.9082	0.9115	0.9126	0.9094	0.9098	0.9158	0.9103
75	0.9100	0.9095	0.9092	0.9100	0.9135	0.9108	0.9133	0.9134	0.9116
80	0.9099	0.9102	0.9099	0.9169	0.9141	0.9113	0.9123	0.9164	0.9121
85	0.9094	0.9110	0.9107	0.9156	0.9146	0.9110	0.9164	0.9162	0.9119
90	0.9113	0.9110	0.9107	0.9170	0.9151	0.9123	0.9168	**0.9171**	0.9127
Min	0.9051	0.9062	0.9057	0.9004	0.9106	0.9071	0.9083	0.9133	0.9076
Max	0.9113	0.9110	0.9107	0.9170	0.9151	0.9123	0.9168	0.9171	0.9127
RMSE ($\mu g/m^3$)									
50	9.4115	1.8329	1.8372	9.6029	1.7913	1.8244	8.8518	1.7615	1.8197
55	4.8536	1.8102	1.8138	9.2754	1.7661	1.8003	9.7894	1.7334	1.7916
60	9.6767	1.7786	1.7817	9.6999	1.7378	1.7694	7.6149	1.7179	1.7601
65	9.5495	1.7465	1.7493	5.4772	1.6963	1.7370	9.5644	1.6853	1.7288
70	8.4649	1.7160	1.7186	8.8848	1.6792	1.7077	6.3041	1.6488	1.6994
75	9.5555	1.7032	1.7059	5.4347	1.6674	1.6919	8.6865	1.6676	1.6847
80	9.5271	1.6877	1.6904	9.5254	1.6518	1.6775	7.6736	1.6307	1.6703
85	8.0018	1.6648	1.6671	9.4256	1.6360	1.6683	9.4138	1.6209	1.6602
90	9.3997	1.6648	1.6671	9.4056	1.6275	1.6525	9.3857	**1.6088**	1.6495
Min	4.8536	1.6648	1.6671	5.4347	1.6275	1.6525	6.3041	1.6088	1.6495
Max	9.6767	1.8329	1.8372	9.6999	1.7913	1.8244	9.7894	1.7615	1.8197
WI									
50	0.6463	0.9489	0.9485	0.6413	0.9515	0.9495	0.6650	0.9533	0.9498
55	0.8190	0.9490	0.9488	0.6473	0.9518	0.9496	0.6325	0.9538	0.9502
60	0.6291	0.9492	0.9490	0.6313	0.9518	0.9498	0.6993	0.9531	0.9503
65	0.6296	0.9498	0.9496	0.7813	0.9530	0.9505	0.6323	0.9537	0.9507
70	0.6603	0.9503	0.9502	0.6464	0.9527	0.9509	0.7442	0.9546	0.9515
75	0.6238	0.9509	0.9508	0.7820	0.9532	0.9516	0.6527	0.9532	0.9520
80	0.6228	0.9513	0.9510	0.6253	0.9536	0.9519	0.6870	0.9549	0.9524
85	0.6722	0.9517	0.9516	0.6262	0.9538	0.9518	0.6266	0.9548	0.9522
90	0.6252	0.9517	0.9516	0.6271	0.9541	0.9525	0.6274	**0.9553**	0.9528
Min	0.6228	0.9489	0.9485	0.6253	0.9515	0.9495	0.6266	0.9531	0.9498
Max	0.8190	0.9517	0.9516	0.7820	0.9541	0.9525	0.7442	0.9553	0.9528
NSE									
50	0.6124	0.9409	0.9404	0.6034	0.9420	0.9407	0.6344	**0.9423**	0.9407
55	0.7981	0.9404	0.9400	0.6073	0.9415	0.9403	0.5867	0.9420	0.9404
60	0.5765	0.9397	0.9394	0.5768	0.9408	0.9397	0.6668	0.9411	0.9398
65	0.5709	0.9393	0.9390	0.7573	0.9406	0.9393	0.5711	0.9407	0.9394
70	0.6066	0.9388	0.9385	0.5859	0.9399	0.9387	0.7077	0.9403	0.9390
75	0.5533	0.9392	0.9389	0.7479	0.9402	0.9392	0.5926	0.9399	0.9394
80	0.5490	0.9393	0.9391	0.5492	0.9404	0.9393	0.6365	0.9407	0.9395
85	0.6155	0.9391	0.9389	0.5480	0.9400	0.9390	0.5484	0.9403	0.9393
90	0.5477	0.9391	0.9389	0.5479	0.9401	0.9392	0.5481	0.9404	0.9392
Min	0.5477	0.9388	0.9385	0.5479	0.9399	0.9387	0.5481	0.9399	0.939
Max	0.7981	0.9409	0.9404	0.7573	0.942	0.9407	0.7077	0.9423	0.9407

Bold values indicate the highest accuracy.

Table 4: The performance of the proposed models through the testing phase.

	MAE								
Length of data %	GBR-M1	ELM-M1	GMDHNN-M1	GBR-M2	ELM-M2	GMDHNN-M2	GBR-M3	ELM-M3	GMDHNN-M3
50	0.8557	0.8585	0.8688	0.8516	0.8500	0.8753	0.8739	0.8578	0.8754
45	1.5797	0.8500	0.8592	**0.8415**	0.8439	0.8633	0.8636	0.8473	0.8647
40	0.8550	0.8521	0.8599	0.8520	0.8434	0.8623	0.9790	0.8472	0.8649
35	0.8594	0.8565	0.8660	1.4210	0.8466	0.8678	0.8551	0.8491	0.8666
30	0.8901	0.8698	0.8774	0.8610	0.8597	0.8771	1.2031	0.8583	0.8767
25	0.8709	0.8707	0.8775	1.3650	0.8610	0.8773	0.8659	0.8681	0.8752
20	0.8889	0.8848	0.8879	0.8733	0.8697	0.8884	0.9516	0.8670	0.8863
15	0.9316	0.8882	0.8925	0.8797	0.8742	0.8923	0.8787	0.8713	0.8877
10	0.8879	0.8890	0.8947	0.8797	0.8785	0.8927	0.8873	0.8769	0.8904
Min	0.8550	0.8500	0.8592	0.8415	0.8434	0.8623	0.8551	0.8472	0.8647
Max	1.5797	0.8890	0.8947	1.4210	0.8785	0.8927	1.2031	0.8769	0.8904
R									
50	0.9153	0.9153	0.9156	0.9168	0.9191	0.9157	0.9169	0.9184	0.9162
45	0.9143	0.9161	0.9163	0.9169	0.9200	0.9173	0.9160	0.9190	0.9175
40	0.9164	0.9167	0.9168	0.9169	**0.9207**	0.9179	0.9174	0.9206	0.9182
35	0.9158	0.9161	0.9161	0.9078	0.9198	0.9172	0.9166	0.9204	0.9174
30	0.9143	0.9144	0.9144	0.9164	0.9187	0.9157	0.9150	0.9181	0.9162
25	0.9113	0.9115	0.9115	0.9113	0.9145	0.9128	0.9129	0.9145	0.9130
20	0.9076	0.9083	0.9086	0.9101	0.9122	0.9100	0.9087	0.9141	0.9103
15	0.9098	0.9110	0.9112	0.9128	0.9151	0.9123	0.9138	0.9162	0.9124
10	0.9000	0.9001	0.9004	0.9022	0.9042	0.9018	0.9006	0.9048	0.9022
Min	0.9000	0.9001	0.9004	0.9022	0.9042	0.9018	0.9006	0.9048	0.9022
Max	0.9164	0.9167	0.9168	0.9169	0.9207	0.9179	0.9174	0.9206	0.9182
RMSE									
50	1.4567	1.4499	1.4481	1.4386	1.4191	1.4473	1.4741	1.4258	1.4439
45	2.2999	1.4361	1.4350	1.4373	**1.4049**	1.4274	1.4374	1.4136	1.4264
40	1.4444	1.4420	1.4415	1.4407	1.4084	1.4329	1.6126	1.4097	1.4306
35	1.4633	1.4607	1.4606	2.1625	1.4291	1.4516	1.4570	1.4243	1.4497
30	1.5409	1.4926	1.4925	1.4915	1.4558	1.4814	1.9040	1.4610	1.4781
25	1.4906	1.4884	1.4886	2.0958	1.4641	1.4787	1.5030	1.4641	1.4764
20	1.5073	1.5015	1.4996	1.4875	1.4708	1.4886	1.6200	1.4555	1.4863
15	1.5949	1.4979	1.4966	1.4840	1.4650	1.4882	1.4753	1.4558	1.4871
10	1.5338	1.5327	1.5305	1.5183	1.5022	1.5206	1.5284	1.4974	1.5175
Min	1.4444	1.4361	1.4350	1.4373	1.4049	1.4274	1.4374	1.4097	1.4264
Max	2.2999	1.5327	1.5305	2.1625	1.5022	1.5206	1.9040	1.4974	1.5175
WI									
50	0.9527	0.9543	0.9541	0.9544	0.9564	0.9543	0.9499	0.9560	0.9545
45	0.8508	0.9547	0.9546	0.9533	0.9568	0.9551	0.9549	0.9564	0.9551
40	0.9549	0.9550	0.9548	0.9556	**0.9573**	0.9555	0.9366	0.9571	0.9555
35	0.9546	0.9547	0.9545	0.8763	0.9569	0.9552	0.9554	0.9571	0.9551
30	0.9473	0.9537	0.9536	0.9519	0.9561	0.9543	0.9094	0.9559	0.9546
25	0.9522	0.9522	0.9521	0.8812	0.9539	0.9527	0.9485	0.9536	0.9528
20	0.9502	0.9505	0.9505	0.9516	0.9527	0.9513	0.9355	0.9535	0.9514
15	0.9404	0.9520	0.9520	0.9530	0.9543	0.9525	0.9534	0.9546	0.9524
10	0.9461	0.9461	0.9462	0.9474	0.9483	0.9468	0.9461	0.9484	0.9471
Min	0.8508	0.9461	0.9462	0.8763	0.9483	0.9468	0.9094	0.9484	0.9471
Max	0.9549	0.9550	0.9548	0.9556	0.9573	0.9555	0.9554	0.9571	0.9555
NSE									
50	0.9340	0.9338	0.9330	0.9343	0.9344	0.9325	0.9326	0.9338	0.9325
45	0.8769	0.9338	0.9330	0.9344	0.9342	0.9327	0.9327	0.9340	0.9326
40	0.9343	0.9346	0.9340	0.9346	**0.9371**	0.9338	0.9248	0.9349	0.9336
35	0.9352	0.9354	0.9347	0.8929	0.9362	0.9346	0.9355	0.9360	0.9347
30	0.9345	0.9360	0.9354	0.9366	0.9367	0.9354	0.9115	0.9368	0.9355
25	0.9335	0.9335	0.9330	0.8957	0.9342	0.9330	0.9339	0.9337	0.9332
20	0.9310	0.9313	0.9310	0.9322	0.9325	0.9310	0.9261	0.9327	0.9312
15	0.9294	0.9327	0.9323	0.9333	0.9337	0.9324	0.9334	0.9340	0.9327
10	0.9282	0.9281	0.9276	0.9288	0.9289	0.9278	0.9282	0.9291	0.9280
Min	0.8769	0.9281	0.9276	0.8929	0.9289	0.9278	0.9115	0.9291	0.928
Max	0.9352	0.936	0.9354	0.9366	0.9371	0.9354	0.9355	0.9368	0.9355

Bold values indicate best MAE value.

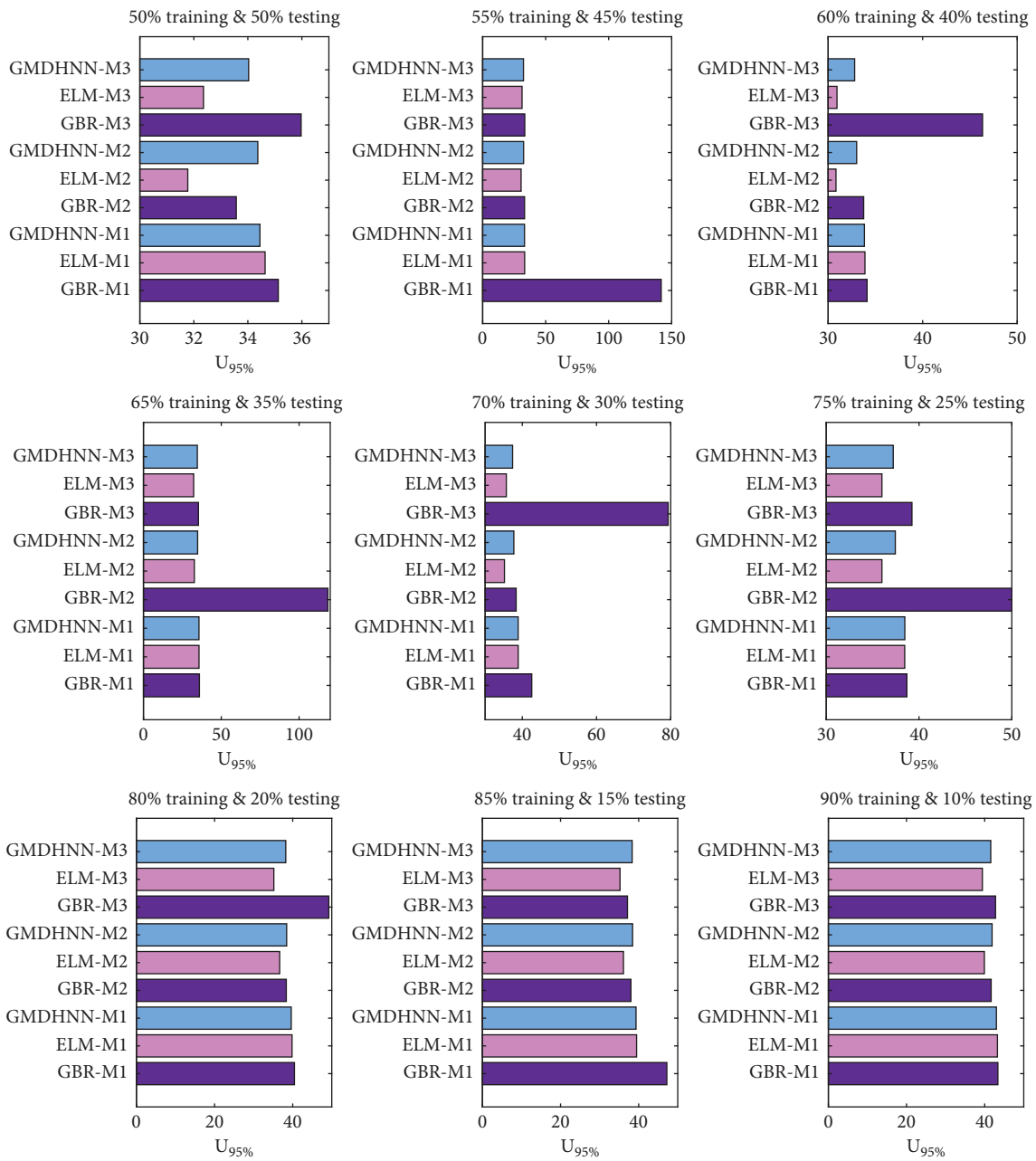


Figure 9: Uncertainty at 95% for different splits and input lags during the testing phase.

Table 5: The reliability analysis (RA%) results.

Train, test	Model	RA%	
		Training	Testing
60%, 40%	ELM	77.34	75.78
	GBR	76.21	71.19
	GMDHNN	58.123	55.16
55%, 45%	ELM	73.74	74.21
	GBR	77.47	69.76
	GMDHNN	59.123	52.98

$$\mathrm{RE}_i = \frac{\mathrm{PM}_{2.5,\mathrm{Obseved},i} - \mathrm{PM}_{2.5,\mathrm{Predicted},i}}{\mathrm{PM}_{2.5,\mathrm{Obseved},i}}. \quad (17)$$

According to value RE_i, if the RE_i fall within the allowable range, the S_i is given 1; otherwise, it is given 0. The allowable range is $\pm$ 20% (0.2)

Based on a specialized technique known as RA, this study evaluated the prediction accuracy of the applied models. Table 5 shows the results of these metrics for every model throughout the training and testing stages. In this work, we select two data division combinations. The first combination involves 60% of the data points used for training and 40% used for testing; however, in the second combination, 55% of the data records are regarded as training data, and the rest are used for validation (testing). According to the obtained result, the ELM generally has a superior performance to other models, achieving the

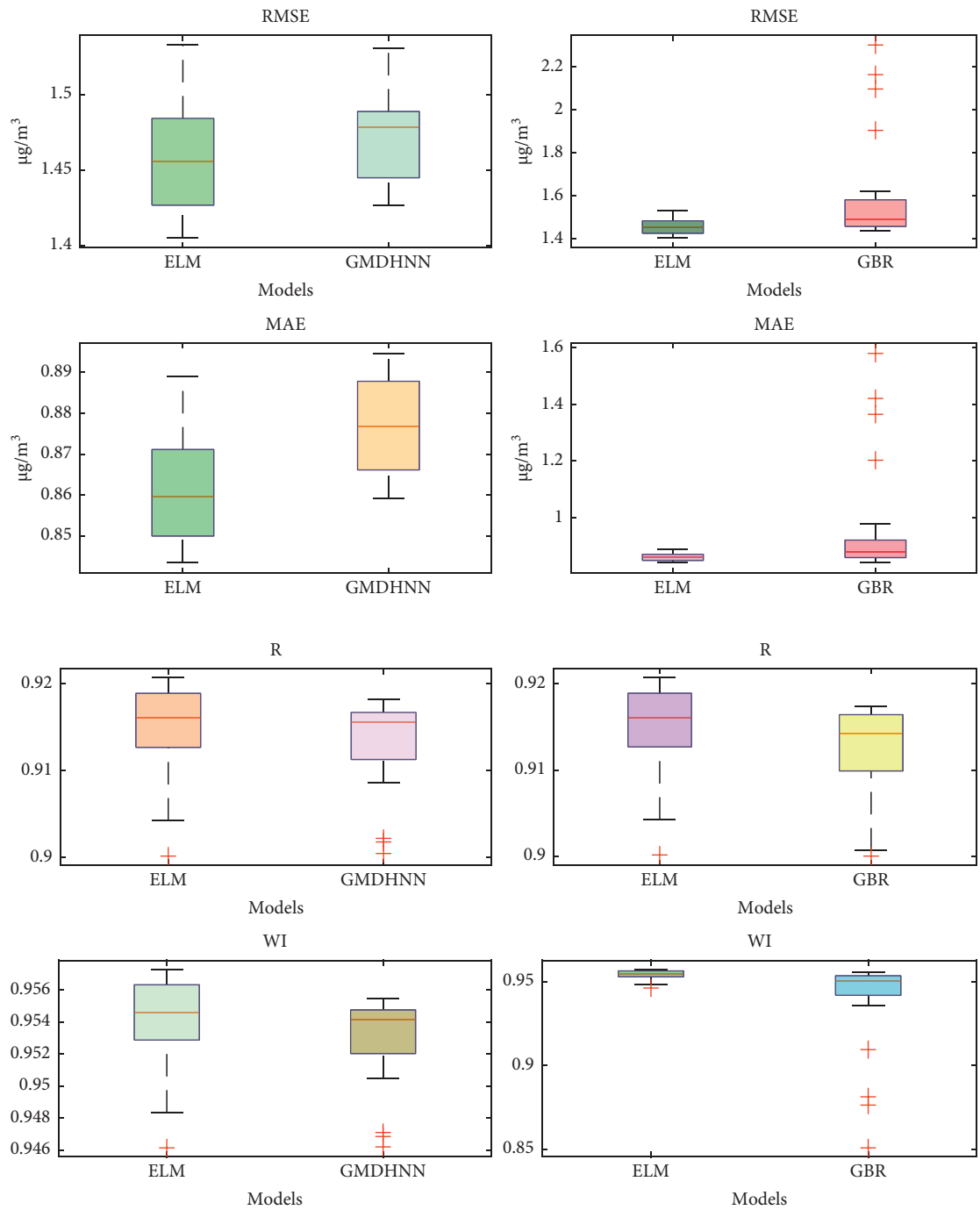


Figure 10: Comparing the ELM models' general performance over GMDHNN and GBR over the testing phase.

highest RA value in the training and testing phase. For example, the ELM in the first combination obtains the higher RA value of 77.34%, followed by GBRT of 76.21%, and finally, the GMDHNN produces the lowest RA with 58.123%. As a result, both ELM and GBR models show a satisfactory performance during the training stage than GMDHNN models. Concerning the testing stage, the results confirm that the ELM is the best model for estimating $PM_{2.5}$, having the highest RA value of 75.78%, followed by GBR (71.19%) and GMDHNN (55.16%). The RA results proved that the ELM is more efficient in estimating the hourly $PM_{2.5}$ than other models. Besides, the appraisal of the models with the help of RA also revealed that the best combination is when the training data records make up 60% of the dataset.

The proposed models are also evaluated graphically using the box plot, violin diagram, and Taylor diagram (see Figures 10–12). According to Figure 10, the overall performance of ELM shows more precise estimates of $PM_{2.5}$ than the other models. Furthermore, median and interquartile range (IQR) errors are reduced more than in the GMDHNN and GBR models.

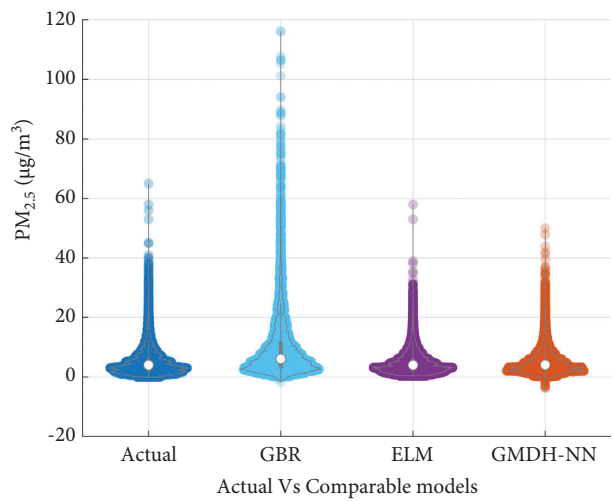


Figure 11: Violin presentation shows the similarity between the actual and predicted $PM_{2.5}$ for the proposed models during the testing phase.

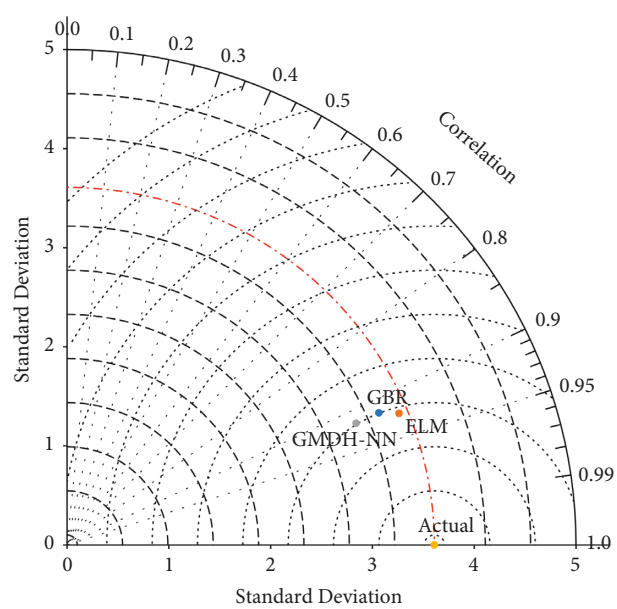


Figure 12: Taylor diagram shows the performance of the proposed models during the testing phase.

Figure 11 represents the violin diagram that integrates a boxplot and density plot to illustrate the data distribution shape. This figure is created for more visualization comparison using testing dataset for the best input combination (training 60% and testing 40%). According to the violin diagram, the ELM model can efficiently mimic the actual data distribution and provide more agreement between the actual and the predicted records. Although the GMDH – NN model outcomes are similar to the actual data distribution, it generates negative values that affect the model's performance. On the other hand, the GBR model performs poorly in mimicking the actual data distribution and generates negative values. Figure 12 represents the Taylor diagram, a polar plot created based on correlation coefficient, standard deviation, and root mean square error to evaluate the models' performance. According to Figure 12, the ELM model can simulate $PM_{2.5}$ closer to the actual one compared to the other models.

Overall, the results obtained in this study prove that the ELM model is more reliable in estimating the hourly $PM_{2.5}$ as well as more flexible in resisting the changes in the data features and length of training data.

Figure 13 gives valuable insight into the practical implementations of the proposed approaches. In particular, data such as $PM_{2.5}$, PM_{10}, CO, O_3, SO_2, and NO_2, as well as temperature, humidity, and air pressure, are extracted from

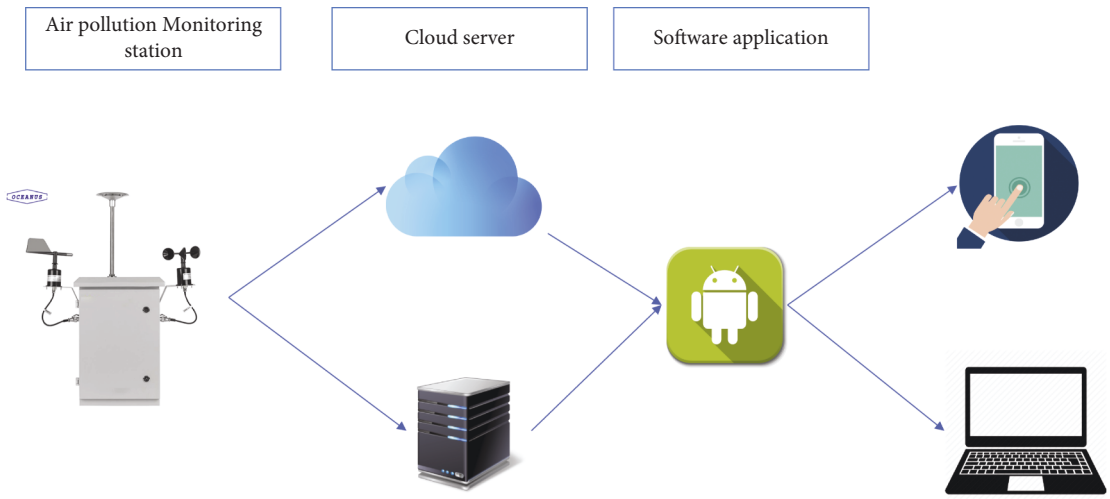


Figure 13: Steps for the practical implementation of AI models in $PM_{2.5}$ prediction.

Figure 14: The interface of the CYAM application [77].

air pollution stations. Next, data are processed at cloud servers using machine learning approaches. Finally, the prediction results can be accessed through a software application interface in real time. A complete implementation example is shown in Figure 14, which shows City Air Quality Management (CYAM) application from Siemens [77]. CYAM is an AI-based worldwide application that can be used on multi-platforms and combines the latest air pollution measurements, such as $PM_{2.5}$, PM_{10}, and NO_2, with the latest AI approach in order to predict the concentrations of $PM_{2.5}$ for the coming days. The CYAM application can predict the pollution for three days ahead with 90% of accuracy and 80% for five days ahead.

4. Conclusion

Three AI models, namely ELM, GMDHNN, and GBR, have been used to predict the hourly $PM_{2.5}$ concentrations over Dorset station, located in Canada. The case study covers the period from 2001 to 2020. The accurate estimations of hourly air pollutants via AI models require a proper input data feature and enough data records for model training. In this study, three input combinations are used via partial autocorrelation function (PACF), and nine data length scenarios are used to validate the models to select the best model that can efficiently adapt to the changes. The finding of this study can be illustrated as follows:

(i) The ELM model generally performs better in estimating $PM_{2.5}$ than the comparable models producing fewer errors (MAE ≈ 0.9710 to 1.099; RMSE ≈ 1.6088 to 1.8329).

(ii) The flexibility of the ELM model in dealing with changes in the size of training data and different training conditions is significantly remarkable. The results showed that the ELM model demands fewer input vectors when the testing data size ranges from 50% to 25% of entire data observations. However, the model requires additional input features in different cases, primarily when the training data represent 80% to 90%.

(iii) All the used models except the ELM do not provide a rational pattern consistent with the changes that occur in the training process.

(iv) The results of this study uncover that the optimal training data, which can provide more accurate estimates, represent 60% of the obtained records.

This study recommends to

(i) Applying the proposed methodology to find the optimal training and testing ratios for other series of pollution like ozone, nitrogen dioxide, sulfur dioxide, and carbon monoxide

(ii) Using the feature selection method instead of PACF and ACF to select the best inputs

(iii) Investigating using deep learning models (i.e., LSTM) to predict the $PM_{2.5}$ concentrations

Abbreviations

XGBoost: Extreme gradient boosting
RPE: Relative prediction error
SVM: Support vector machine
ANN: Artificial neural network
R: Correlation coefficient
R^2: Coefficient of determination
MSE: Mean square error
MAE: Mean absolute error
RMSE: Root mean square error
IA: Index of agreement
MAPE: Mean absolute percentage error
RMSPE: Root mean square prediction error
MPE: Mean prediction error
EEMD- Empirical mode decomposition–
GRNN: General regression neural network
ANFIS: Adaptive Neuro – Fuzzy inference system
MLR: Multiple linear regression
GTWR: Geographically and temporally weighted regression
LR: Linear regression
GRNN: General regression neural network
RBF: Radial basis function
SVR: Support vector regression
PCR: Principal component regression
ARIMA: Autoregressive integrated moving average
NELRM: Nonlinear exposure – lag – response model
RF: Random forest.

Data Availability

The data are available upon request to the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors thank Al-Maarif University College for funding this research.

References

[1] M. Zamani Joharestani, C. Cao, X. Ni, B. Bashir, and S. Talebiesfandarani, "$PM_{2.5}$ prediction based on random forest, XGBoost, and deep learning using multisource remote sensing data," *Atmosphere*, vol. 10, no. 7, p. 373, 2019.

[2] M. Ghoneim and S. M. Hamed, "Towards a smart sustainable city: air pollution detection and control using internet of things," in *2019 5th International Conference on Optimization and Applications (ICOA)*, pp. 1–6, Kenitra, Morocco, 2019.

[3] M. M. Aljumaily, M. A. Alsaadi, N. A. Binti Hashim et al., "Superhydrophobic nanocarbon-based membrane with antibacterial characteristics," *Biotechnology Progress*, vol. 36, no. 3, p. e2963, 2020.

[4] J. I. R. Molano, L. M. O. Bobadilla, and M. P. R. Nieto, "Of cities traditional to smart cities," in *2018 13th Iberian Conference on Information Systems and Technologies (CISTI)*, pp. 1–6, Caceres, Spain, 2018.

[5] W. Sun and J. Sun, "Daily PM2.5 concentration prediction based on principal component analysis and LSSVM optimized by cuckoo search algorithm," *Journal of Environmental Management*, vol. 188, pp. 144–152, 2017.

[6] K. Gan, S. Sun, S. Wang, and Y. Wei, "A secondary-decomposition-ensemble learning paradigm for forecasting PM2.5 concentration," *Atmospheric Pollution Research*, vol. 9, no. 6, pp. 989–999, 2018.

[7] J. Du, F. Qiao, and L. Yu, "Temporal characteristics and forecasting of PM2.5 concentration based on historical data in Houston, USA," *Resources, Conservation and Recycling*, vol. 147, pp. 145–156, 2019.

[8] T. Washington, *HM and EU of the Cost of Air Pollution: Strengthening the Economic Case for Action*, Evaluation University of Washington, Seattle, Washington DC, USA, 2016.

[9] Y. Qi, Q. Li, H. Karimian, and D. Liu, "A hybrid model for spatiotemporal forecasting of PM2.5 based on graph convolutional neural network and long short-term memory," *Science of the Total Environment*, vol. 664, pp. 1–10, 2019.

[10] A. H. Al Hanai, D. S. Antkiewicz, J. D. Hemming et al., "Seasonal variations in the oxidative stress and inflammatory potential of PM2.5 in Tehran using an alveolar macrophage model; the role of chemical composition and sources," *Environment International*, vol. 123, pp. 417–427, 2019.

[11] J. Evans, A. van Donkelaar, R. V. Martin et al., "Estimates of global mortality attributable to particulate air pollution using satellite imagery," *Environmental Research*, vol. 120, pp. 33–42, 2013.

[12] D. Rojas-Rueda, A. de Nazelle, O. Teixidó, and M. J. Nieuwenhuijsen, "Health impact assessment of increasing public transport and cycling use in Barcelona: a morbidity and burden of disease approach," *Preventive Medicine*, vol. 57, no. 5, pp. 573–579, 2013.

[13] J. Lelieveld, J. S. Evans, M. Fnais, D. Giannadaki, and A. Pozzer, "The contribution of outdoor air pollution sources to premature mortality on a global scale," *Nature*, vol. 525, no. 7569, pp. 367–371, 2015.

[14] Y. Sathe, S. Kulkarni, P. Gupta, A. Kaginalkar, S. Islam, and P. Gargava, "Application of moderate resolution imaging spectroradiometer (MODIS) aerosol optical depth (AOD) and weather research forecasting (WRF) model meteorological data for assessment of fine particulate matter (PM2.5) over India," *Atmospheric Pollution Research*, vol. 10, no. 2, pp. 418–434, 2019.

[15] G. Zhou, J. Xu, Y. Xie et al., "Numerical air quality forecasting over eastern China: an operational application of WRF-Chem," *Atmospheric Environment*, vol. 153, pp. 94–108, 2017.

[16] P. Perez and E. Gramsch, "Forecasting hourly PM2.5 in Santiago de Chile with emphasis on night episodes," *Atmospheric Environment*, vol. 124, pp. 22–27, 2016.

[17] B. Lv, W. G. Cobourn, and Y. Bai, "Development of nonlinear empirical models to forecast daily PM2.5 and ozone levels in three large Chinese cities," *Atmospheric Environment*, vol. 147, pp. 209–223, 2016.

[18] B. Lyu, Y. Zhang, and Y. Hu, "Improving PM2.5 air quality model forecasts in China using a bias-correction framework," *Atmosphere*, vol. 8, no. 12, p. 147, 2017.

[19] L. Zhang, J. Lin, R. Qiu et al., "Trend analysis and forecast of PM2.5 in Fuzhou, China using the ARIMA model," *Ecological Indicators*, vol. 95, pp. 702–710, 2018.

[20] S. Mahajan, L.-J. Chen, and T.-C. Tsai, "Short-term PM2.5 forecasting using exponential smoothing method: a comparative analysis," *Sensors*, vol. 18, no. 10, p. 3223, 2018.

[21] A. B. Chelani, "Estimating PM2.5 concentration from satellite derived aerosol optical depth and meteorological variables using a combination model," *Atmospheric Pollution Research*, vol. 10, no. 3, pp. 847–857, 2019.

[22] H. J. Fernando, M. Mammarella, G. Grandoni et al., "Forecasting PM10 in metropolitan areas: efficacy of neural networks," *Environmental Pollution*, vol. 163, pp. 62–67, 2012.

[23] W. Qiao, W. Tian, Y. Tian, Q. Yang, Y. Wang, and J. Zhang, "The forecasting of PM2.5 using a hybrid model based on wavelet transform and an improved deep learning algorithm," *IEEE Access*, vol. 7, pp. 142814–142825, 2019.

[24] H. Liu and C. Chen, "Prediction of outdoor PM2.5 concentrations based on a three-stage hybrid neural network model," *Atmospheric Pollution Research*, vol. 11, no. 3, pp. 469–481, 2020.

[25] T. Li, H. Shen, Q. Yuan, and L. Zhang, "Deep learning for ground-level $PM_{2.5}$ prediction from satellite remote sensing data," in *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, pp. 7581–7584, Valencia, Spain, 2018.

[26] Y. Wang, "Regional-level prediction model with advection PDE model and fine particulate matter (PM 2.5) concentration data," *Physica Scripta*, vol. 95, no. 3, Article ID 035204, 2020.

[27] M. M. Hameed, F. Khaleel, M. A. Abed, D. Khaleel, and M. K. Alomar, "An effective predictive model for daily evapotranspiration based on a limited number of meteorological parameters," in *2021 3rd International Sustainability and Resilience Conference: Climate Change*, pp. 495–499, Sakheer, Bahrain, 2021.

[28] M. M. Hameed, M. K. AlOmar, F. Khaleel, and N. Al-Ansari, "An extra tree regression model for discharge coefficient prediction: novel, practical applications in the hydraulic sector and future research directions," *Mathematical Problems in Engineering*, vol. 2021, Article ID 7001710, 19 pages, 2021.

[29] M. M. Hameed, F. Khaleel, and D. Khaleel, "Employing a robust data-driven model to assess the environmental damages caused by installing grouted columns," in *2021 Third International Sustainability and Resilience Conference: Climate Change*, pp. 305–309, Sakheer, Bahrain, 2021.

[30] A. Dikshit, B. Pradhan, and M. Santosh, "Artificial neural networks in drought prediction in the 21st century–A scientometric analysis," *Applied Soft Computing*, vol. 114, Article ID 108080, 2022.

[31] G. Hinton, "Deep belief networks," *Scholarpedia*, vol. 4, no. 5, p. 5947, 2009.

[32] F. Khaleel, M. M. Hameed, D. Khaleel, and M. K. AlOmar, "Applying an efficient AI approach for the prediction of bearing capacity of shallow foundations," in *International Conference on Emerging Technology Trends in Internet of Things and Computing*, pp. 310–323, Berlin, Germany, 2022.

[33] P. Du, J. Wang, W. Yang, and T. Niu, "A novel hybrid fine particulate matter (PM2.5) forecasting and its further application system: case studies in China," *Journal of Forecasting*, vol. 41, no. 1, pp. 64–85, 2022.

[34] A. Stojić, G. Jovanovic, S. Stanisic et al., "The PM2.5-bound polycyclic aromatic hydrocarbon behavior in indoor and outdoor environments, part II: explainable prediction of benzo [a] pyrene levels," *Chemosphere*, vol. 289, Article ID 133154, 2022.

[35] P. Liu, E. Yao, T. Liu, L. Kong, X. Tang, and G. Tan, "Improvement of AI forecast of gridded PM2.5 forecast in China through ConvLSTM and Attention," *CCF Trans. High Perform. Comput.*vol. 4, pp. 104–119, 2022.

[36] P. Zhang, L. Yang, W. Ma, N. Wang, F. Wen, and Q. Liu, "Spatiotemporal estimation of the PM2.5 concentration and human health risks combining the three-dimensional landscape pattern index and machine learning methods to optimize land use regression modeling in Shaanxi, China," *Environmental Research*, vol. 208, Article ID 112759, 2022.

[37] Q. Xiao, G. Geng, J. Cheng et al., "Evaluation of gap-filling approaches in satellite-based daily PM2.5 prediction models," *Atmospheric Environment*, vol. 244, Article ID 117921, 2021.

[38] I. Yeo, Y. Choi, Y. Lops, and A. Sayeed, "Efficient PM2.5 forecasting using geographical correlation based on integrated deep learning algorithms," *Neural Computing & Applications*, vol. 33, no. 22, pp. 15073–15089, 2021.

[39] M. H. Nguyen, P. Le Nguyen, K. Nguyen, V. A. Le, T. H. Nguyen, and Y. Ji, "PM25 prediction using genetic algorithm-based feature selection and encoder-decoder model," *IEEE Access*, vol. 9, pp. 57338–57350, 2021.

[40] S. Chae, J. Shin, S. Kwon, S. Lee, S. Kang, and D. Lee, "PM_{10} and PM_{25} real-time prediction models using an interpolated convolutional neural network," *Scientific Reports*, vol. 11, no. 1, pp. 11952–11959, 2021.

[41] E. Kristiani, H. Lin, J.-R. Lin, Y.-H. Chuang, C.-Y. Huang, and C.-T. Yang, "Short-term prediction of PM2.5 using LSTM deep learning methods," *Sustainability*, vol. 14, no. 4, p. 2068, 2022.

[42] J. Ni, Y. Chen, Y. Gu, X. Fang, and P. Shi, "An improved hybrid transfer learning-based deep learning model for PM2.5

concentration prediction," *Applied Sciences*, vol. 12, no. 7, p. 3597, 2022.

[43] K. Huang, Q. Xiao, X. Meng et al., "Predicting monthly high-resolution PM2.5 concentrations with random forest model in the North China Plain," *Environmental Pollution*, vol. 242, pp. 675–683, 2018.

[44] A. Masood and K. Ahmad, "A review on emerging artificial intelligence (AI) techniques for air pollution forecasting: fundamentals, application and performance," *Journal of Cleaner Production*, vol. 322, Article ID 129072, 2021.

[45] Z. Shang and J. He, "Predicting hourly mathbf {PM}-{2.5} concentrations based on random forest and ensemble neural network," in *Proceedings 2018 Chinese Automation Congress, CA 2018*, pp. 2341–2345, 2019.

[46] W. Yuchi, E. Gombojav, B. Boldbaatar et al., "Evaluation of random forest regression and multiple linear regression for predicting indoor fine particulate matter concentrations in a highly polluted city," *Environmental Pollution*, vol. 245, pp. 746–753, 2019.

[47] J. Murillo-Escobar, J. P. Sepulveda-Suescun, M. A. Correa, and D. Orrego-Metaute, "Forecasting concentrations of air pollutants using support vector regression improved with particle swarm optimization: case study in Aburrá Valley, Colombia," *Urban Climate*, vol. 29, Article ID 100473, 2019.

[48] M. Najafzadeh and F. Saberi-Movahed, "GMDH-GEP to predict free span expansion rates below pipelines under waves," *Marine Georesources & Geotechnology*, vol. 37, no. 3, pp. 375–392, Mar. 2019.

[49] M. Najafzadeh and H. M. Azamathulla, "Neuro-fuzzy GMDH to predict the scour pile groups due to waves," *Journal of Computing in Civil Engineering*, vol. 29, no. 5, Article ID 4014068, 2015.

[50] M. M. Hameed, M. A. Abed, N. Al-Ansari, and M. K. Alomar, "Predicting compressive strength of concrete containing industrial waste materials: novel and hybrid machine learning model," *Advances in Civil Engineering*, vol. 2022, Article ID 5586737, 19 pages, 2022.

[51] S. Moisan, R. Herrera, and A. Clements, "A dynamic multiple equation approach for forecasting PM2.5pollution in Santiago, Chile," *International Journal of Forecasting*, vol. 34, no. 4, pp. 566–581, 2018.

[52] A. Masood and K. Ahmad, "A model for particulate matter (PM2.5) prediction for Delhi based on machine learning approaches," *Procedia Computer Science*, vol. 167, pp. 2101–2110, 2020.

[53] J. Amanollahi and S. Ausati, "PM2.5 concentration forecasting using ANFIS, EEMD-GRNN, MLP, and MLR models: a case study of Tehran, Iran," *Air Quality, Atmosphere & Health*, vol. 13, no. 2, pp. 161–171, 2020.

[54] M. Mirzaei, J. Amanollahi, and C. G. Tzanis, "Evaluation of linear, nonlinear, and hybrid models for predicting PM2.5 based on a GTWR model and MODIS AOD data," *Air Quality, Atmosphere & Health*, vol. 12, no. 10, pp. 1215–1224, 2019.

[55] S. S. Ganesh, P. Arulmozhivarman, and V. S. N. R. Tatavarti, "Prediction of PM2.5 using an ensemble of artificial neural networks and regression models," *Journal of Ambient Intelligence and Humanized Computing*, 2018.

[56] Q. Zhou, H. Jiang, J. Wang, and J. Zhou, "A hybrid model for PM2.5 forecasting based on ensemble empirical mode decomposition and a general regression neural network," *Science of the Total Environment*, vol. 496, pp. 264–274, 2014.

[57] X. Mao, T. Shen, and X. Feng, "Prediction of hourly ground-level PM2.5 concentrations 3 days in advance using neural networks with satellite data in eastern China," *Atmospheric Pollution Research*, vol. 8, no. 6, pp. 1005–1015, 2017.

[58] Z.-Y. Chen, T. H. Zhang, R. Zhang et al., "Extreme gradient boosting model to estimate PM2.5 concentrations with missing-filled satellite data in China," *Atmospheric Environment*, vol. 202, pp. 180–189, 2019.

[59] X. Hu, J. H. Belle, X. Meng et al., "Estimating PM2.5 concentrations in the conterminous United States using the random forest approach," *Environmental Science and Technology*, vol. 51, no. 12, pp. 6936–6944, 2017.

[60] Z. A. Al Sudani and G. S. A. Salem, "Evaporation rate prediction using advanced machine learning models: a comparative study," *Advances in Meteorology*, vol. 2022, Article ID 1433835, 13 pages, 2022.

[61] M. M. Hameed, M. K. AlOmar, W. J. Baniya, and M. A. AlSaadi, "Prediction of high-strength concrete: high-order response surface methodology modeling approach," *Engineering with Computers*, vol. 38, no. S2, pp. 1655–1668, 2022.

[62] N. Nabipour, S. N. Qasem, E. Salwana, and A. Baghban, "Evolving LSSVM and ELM models to predict solubility of non-hydrocarbon gases in aqueous electrolyte systems," *Measurement*, vol. 164, Article ID 107999, 2020.

[63] "Health impacts of air pollution in Canada," 2021, https://www.canada.ca/en/health-canada/services/publications/healthy-living/2021-health-effects-indoor-air-pollution.html.

[64] M. K. AlOmar, M. M. Hameed, and M. A. AlSaadi, "Multi hours ahead prediction of surface ozone gas concentration: robust artificial intelligence approach," *Atmospheric Pollution Research*, vol. 11, no. 9, pp. 1572–1587, 2020.

[65] G. Bin Huang, Q. Y. Zhu, and C. K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, 2006.

[66] D. Wang and G.-B. Huang, "Protein sequence classification using extreme learning machine," in *Proceedings. 2005 IEEE International Joint Conference on Neural Networks*, pp. 1406–1411, Montreal, Que, 2005.

[67] G.-B. Huang, Q.-Y. Zhu, and C. K. Siew, "Real-time learning capability of neural networks," *IEEE Transactions on Neural Networks*, vol. 17, no. 4, pp. 863–878, 2006.

[68] A. G. Ivakhnenko, "Heuristic self-organization in problems of engineering cybernetics," *Automatica*, vol. 6, no. 2, pp. 207–219, 1970.

[69] J. H. Friedman, "Greedy function approximation: a gradient boosting machine 1 function estimation 2 numerical optimization in function space," *North*, vol. 1, no. 3, pp. 1–10, 1999.

[70] J. H. Friedman, "Stochastic gradient boosting," *Computational Statistics & Data Analysis*, vol. 38, no. 4, pp. 367–378, 2002.

[71] M. K. Alomar, M. M. Hameed, N. Al-Ansari, and M. A. Alsaadi, "Data-Driven model for the prediction of total dissolved gas: robust artificial intelligence approach," *Advances in Civil Engineering*, vol. 2020, Article ID 6618842, 20 pages, 2020.

[72] M. Despotovic, V. Nedic, D. Despotovic, and S. Cvetanovic, "Review and statistical analysis of different global solar radiation sunshine models," *Renewable and Sustainable Energy Reviews*, vol. 52, pp. 1869–1880, 2015.

[73] A. Yafouz, A. N. Ahmed, N. Zaini, M. Sherif, A. Sefelnasr, and A. El-Shafie, "Hybrid deep learning model for ozone

concentration prediction: comprehensive evaluation and comparison with various machine and deep learning algorithms," *Engineering Applications of Computational Fluid Mechanics*, vol. 15, no. 1, pp. 902–933, 2021.

[74] M. M. Hameed, M. K. AlOmar, S. F. Mohd Razali et al., "Application of artificial intelligence models for evapotranspiration prediction along the southern coast of Turkey," *Complexity*, vol. 2021, Article ID 8850243, 20 pages, 2021.

[75] M. M. Hameed, F. Khaleel, M. K. AlOmar, S. F. Mohd Razali, and M. A. AlSaadi, "Optimising the selection of input variables to increase the predicting accuracy of shear strength for deep beams," *Complexity*, vol. 2022, Article ID 6532763, 23 pages, 2022.

[76] F. Saberi-Movahed, M. Najafzadeh, and A. Mehrpooya, "Receiving more accurate predictions for longitudinal dispersion coefficients in water pipelines: training group method of data handling using extreme learning machine conceptions," *Water Resources Management*, vol. 34, no. 2, pp. 529–561, 2020.

[77] "Artificial intelligence improves air quality," 2019, https://new.siemens.com/global/en/company/stories/infrastructure/2019/artificial-intelligence-improves-air-quality.html.

Hindawi
Advances in Meteorology
Volume 2022, Article ID 4873393, 10 pages
https://doi.org/10.1155/2022/4873393

Research Article

Blue-Green Space Changes of Baiyangdian Wetland in Xiong'an New Area, China

Chunlei Zhao,[1,2] **Shuan Qian**,[3] **Chengzhen Meng**,[1,2] **Yufei Chang**,[1,2] **Wenzhe Guo**,[4,5] **Sha Wang**,[2,6] **and Yinglong Sun**[3]

[1]*Hebei Provincial Institute of Meteorological Sciences, Shijiazhuang 050021, China*
[2]*Key Laboratory of Meteorology and Ecological Environment of Hebei Province, Shijiazhuang 050021, China*
[3]*National Meteorological Center, Beijing 100081, China*
[4]*Hengshui Meteorological Bureau, Hengshui 053000, China*
[5]*Raoyang National Climatological Observatory, Hengshui 053900, China*
[6]*Qinhuangdao Meteorological Bureau, Qinhuangdao 066000, China*

Correspondence should be addressed to Shuan Qian; shqian19@163.com

Received 11 May 2022; Revised 24 August 2022; Accepted 10 September 2022; Published 27 September 2022

Academic Editor: Upaka Rathnayake

As a regulator of ecological environment, Baiyangdian Wetland is in a pivotal position in constructing the blue-green space (BGS) of Xiong'an New Area in China. This study aims to reveal the spatiotemporal changes of the BGS in Baiyangdian Wetland from 2016 to 2021. It uses Google Earth Engine (GEE) to calculate NDVI and NDWI based on Sentinel-2 Satellite remote sensing data and extracts the blue-green space by a classification model driven by NDVI and NDWI. Moreover, the land-use transfer matrix and landscape pattern indices are applied for evaluating the BGS changes in the wetland. According to the results, vegetation in the wetland shows no obvious spatial transfer. From 2016 to 2020, the BGS proportion in the wetland showed a stable increase, with the blue space getting larger by 10.8%. The indicators of the Number of Patches (NP), Patch Density (PD), Largest Patch Index (LPI), Contagion, and Landscape Shape Index (LSI) of the wetland decreased, suggesting a better ecological environment since the establishment of Xiong'an New Area in 2017. Based on the results, the author makes the following conclusion: the construction of BGS in Baiyangdian Wetland results in a well-organized ecological environment. The study provides a reference for building Xiong'an New Area and monitoring BGS changes in other regions.

1. Introduction

Blue-green space (BGS) refers to the blue space and green space in an area. Blue space refers to the distribution of both natural and artificial water bodies in the area. Green space refers to areas that are covered by vegetation [1]. The blue-green space serves as a source of oxygen and a carbon sink in the ecological cycle. It regulates local climate and improves local environment. The enhancement of the blue-green space surrounding the urban area could effectively reduce urban heat island effects and improve air quality, benefiting the residential environment [2–6].

Xiong'an New Area, south of Beijing, was established in 2017 as a national new area and part of the millennium plan. This project is of major national significance and is a model and touchstone for high-quality development across China. The success of Xiong'an New Area marks that China is moving toward a cleaner and greener development path. For Xiong'an, "ecological and green development" is a high priority. Baiyangdian is the core water area of Xiong'an New Area, also known as the "kidney of North China". It is the largest natural freshwater lake in the North China Plain. The ecological environment of Baiyangdian is vital to the strategy of building a blue and green ecological city in the Xiong'an new area. To make the proportion of blue-green space stand at 70% of the Xiong'an New Area by 2035, a detailed study needs to be carried out on the Baiyangdian Wetland and the ecological environment changes in Xiong'an New Area.

Wetland landscape patterns and their changes directly affect the biogeochemical cycle, hydrological cycle, ecosystem services, and biodiversity protection. They have a profound impact on the blue-green spatial distribution of the wetland [7–10]. Since the 1970s, many researchers on wetlands have used landscape ecology to study the wetland pattern in different regions. The dynamic change of wetland landscape patterns has become the focus of wetland research [11–14]. Based on this, numerous academics have studied and analyzed the landscape of Baiyangdian Wetland, paying particular attention to its distinctive landscape structure, function, and pattern distribution [15–20]. For example, Yang et al. analyzed the landscapes' spatial and temporal changes in Baiyangdian Wetland and the driving factors. Moreover, they discussed future changes in the landscapes based on relevant planning [15] and their influences. Tang et al. studied the trend of aquatic plant area and biomass changes in Baiyangdian Wetland during different periods [16]. Wu et al. analyzed the dynamic changes in the water level of Baiyangdian Wetland with deep learning. Their study improved the accuracy of water-level monitoring [17]. By extracting the information of Baiyangdian Wetland with Normalized Difference Water Index (NDWI), Lin et al. studied the relationship between wetland area and the water level [18]. Zhao et al. studied the relationship between the precipitation and the lowest water level of Baiyangdian Lake and established a water-level prediction model [19, 20]. Previous studies mainly focused on the impact of the water environment on Baiyangdian Wetland. Their study directions included water quantity, water quality, wetland degradation, and ecosystem services. Nevertheless, Xiong'an New Area is built on blue water and green trees with the lake integrating with the city. The BGS changes in water and vegetation in Baiyangdian Wetland are the most urgent problem to be solved and are important to the development of the Xiong'an New Area.

Generally, research on long-time series changes with satellite remote sensing images requires massive data and complex computing. Data preprocessing is a time-consuming and laborious task. Google Earth Engine (GEE), a cloud computing data analysis platform, provides a solution. This paper has adopted GEE to analyze the Sentinel-2 Satellite remote sensing images of Baiyangdian Wetland. The images are from June to August of 2016~2021 with a 10-meter spatial resolution. The algorithm of NDVI-NDWI decision trees is used to extract information about the Baiyangdian Wetland. The results are analyzed to evaluate changes in BGS and types of Baiyangdian Wetland. In this way, the data on BGS changes in Baiyangdian Wetland before (2016) and after (2017) the establishment of Xiong'an New Area can be obtained. The data will provide technical support for the ecological protection and restoration of Baiyangdian Wetland.

2. Material and Methods

2.1. Overview of the Study Area. Baiyangdian, the largest freshwater lake in northern China, is located in the south branch of Daqing River (Figure 1). Belonging to the Haihe River Basin, it is at the center of Hebei Province and the Beijing-Tianjin-Hebei region. The wetland brings together 143 lakes of different sizes, such as Baiyangdian Lake, Zaozha Lake, Mapeng Lake, and Yaohulu Lake. Baiyangdian is rich in aquatic species such as plankton, benthic animals, and fish [16]. Baiyangdian water area effectively maintains the ecological balance of Xiong'an New Area [21]. Interweaving land with water, the wetland is also home to terrestrial plants, aquatic plants, and animals. This complex ecosystem presents a beautiful scenery of "blue water and green vegetation" in the growing season.

2.2. Data Source. GEE is jointly developed by Google, Carnegie Mellon University, and U.S. Geological Survey. This platform provides scientific analysis of geographic data based on big data and cloud computing. With high-performance computing resources, it is a good tool for processing huge amounts of geospatial data [22–24]. GEE platform contains archived remote sensing data, which is supported by NASA, U.S. Geological Survey, and NOAA, as well as PB data. The geographic information data is of different kinds and has more than 40 years of historical sequence. GEE data is updated in real-time, follows standardized storage forms, and is easy to call [25]. The remote sensing images of the study area are from the Sentinel-2 Satellite and downloaded from GEE. The data has a 10-meter spatial resolution, and its historical sequence is from June to August of 2016~2021. To discard images with many clouds and retain those with less than 20% cloud amount, this paper has developed a cloud-removing algorithm based on the bitmask information of the Sentinel-2 Satellite data. After the pixels with clouds have been eliminated in preprocessed images to the greatest extent, pure ground information is retained. NDVI data and NDWI data are derived from preprocessed images. By calculating, pixel by pixel, the average value of multitemporal data in the image stack, the average value of NDVI data, and NDWI data, both of which are from June to August of 2016~2021, can be synthesized. Finally, NDVI and NDWI results are exported from GEE for the analysis of the BGS pattern and evolution of Baiyangdian Wetland. Normalized Difference Vegetation Index (NDVI), which is a radiation quantification parameter, reflects the abundance and activity of green vegetation. Its physical basis is that vegetation chlorophyll shows strong absorption in the red-light spectrum and high reflectance in the near-infrared spectrum. Using the spectral reflectance of vegetation, the normalized difference between near-infrared and red-light spectrums can be computed in remote sensing images, namely, $NDVI = (NIR - R)/(NIR + R)$. Therefore, this study uses NDVI to inverse the green space in Baiyangdian Wetland. Normalized Differential Water Index (NDWI) is a parameter improved based on the NDVI algorithm. Its physical basis is the reflectance of vegetation and water in near-infrared and visible light spectrums. The normalized difference between near-infrared and visible light spectrums in remote sensing images is calculated based on the equation $NDWI = (Green - NIR)/(Green + NIR)$. This method inhibits the vegetation information in images and is good at distinguishing the phytoplankton on the water surface from the vegetation on the water shore. In this way, enhanced water information can be acquired. Therefore, this study uses NDWI to inverse the blue space in Baiyangdian Wetland.

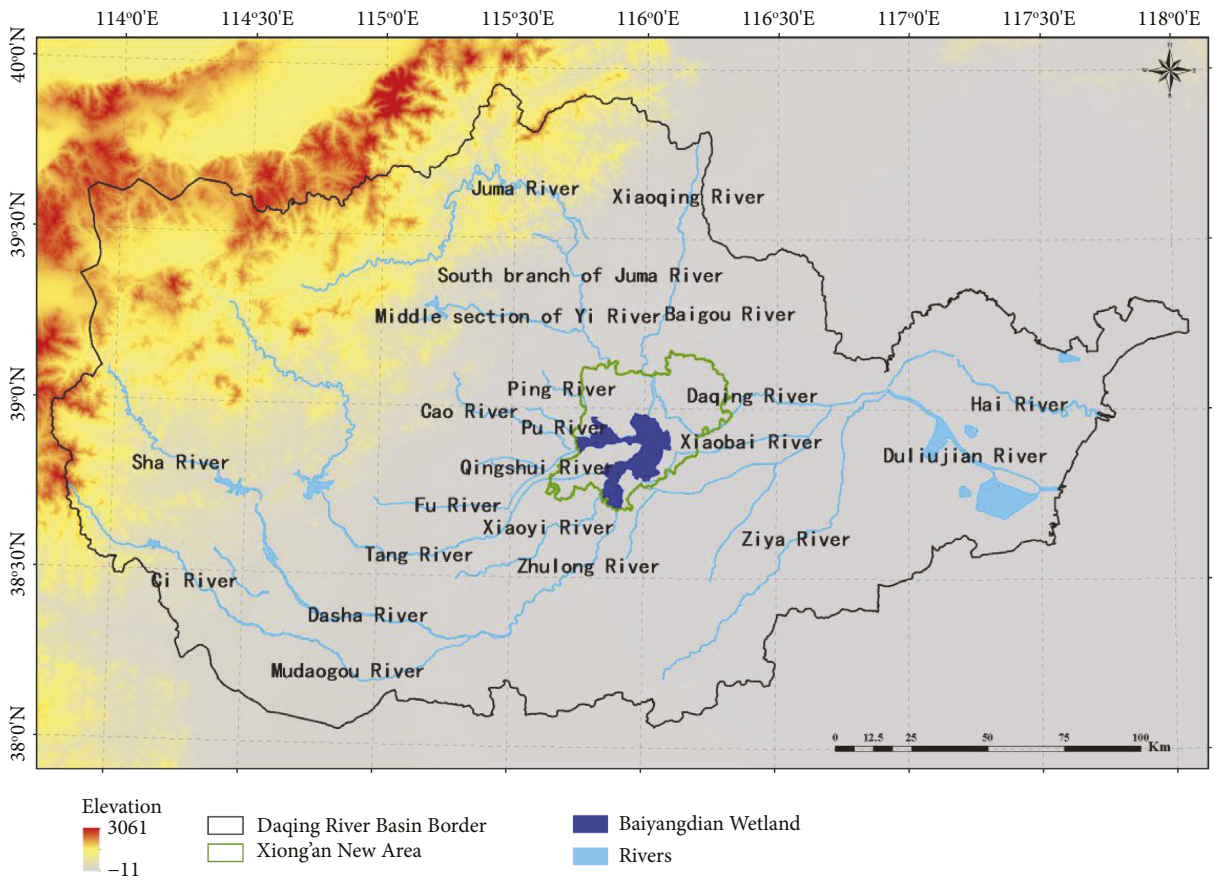


Figure 1: Overview of the study area.

Images of GF Satellite from the China Centre for Resources Satellite Data and Application (CRESDA) have higher resolution than the Sentinel-2 Satellite. GF-2 Satellite includes both 1-meter-resolution panchromatic images and 4-meter-resolution multispectral images. But GF Satellites have the coverage data for Baiyangdian Wetland from June to August, so they were not used in monitoring the BGS changes from 2016 to 2021.1-meter-resolution images of GF-2 Satellite which is "true value" are generated to only validate the remote sensing results from the Sentinel-2 Satellite which can cover all of Baiyangdian Wetland every year and it is used to obtain yearly BGS changes.

2.3. Methods

2.3.1. Wetland Classification Model. Baiyangdian Wetland has various ground objects, with closely related spectrums. For example, ground objects of the same type may show great similarity in satellite images but are different in spectrums [26]. Conventional supervised classification involves a heavy workload and is influenced by human subjectivity. Moreover, this method also needs a highly pure classification template, as well as a typical, representative, and accurate training area. It has limited application scope and is indispensable for artificial sample selection and prior knowledge of the place. In comparison, NDVI and NDWI indices are more sensitive to water and vegetation. SWIR is the short-wave infrared band channel of Sentinel-2 data, which is less scattered by the atmosphere and has a strong penetration ability. It can effectively reflect the reflection of vegetation and other ground objects and efficiently classifies waterbody, vegetation, and urban area. Therefore, as characteristics of water, urban and rural construction space and topographic shadow are easily confused, NDVI and NDWI are suitable to analyze the Baiyangdian Wetland, and the SWIR channel is utilized in this classification model to improve the accuracy of wetland ground object information. Considering the image characteristics of the study area, this paper has selected NDVI, NDWI data, and SWIR channel for analysis.

BGS information extraction model for Baiyangdian Wetland is based on the NDVI-NDWI method (Figure 2). It utilizes the changing pattern of Sentinel-2 data based on information spectrums and adopts the spectral reflectance of ground object information to categorize wetland information. The model aims to extract the results of open water, vegetation as well as urban and rural construction space. The thresholds 1-7 are the identification thresholds of NDVI and NDWI, respectively. SWIR denotes the channel reflectance of the shortwave and infrared spectrum in the Sentinel-2 data.

2.3.2. Validation of Classification Results. There are mainly two types of validation data to evaluate the results of remote

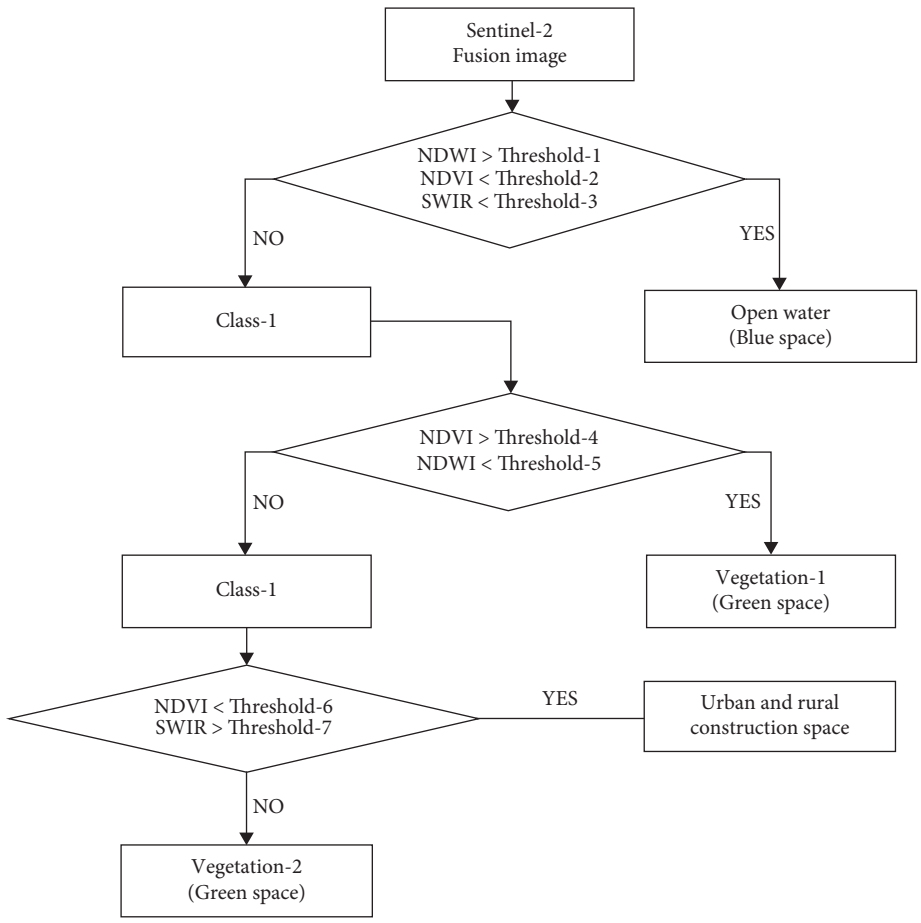


Figure 2: BGS information extraction model for Baiyangdian Wetland.

sensing classification. One is measured sampling data. The other is high-resolution satellite remote sensing data that reflects ground conditions [27]. Field investigation and sampling are difficult for some reasons. First, Baiyangdian Wetland involves a large area and various ground objects. Secondly, the wetland is often influenced by ecological flood dike, estuary wetland construction projects, and upstream floods. Third, the opening hours of Baiyangdian Wetland to the public are short. In comparison, GF-2 data has higher spatial resolution than data of Sentinel-2, GF-2, and Sentinel-2 of the same day, so it can be used during validation as the "true value" of ground. Specifically, supervised classification is used for categorizing the data before the visual interpretation for test. In this way, the classification results are extracted from GF-2 data (Figure 2). Due to the different spatial resolution of GF-2 and Sentinel-2 data, the aggregation resampling method is adopted to resample the GF-2 data into the same grid size as the Sentinel-2 data.

The classification data of 2019 from Sentinel-2 is compared with the data of the same period by GF-2 inversion (Figure 3). According to the results (Table 1), the contour of the classification data between Sentinel-2 data and GF-2 which is the "true value" is closer. The overall distribution of BGS and construction space is relatively accurate. The inversion result is the closest when extracting the open water.

The model built in Figure 2 can precisely extract data of small and scattered water areas. When extracting the transition between water and vegetation, the model's inversion result is close to the true value. In other words, the model can reflect the contour of the transition area.

To evaluate the accuracy, correlation coefficient (R), mean error (ME), and root mean square error (RMSE) are used for comparative analysis. For the statistical analysis of inversion results, 1 is set as the digital number (DN) of blue space, 2 for green space, and 3 for construction space in urban and rural areas. According to Table 1, R between the classification result and the "true value" reaches 0.948, passing the 99% significance test. Mean bias (MB), mean error (ME), and root mean square error (RMSE) are 0.06, 0.203, and 0.144, respectively. Thus, the classification results of the proposed BGS information extraction model are highly accurate.

2.3.3. Landscape Pattern Index. The landscape pattern index is a basic means to study the landscape pattern, usually from the type level scale and the landscape level scale [15]. Based on the understanding of the landscape pattern index, the indicators of patch number (NP), patch density (PD), and the contagion index (CONTAG) were selected in the ecological process to evaluate the BGS and construction space in

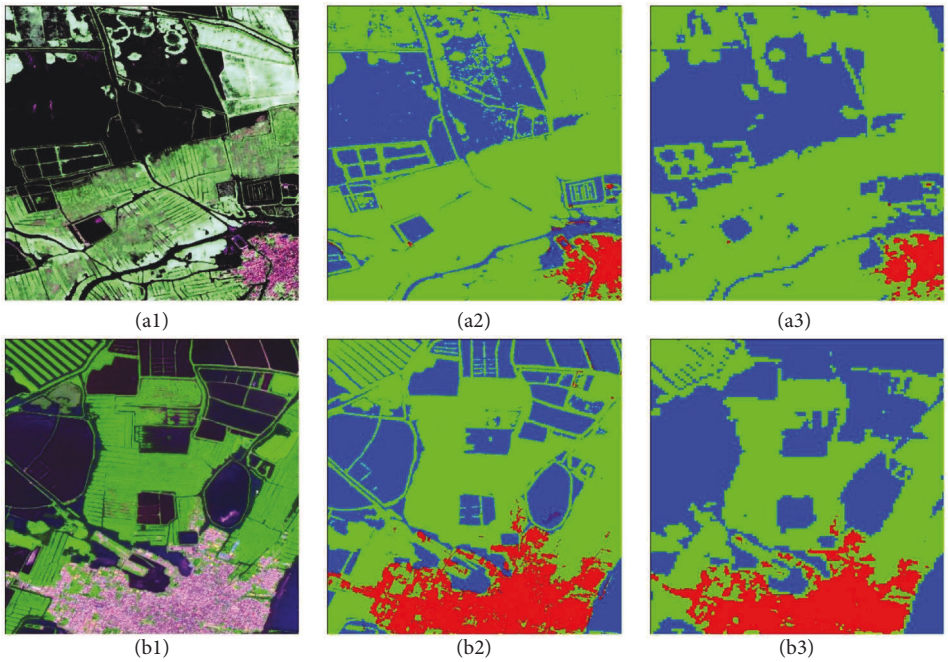


Figure 3: Comparison between Baiyangdian Wetland sample images and classification data of the model built based on GF-2 and Sentinel-2 data (a1 is region 1, a2 is the classification result of the model built based on GF-2 data in region 1, a3 is the classification result of the model built based on Sentinel-2 data in region 1, b1 is region 2, b2 is the classification result of the model built based on GF-2 data in region 2, and b3 is the classification result of the model built based on Sentinel-2 data in region 2).

Table 1: Accuracy test of the classification results of the model.

	R	*P*	MB	ME	RMSE
Test	0.948	0.001	0.060	0.203	0.144

Baiyangdian Wetland. These indices can reflect changes in the aggregation and fragmentation of BGS, and thus assess the ecological function of BGS. Landscape pattern indices can quantitatively describe the changing pattern of the landscape. By analyzing the relationship between landscape pattern and landscape process, we can further study the landscape structure and spatial pattern in the study area and know about the precise dynamic changes of landscape pattern in the Baiyangdian Wetland [28]. During the study of small-scale BGS, such as the BGS of wetlands, the biggest external influential factor for the structure is human activity. Fragmentation and landscape diversity are also important features [26]. Therefore, this study has calculated the spatial heterogeneity, fragmentation degree, and patch complexity of Baiyangdian Wetland landscape. It has also analyzed the dynamic changes of BGS in Baiyangdian Wetland from the horizontal orientation.

3. Results

3.1. BGS Changes of Baiyangdian Wetland. According to the classification results (Figure 4) from Sentinel-2 data, the green space is mainly concentrated in the north, the northwest, and the southwest of the wetland. The overall distribution pattern is as follows: the north has more green space than the south and the west has more green space than the east. The blue space is mainly scattered in the southeast and northeast of the wetland.

As for annual changes in Baiyangdian Wetland, the BGS proportion increased steadily from 2016 to 2020, among which the proportion of blue space increased with some fluctuations. However, in 2021, due to the construction of the ecological function area in the northwest, some green space was transformed into urban and rural construction space, resulting in a small decrease in the BGS proportion (Figure 5).

From 2016 to 2021, due to constant ecological replenishment of water, the proportion of blue space in Baiyangdian Wetland increased by 10.8%, with an area of 64.8 km^2. The proportion of green space decreased by 3.3% (Table 2), without obvious spatial transfer (still concentrated in the northern, northwestern, and southwestern regions).

3.2. Landscape Pattern Changes of Baiyangdian Wetland. In the landscape pattern analysis, higher fragmentation, which means increases in numbers of patches (NP) and patches density (PD) [29], indicates a worse ecological function of the landscape. The landscape shape index (LSI)

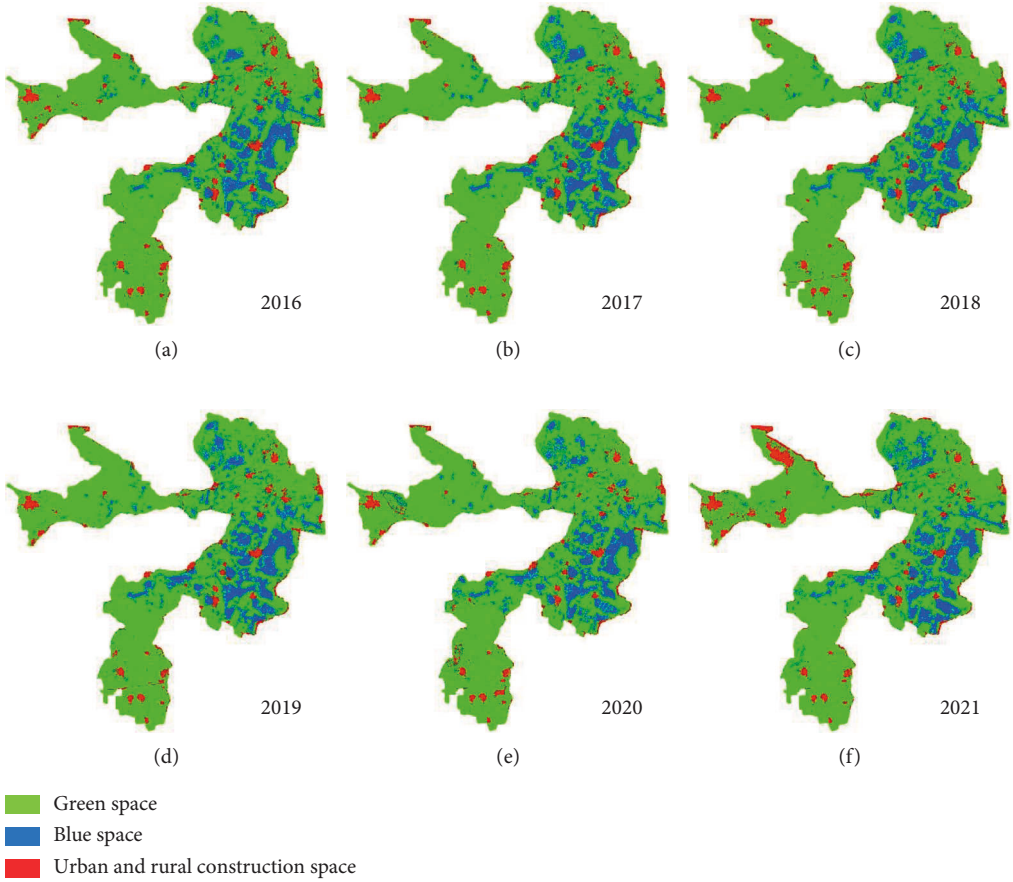


Figure 4: BGS distribution changes of Baiyangdian Wetland from 2016 to 2021. (a) 2016. (b) 2017. (c) 2018. (d) 2019. (e) 2020. (f) 2021.

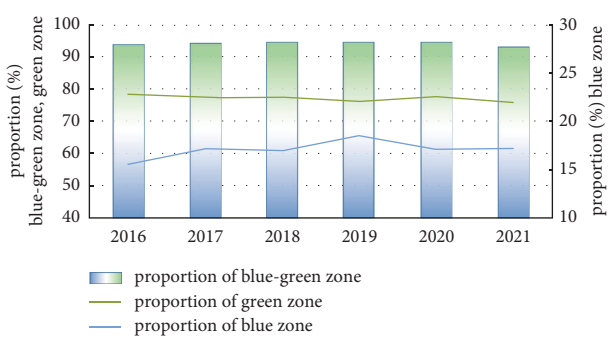


Figure 5: BGS changes of Baiyangdian Wetland from 2016 to 2021.

Table 2: Transfer matrix of BGS in Baiyangdian Wetland from 2016 to 2021 (km^2).

	Green space	Blue space	Urban and rural construction space	2016
Green space	267.61	17.54	9.74	294.90
Blue space	11.32	47.09	0.01	58.42
Urban and rural construction space	6.31	0.12	17.81	24.24
2021	285.24	64.09	27.56	-

of Baiyangdian Wetland from 2016 to 2021 is shown in Figure 6. According to the figure, NP and PD decreased from 2016 to 2021, despite some fluctuations. NP dropped from 8,904 to 7,755, and PD falls from 24.4942 to 21.3334. These results suggest lower fragmentation of Baiyangdian Wetland. LSI shows the same trend as NP and PD, indicating simpler landscape shapes and better ecological function of the landscape.

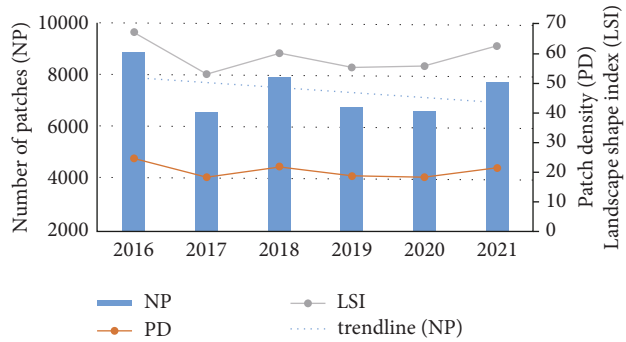


Figure 6: Changes in landscape pattern indices (NP, PB, and LSI) of Baiyangdian Wetland from 2016 to 2021.

3.3. BGS Changes of Baiyangdian Wetland. Figure 7 shows the largest patch index (LPI) of the BGS and construction space in urban and rural areas of Baiyangdian Wetland from 2016 to 2021. According to the figure, PD shows a decreasing trend. This indicates that systematic ecological management helps the landscape of Baiyangdian Wetland change from complicated and discontinuous patches to homogeneous and continuous BGS. As the fragmentation of the landscape declines, the ecosystem becomes more efficient and stable. The urban and rural construction space increased slightly in 2021 because of the enhanced ecological land.

4. Discussion

This paper provides a scientific reference for monitoring and evaluating the ecological environment of Baiyangdian Wetland and wetlands with similar features. The author finds that since the establishment of Xiong'an New Area, a well-organized and balanced wetland ecosystem has been built.

4.1. Baiyangdian Blue-Green Space Division. The concept of blue-green space (BGS) proposed in this study represents the two most significant driving factors of the ecological environment: vegetation and open water. However, different land-use classification systems will affect the results of blue-green spatial (BGS) change analysis, especially on the BGS division in wetlands. Current blue-green space research mainly analyzes urban evolution. For example, Cui Jie, Yang Liuqi, Song Shuang, et al analyzed the evolution and driving factors of street view and blue-green spatial landscape patterns of central cities in Xuzhou, Wuhan, and Northeast China [30–32]. However, their division of the blue-green space was too general, and the wetland classes were categorized as blue space, and the aquatic vegetation in the wetland was not separated. This limitation leads to the deviation of the division of the blue-green space. In their research on Baiyangdian Wetland, Zhu Jinfeng and Liu Chunlan divided the paddy field, reed field, and lotus pond into vegetation classes [33, 34]. Based on the definition of wetlands and the classification in Baiyangdian Wetland, this paper suggests that paddy fields and aquatic vegetation should be regarded as important components of green space. The blue space in the Baiyangdian Wetlands specifically refers to the water body, which contains open water lakes, wide waterways, ditches, and fishponds. The green space in wetland includes aquatic plants, terrestrial plants, cultivated land, and forest land.

4.2. Reasons for the Change of Blue-Green Space in Baiyangdian. Nevertheless, by the end of the study period, the wetland ecological restoration project was still under construction. As a result, there was a slight decrease in the proportion of BGS in the satellite images of Baiyangdian. A part of the green space in northwest Baiyangdian Wetland was converted into urban and rural construction space. To restore the ecology of the wetland area, a project of returning farmland to water body has been carried out in Pinghe River. Thus, the pattern of the wetland area changes, and the green space transferring to construction space is partial and temporal.

Since the establishment of Xiong'an New Area, ecological water replenishment has been implemented multiple times. The construction of ecological treatment projects and restoration will help change local blue-green spatial distribution in the future.

The BGS changes of Baiyangdian Wetland are greatly affected by weather and climate. Future studies could consider the influence of weather and climate on the BGS changes. Meanwhile, how human factors change the BGS of Baiyangdian Wetland should also be studied. The factors include economic development, water diversion, and water replenishment.

4.3. Advantages and Disadvantages of Classification Methods. Many researchers have used 30-meter resolution Landsat satellite data to classify wetland landscape patterns in long-term series to study the dynamic changes [35–38]. However, such a method has limited classification accuracy due to the medium-resolution Landsat images. To more efficiently divide the landscape of the Baiyangdian wetland into detailed categories, the decision tree method of NDVI-NDWI-SWIR has been created by a 10-meter resolution Sentinel-2 satellite. The results of validation with 0.8-meter resolution Gaofen-2 satellite data show that the classification accuracy is high. Compared with prior studies, this study focuses on the period after the establishment of the Xiong'an New Area. It adopts targeted measures to show the changes in Baiyangdian during its transformation into an ecological-

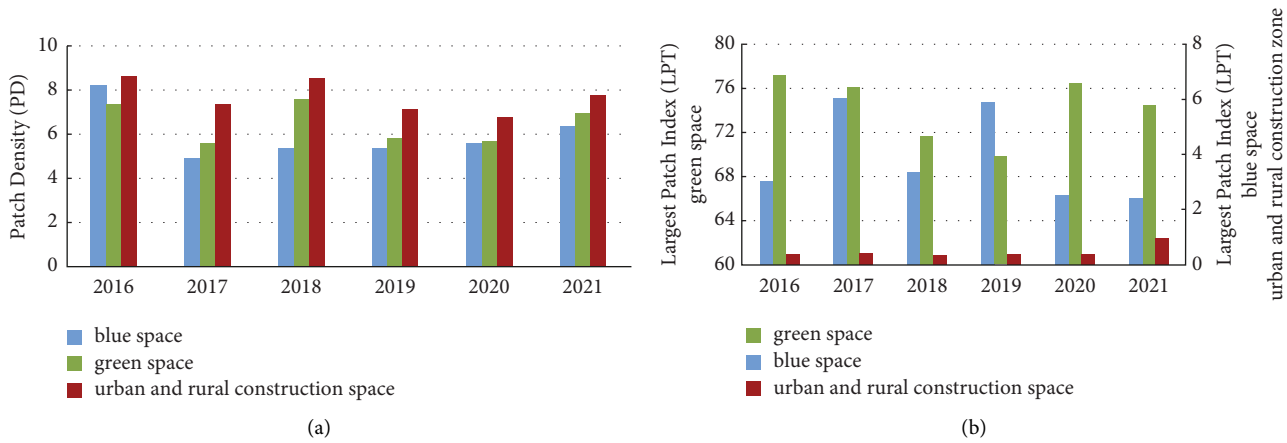


Figure 7: Patch properties of BGS and construction space in urban and rural areas of Baiyangdian Wetland from 2016 to 2021.

friendly city. The results of this study could provide practical references for the planning and management of Baiyangdian Wetland. However, as the Xiong'an New Area was established in recent years, most of the analyses are based on limited historical series data. There are still limitations to overcome in the analysis of landscape patterns. During the ecological restoration of the Baiyangdian Wetland and the construction of the Xiong'an New Area, further studies need to be carried out on the spatial distribution of blue-green space, the optimization of regional landscape patterns, and the construction of ecological patterns in the Xiong'an New Area. With the development of multisource remote sensing technology, more high-resolution data can be selected for the blue-green space inversion to improve classification accuracy.

5. Conclusion

Baiyangdian Wetland has become the core area of Xiong'an New Area since 2016. Based on the model of extracting the BGS information and calculation of landscape pattern indices, this paper has analyzed the spatiotemporal changes of the BGS in the Baiyangdian Wetland. The analysis shows that:

(1) The correlation coefficient between the classification result of Sentinel-2 in the 10-meter resolution and validation data of GF-2 in the 1-meter resolution is as high as 0.948, passing the 99% significance test. The MB, ME, and RMSE are 0.06, 0.203, and 0.144, respectively. The model applied in this study area has high accuracy and meets the demand of spatial analysis. Therefore, this model can be used for obtaining the detailed BGS and landscape changes in Baiyangdian Wetland from Sentinel-2 data.

(2) There is no obvious spatial transfer of vegetation in Baiyangdian Wetland from 2016 to 2021. Most green space concentrates in the north, the northwest, and the southwest of the wetland. The blue space is scattered in the southeast and the northeast of the wetland.

(3) From 2016 to 2021, NP, PD, LPI, and LSI of Baiyangdian Wetland all decreased, indicating a better ecological environment and more stable ecosystem of Baiyangdian Wetland since the establishment of Xiong'an New Area in 2017.

(4) The BGS proportion in Baiyangdian Wetland increased steadily from 2016 to 2020, among which the proportion of blue space increased by 10.8%. However, 2021 witnessed a slight decrease in BGS and a minor increase in the construction space in urban and rural areas. This was the result of the ecological construction. Baiyangdian Wetland ecology restoration project is still under construction. Its progress and changes will be monitored continuously.

Data Availability

The data that support the findings of this study are available upon reasonable request from the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This article was supported by the China Meteorological Administration innovation and Development Project (CXFZ2022J015); Innovative Team of Remote Sensing Application in Ecology (Doc No. Hebei Meteorological Bureau official letter [2021]186); Ecological Meteorological Monitoring and Evaluation Innovation Project of National Meteorological Center (NMCEMT20215409); and Scientific Research and Development Project of Hebei Meteorological Bureau (20ky09).

References

[1] H. Xu, T. Jin, and W. Liu, "Study on the scale and landscape pattern evolution characteristics of blue-green space in Suzhou-Wuxi-Changzhou metropolitan area, China," *Journal of Nanjing Forestry University (Natural Sciences Edition)*, vol. 46, no. 01, pp. 219–226, 2022.

[2] J. Sadler, A. Bates, and J. Hale, *Bringing Cities Alive: The Importance of Urban Green Spaces for People and Biodiversity. GASTON K J. Urban Ecology*, pp. 230–260, Cambridge University Press, Cambridge, UK, 2010.

[3] N. Kabisch and D. Haase, "Green spaces of European cities revisited for 1990–2006," *Landscape and Urban Planning*, vol. 110, pp. 113–122, 2013.

[4] K. R. Gunawardena, M. J. Wells, and T. Kershaw, "Utilising green and blue space to mitigate urban heat island intensity," *Science of The Total Environment*, vol. 584-585, pp. 1040–1055, 2017.

[5] Q. Q. Zhou, C. C. Konijnendijk Van Den Bosch, Z. G. Chen et al., "China's green space system planning: development, experiences, and characteristics," *Urban Forestry and Urban Greening*, vol. 60, Article ID 127017, 2021.

[6] F. H. Kong, H. W. Yin, P. James, L. R. Hutyra, and H. S. He, "Effects of spatial pattern of greenspace on urban cooling in a large metropolitan area of eastern China," *Landscape and Urban Planning*, vol. 128, pp. 35–47, 2014.

[7] A. Lausch, T. Blaschke, D. Haase et al., "Understanding and quantifying landscape structure-A review on relevant process characteristics, data models and landscape metrics," *Ecological Modelling*, vol. 295, pp. 31–41, 2015.

[8] R. Hao, D. Yu, Y. Liu et al., "Impacts of changes in climate and landscape pattern on ecosystem services," *Science of the Total Environment*, vol. 579, pp. 718–728, 2017.

[9] S. M. Sterling, A. Ducharne, and J. Polcher, "The impact of global land-cover change on the terrestrial water cycle," *Nature Climate Change*, vol. 3, no. 4, pp. 385–390, 2013.

[10] Q. Quan, W. Liang, D. Yan, and J. Lei, "Influences of joint action of natural and social factors on atmospheric process of hydrological cycle in Inner Mongolia, China," *Urban Climate*, vol. 41, Article ID 101043, 2022.

[11] M. R. Su, Y. Zheng, Y. Hao et al., "The influence of landscape pattern on the risk of urban water-logging and flood disaster," *Ecological Indicators*, vol. 92, pp. 133–140, 2018.

[12] D. C. Schneide, "The rise of the concept of scale in ecology: the concept of scale is evolving from verbal expression to quantitative expression," *BioScience*, vol. 51, no. 7, pp. 545–553, 2001.

[13] B. Junhong, O. Hua, C. Baoshan, W. Qinggai, and C. Hui, "Changes in landscape pattern of alpine wetlands on the Zoige Plateau in the past four decades," *Acta Ecologica Sinica*, vol. 28, no. 5, pp. 2245–2252, 2008.

[14] L. B. Huang, J. H. Bai, D. H. Yan, B. Chen, R. Xiao, and H. Gao, "Changes of wetland landscape patterns in Dadu River catchment from 1985 to 2000, China," *Frontiers of Earth Science*, vol. 6, no. 3, pp. 237–249, 2012.

[15] M. Yang, J. G. Gong, Y. Zhao, H. Wang, and C. P. Zhao, "Analysis of dynamic changes and trends in the landscape pattern of the Baiyangdian Region," *Acta Ecologica Sinica*, vol. 40, no. 20, pp. 7165–7174, 2020.

[16] C. Tang, D. Chen, Y. Yi, Y. Zhou, and S. Zhang, "Effects of ecological water supplement on vegetation dynamics in Lake Baiyangdian," *Journal of Lake Sciences*, vol. 34, no. 4, 2022.

[17] J. J. Wu, Z. W. Hu, X. Luo, G. F. Wu, and C. Wang, "Mapping long-term water dynamics of Baiyangdian Lake via deep learning method," *Environment and Ecology*, vol. 3, no. 11, pp. 35–42, 2021.

[18] H. B. Lin, X. L. Wu, F. Liu, and J. Bian, "Wetland resources monitoring for Baiyangdian lake by remote sensing technology," *Journal of Central South University of Forestry and Technology*, vol. 32, no. 4, pp. 27–130, 2012.

[19] C. L. Zhao, S. Qian, Q. Huang, G. M. Wu, and C. Z. Meng, "Prediction of minimum water level in Baiyangdian Lake based on precipitation," *Chinese Journal of Eco-Agriculture*, vol. 27, no. 8, pp. 1238–1244, 2019.

[20] C. Z. Meng, C. L. Zhao, Z. Y. Le, R. Y. Li, and M. M. Lu, "Information extraction of Baiyangdian wetland based on GF-2 remote sensing data," *IOP Conference Series: Earth and Environmental Science*, vol. 351, no. 1, Article ID 012042, 2019.

[21] B. Meng, X. S. Zhi, C. Wang, S. J. Yin, Y. M. Zhou, and Q. Zhao, "Remote sensing monitoring of dry riverbeds in Baiyangdian Lake Basin," *Science Technology and Engineering*, vol. 22, no. 9, pp. 3827–3833, 2022.

[22] J. Xiong, P. Thenkabail, J. Tilton et al., "Nominal 30m cropland extent map of continental Africa by integrating pixel-based and object-based algorithms using sentinel-2 and landsat-8 data on google earth engine," *Remote Sensing*, vol. 9, no. 10, pp. 1065–1069, 2017.

[23] H. Tian, Y. Wang, T. Chen, L. Zhang, and Y. Qin, "Early-Season mapping of winter crops using sentinel-2 optical imagery," *Remote Sensing*, vol. 13, no. 19, p. 3822, 2021.

[24] H. Tian, J. Pei, J. Huang et al., "Garlic and winter wheat identification based on active and passive satellite imagery and the google earth engine in northern China," *Remote Sensing*, vol. 12, no. 21, p. 3539, 2020.

[25] A. Shelestov, M. Lavreniuk, N. Kussul, A. Novikov, and S. Skakun, "Large scale crop classification using Google earth engine platform," *Geoscience and Remote Sensing Symposium*, IEEE, 2017.

[26] Z. F. Shen, J. L. Li, and X. J. Yu, "Water information extraction of Baiyangdian wetland based on the collaborative computing method," *Journal of Geo-Information Science*, vol. 18, no. 5, pp. 690–698, 2016.

[27] R. Jin, X. Li, and M. G. Ma, "Key methods and experiment verification for the validation of quantitative remote sensing products," *Advances in Earth Science*, vol. 32, no. 6, pp. 630–642, 2017.

[28] T. T. Han, J. W. Luan, T. T. Shao, and X. F. Chen, "Research on wetland landscape pattern changes and driving factors in Wuhan city," *Geographic Information System*, vol. 20, no. 3, pp. 12–17, 2022.

[29] W. R. Yang, "Spatiotemporal change and driving forces of urban landscape pattern in Beijing," *Acta Ecologica Sinica*, vol. 35, no. 13, pp. 4357–4366, 2015.

[30] J. Cui, H. Xu, and W. Liu, "Analysis on the evolution of blue-green spatial landscape pattern in Xuzhou Metropolitan Area in the past 20 years," *Landscape Architecture Academic Journal*, vol. 39, no. 7, pp. 75–81, 2022.

[31] L. Yang, Y. Zhou, and J. Luo, "Study on spatial pattern of urban blue-green system and its synergistic relationship with urban development in Wuhan based on spatiotemporal evolution analysis," *Landscape Architecture Academic Journal*, vol. 39, no. 7, pp. 66–74, 2022.

[32] S. Song, M. Shi, and S. Hu, "Evolutions and driving mechanisms of urban blue-green spaces in northeast China: a case study with the urban central district of Harbin City," *Journal of Nanjing Forestry University (Natural Sciences Edition)*, vol. 46, no. 4, pp. 221–229, 2022.

[33] J. Zhu, Y. Zhou, and S. Wang, "Analysis of changes of Baiyangdian wetland from 1975 to 2018 based on remote sensing," *Journal of Remote Sensing*, vol. 23, no. 5, pp. 971–986, 2019.

[34] C. Liu, G. Xie, and H. Huang, "Shrinking and drying up of Baiyangdian lake wetland: a natural or human cause?" *Chinese Geographical Science*, vol. 16, no. 4, pp. 314–319, 2006.

[35] P. Cong, K. Chen, L. Qu, and J. Han, "Dynamic changes in the wetland landscape pattern of the Yellow River delta from 1976 to 2016 based on satellite data," *Chinese Geographical Science*, vol. 29, no. 3, pp. 372–381, 2019.

[36] X. Shen, H. Yang, C. Wang, and J. Li, "Dynamic changes of wetland landscape in the coastal zone of Yancheng from 1990 to 2020 based on remote sensing images," *Journal of Shanghai Ocean University*, vol. 31, no. 4, pp. 972–983, 2022.

[37] B. Yu, Y. Xie, X. Ma, and B. Cui, "Wetland landscape pattern change and its damage degree in Guangdong-Hong Kong-Macau Bay Area in recent 40 years," *Environment and Ecology*, vol. 4, no. 5, pp. 59–68, 2022.

[38] W. Yan, D. Gu, and H. Sun, "Dynamic study on landscape structure of Yancheng coastal wetlands in recent 40 years," *Marine Environmental Science*, vol. 41, no. 3, pp. 437–443, 2022.

Hindawi
Advances in Meteorology
Volume 2022, Article ID 9183882, 14 pages
https://doi.org/10.1155/2022/9183882

Research Article

Spatiotemporal Climate Variation and Analysis of Dry-Wet Trends for 1960–2019 in Jiangsu Province, Southeastern China

Mengsheng Qin,[1] **Liting Zhang**,[1] **Shiquan Wan**,[1] **Yuan Yue**,[2] **Qiong Wu**,[1] **and Lu Xia**[1]

[1]*Yangzhou Municipal Meteorological Bureau, Yangzhou, China*
[2]*Jilin Meteorological Bureau, Changchun, China*

Correspondence should be addressed to Shiquan Wan; wan_sq@qq.com

Received 6 May 2022; Accepted 5 August 2022; Published 27 August 2022

Academic Editor: Upaka Rathnayake

The spatiotemporal characteristics of dry-wet trends were identified and assessed, and the dominant meteorological factors were identified for the climate of Jiangsu province in humid southeastern China for the period 1960–2019. We conducted the research using data for the entire Jiangsu province as well as three major regions in Jiangsu (Huaibei, Jianghuai, and Sunan) with different regional climates. The results showed that decreased precipitation and relative humidity in spring and autumn over the study period were mainly responsible for the dry trends of the climates of all three regions and the entire province. Precipitation had a greater influence in spring and relative humidity in autumn. Decreases in sunshine hours and wind speed were responsible for the summer wet trends of the climates of Huaibei and Jianghuai and the entire province. However, precipitation increased significantly in the summer and was responsible for the increasing wet trend in Sunan. Significantly increased precipitation in winter was primarily responsible for the increasing wetness in Jianghuai and Sunan and the entire province in that season. However, the wet trend in northern Huaibei in winter was mainly caused by the decrease in wind speed over the study period. For the growing season and annually, the positive effects of changes in wind speed, sunshine hours, and precipitation led to increased humidity index in Jianghuai, Sunan, and the entire province. Precipitation showed a decreasing trend that countered the positive effects of decreases in wind speed and sunshine hours, which resulted in a slight decrease in the humidity index in Huaibei for both the growing season and annually. Sensitivity analysis indicated that the humidity index was positively sensitive to precipitation and relative humidity and negatively sensitive to air temperature, wind speed, and sunshine hours in Jiangsu province during 1960–2019. Overall, the humidity index in this region of southeastern China was most sensitive to changes in precipitation followed, in order of sensitivity, by sunshine hours, air temperature, wind speed, and relative humidity. Our findings provide a theoretical basis for adjusting irrigation programs and efficient utilization of water resources at the regional scale in humid southeastern China.

1. Introduction

It is well recognized that extreme weather events are largely caused by global climate change and anthropogenic activity [1], and they severely affect human society and the natural environment [2]. Drought is a significant and influential meteorological phenomenon in most climate zones [3]. Widespread drought causes tremendous economic losses and is a great threat to food, water and energy security, civil society, and natural habitats [2, 4–6]. It is therefore important to study the spatiotemporal characteristics of droughts and their causes; that is, to analyze and predict spatiotemporal variations in indicators of extreme wet or dry conditions and to identify reasons for these variations.

Researchers have developed dry and wet indexes based on air temperature (T_a), wind speed (WS), precipitation (Pre), reference evapotranspiration (ET_o), and other meteorological factors [7]. These indexes include the Palmer drought severity index (PDSI) [8], the standardized precipitation index (SPI) [9], the humidity index (HI) [10], and the standardized precipitation evapotranspiration index (SPEI) [11, 12]. Yang et al. [13] compared the applicability of

these indicators to China's conditions and recommended that the HI be used in this country. Ma and Fu [14] found that HI had performed well in China over the last five decades. HI, which reflects changes in precipitation and air temperature, has been used extensively in studies around the world [15–20].

In the literature review, it is found that the variations in dry-wet trends and their causes have been examined for different seasons, regions, and durations worldwide. Chinese researchers have studied the annual increase in aridity in regions within China such as the Zoige Wetland [21], Bosten Lake [22], Loess Plateau [15], semi-arid or semi-humid areas [6], the entire northern part of China [23], and even China in its entirety [7]. Annual dry trends have been identified in other countries such as Korea [24], Iran [16], Guatemala [25], Iraq [20], Rwanda [26], India [27], Middle East [28], and Iberian Peninsula [29]. In general, these studies attributed the drying mainly to surface warming, decreased precipitation, and increased evapotranspiration [5–7, 15, 16, 20, 21, 27–29]. However, the main drivers of the wet trends in the Yangtze River Delta in eastern China [1] and the Tibetan Plateau, North Xinjiang, and the Tuotuo River of western China [30–32], were found to be increased precipitation and relative humidity (RH) together with a reduction in wind speed and sunshine hours (SSH) [1, 30–32]. It was also observed that the dominant meteorological variables caused opposite dry-wet trends in different months or seasons. For example, Jin et al. [21] found that the Zoige Wetland in southwestern China tended to become drier in spring and winter and wetter in summer and autumn during the period 1961–2016. In the Zhejiang province of southeastern China, in the period 1971–2015, there were dry trends in April, May, and September, a wet trend in August, and no observable dry or wet trends in other months [33]. Muhire et al. [26] reported increased aridity in the rainy season in Rwanda in the period 1961–1992, but a wet trend in other months. In summary, there has not been unanimous scientific agreement on the causes of spatiotemporal variation in aridity or moisture or on the causes of drier or wetter regional climates.

Studies have found that in China (1) both wet [1, 31, 34] and dry trends [6, 15, 21, 22] occurred in different regions and during different seasons; (2) the drivers of wet and dry trends were not only temperature and precipitation but also other meteorological parameters such as WS, SSH, and RH [6, 7, 15, 21, 31, 32, 34]; and (3) most research was concerned with arid areas of northern and western China but ignored wet and dry trends in the climate of humid southeastern China, which was greatly affected by global climate change due to human activity.

We conducted our research on Jiangsu province and proposed the hypothesis that increasing air temperature, mainly induced by global warming and rapid regional urbanization, would lead to a dry trend in this province. The aims of this study were (1) to explore the spatiotemporal changes that determine the humidity index using the observed meteorological data; (2) to determine the contribution of each meteorological parameter to variation in the humidity index; and (3) to perform sensitivity analysis to identify the relative change in HI due to relative changes in the five parameters T_a, WS, RH, SSH, and Pre.

2. Study Area

Jiangsu province (30°45′–35°20′N, 116°18′–121°57′E) is an important economic and agricultural region in southeastern China. It has a high population density of 770 people/km^2 [35, 36] and has witnessed the most rapid urbanization in China [36]. Intensification of human activities, land use changes associated with urbanization, increased industrial emissions, and a sharp increase in water demand have influenced its regional climate and thus affected dry-wet climate trends as well as the forces of global climate change. The province is surrounded by Zhejiang province and the city of Shanghai to the south, Anhui province in the west, Shandong province in the north, and the Yellow Sea in the east; the coastline is over 1000 km long (Figure 1). Jiangsu province is 460 km from north to south and is partitioned into three regions (Huaibei, Jianghuai, and Sunan) by the Huai River and the Yangtze River (Figure 1). The province spans more than 5° of latitude, which leads to differences in the regional climate of three regions, such as an air temperature range of 13–16.5°C and a precipitation range of 800–1200 mm [37]. These differences in regional climate cause differences in moisture between the regions. We therefore examined the moisture in each region as well as in the province as a whole.

3. Materials and Methods

3.1. Data Source. Meteorological data used in the study included daily mean air temperature (T_a, °C), maximum temperature (T_{max}, °C), minimum temperature (T_{min}, °C), wind speed (WS, m/s), sunshine hours (SSH, h), relative humidity (RH, %), and precipitation (Pre, mm) measured at 24 stations during 1960–2019. All data were preprocessed by the China Meteorology Administration according to the 2004 *Standard of Surface Meteorology Observation by the China Meteorology Administration* was shared in the China Meteorological Data Service Center (http://data.cma.cn/).

Jiangsu province has been an important rice-growing region for thousands of years [38]. We conducted our research for individual seasons, spring (March–May), summer (June–August), autumn (September–November), and winter (December–February in the next year) as well as the entire year (January–December) and the rice-growing season (May–October).

3.2. Humidity Index. The humidity index HI is calculated by [39]

$$\mathrm{HI} = \frac{P}{\mathrm{ET}_o}, \tag{1}$$

where P is precipitation (mm) and ET_o is reference evapotranspiration (mm) and was calculated by the FAO 56 Penman–Monteith (PM) model. This model has been recommended by the Food and Agricultural Organization (FAO) and is expressed as [40]

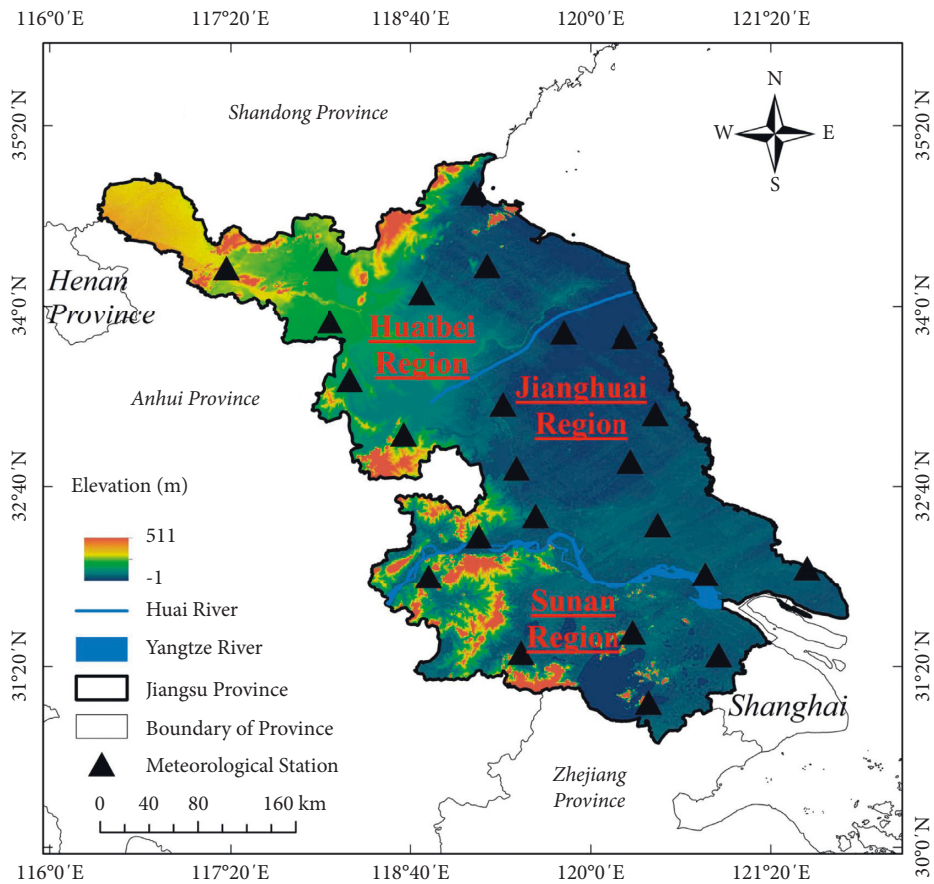


Figure 1: The three regions of Jiangsu province (Huaibei, Jianghuai, and Sunan), rivers, and meteorological stations.

$$\mathrm{ET}_o = \frac{0.408\Delta(R_n - G) + \gamma(900/T + 273)U_2(e_s - e_a)}{\Delta + \gamma(1 + 0.34U_2)}, \quad (2)$$

where R_n is the net radiation at the crop surface (MJ/m^2/d); G is the soil heat flux density (MJ/m^2/d), which can be ignored at a daily scale; T is the mean daily air temperature (°C); U_2 is the wind speed at 2 m height (m/s); e_s is the saturation vapor pressure (kPa); e_a is the actual vapor pressure (kPa); γ is the psychrometric constant (kPa/oC); and Δ is the slope of the vapor pressure curve (kPa/oC).

A set of equations were devised by Allen et al. [40] to calculate the parameters in equation (2), which we used with the daily meteorological data of 24 stations in Jiangsu province. We calculated $\mathrm{ET_o}$ at a daily scale using equation (2), combined observed daily precipitation data with calculated daily $\mathrm{ET_o}$ at seasonal and annual scales, and then used equation (1) to compute seasonal and annual HI for each region and for the entire province.

3.3. Trend Analysis. The trend of long-term climatic data is a critical indicator in analyzing climate change. Many kinds of trends can be defined, such as extrinsic and predetermined trends, intrinsic and adaptive trends, and using empirical mode decomposition for trend extraction [41]. The commonest and most simple trend is a straight line fitted to the data. A trend can be found by identifying changes in the moving average of a data series. In addition to these, trends that emerge from complicated trend extraction methods (such as regression analysis and Fourier-based filtering) are all extrinsic and predetermined trends. In this study, we used linear fitting and the Mann–Kendall test to identify extrinsic predetermined trends [42, 43] to identify changes in meteorological parameters and HI at both seasonal and annual scales. The Mann–Kendall test was also used to determine the significance of the trends at the 0.001, 0.01, 0.05, or 0.1 levels. The Mann–Kendall test is based on the hypothesis that the expected change in the series is a monotonic trend. The null hypothesis H_0 is that in a data series X (X_k, $k = 1, 2, \ldots, n$), X_k is independent and randomly distributed. The alternative hypothesis H_1 is that a monotonic trend exists in X. The statistic S is calculated as

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} \mathrm{sgn}(X_j - X_k), \quad (3)$$

where X_j represents the sequential data values, n is the number of the dataset, and

$$\mathrm{sgn}(X_j - X_k) = \begin{cases} 1, & \text{if } X_j - X_k > 0, \\ 0, & \text{if } X_j - X_k = 0, \\ -1, & \text{if } X_j - X_k < 0. \end{cases} \quad (4)$$

When $n > 10$, S is approximately normally distributed with $E(S) = 0$. The variance of statistic S can be calculated by

$$\mathrm{Var}(S) = \frac{1}{18}\left[n(n-1)(2n+5) - \sum_{w=1}^{v} t_p(t_w - 1)(2t_w + 5)\right], \tag{5}$$

where v is the number of tied groups and t_w is the number of data values in group w.

The standard test statistic (Z) is

$$Z = \begin{cases} \dfrac{s-1}{\sqrt{\mathrm{Var}(s)}}, & \text{if } s > 0, \\ 0, & \text{if } s = 0, \\ \dfrac{s+1}{\sqrt{\mathrm{Var}(s)}}, & \text{if } s < 0. \end{cases} \tag{6}$$

The null hypothesis H_0 is rejected when $|Z| > Z_{1-\alpha/2}$, where $Z_{1-\alpha/2}$ is the standard normal deviation. $|Z| > 1.28$, 1.64, and 2.32 indicate that change trends are significant at > 90% ($p < 0.1$), >95% ($p < 0.05$), and >99% ($p < 0.01$), respectively.

The Theil–Sen estimator [44, 45] was used to evaluate the magnitudes of the parameter trends. These simple but effective statistical methods have been widely used in hydrological-meteorological time series dataset trend analysis [32, 46–49]. The Theil–Sen estimator is calculated as

$$\beta = \mathrm{Median}\left(\frac{X_j - X_k}{j-k}\right), \quad 1 < k < j < n, \tag{7}$$

where β is the estimated magnitude of the slopes of HI trends. $\beta > 0$ represents an increasing trend, and $\beta < 0$ represents a decreasing trend.

3.4. Detrended Method. Linear detrending was used to quantify the contributions of five meteorological variables to the trend of HI at both seasonal and annual scales for each region and for the entire Jiangsu province. Previous studies [50–55] have described this method in three steps. (1) For each meteorological parameter, detrend the parameter time series to remove any overall increasing or decreasing trend. (2) Recalculate HI using one detrended meteorological parameter time series with the other four original parameter time series, and repeat this for each parameter. (3) Compare the recalculated HI with the original HI and use the indicator R to quantify the changes in meteorological variables that affect the trends in HI. R is calculated by

$$R = \sum_{i=1}^{m} \frac{\left(\mathrm{HI}_o^o - \mathrm{HI}_o^R\right)}{\mathrm{HI}_o^o i}, \tag{8}$$

where HI_o^o and HI_o^R are the original and the recalculated HI, respectively, and m is the length of the time series. $R > 0$, $R < 0$, and $R = 0$ indicate the positive, negative, and null effects of changes in the parameter on HI. A greater value of $|R|$ indicates a greater contribution of this variable to the HI trend.

3.5. Sensitivity Analysis. We used the method of sensitivity analysis developed by McCuen et al. [56] to identify the percentage change in HI due to percentage change in each meteorological parameter. This method has been widely used to analyze the sensitivity coefficients that relate ET_o to various meteorological parameters [57–60], and the method has also been used in recent years to examine the sensitivity of the dry-wet index to meteorological parameters [31, 61–63]. The sensitivity coefficient of HI to a meteorological parameter is calculated by

$$S_{xi} = \lim_{\Delta x_i \longrightarrow 0}\left(\frac{\Delta \mathrm{HI}/\mathrm{HI}}{\Delta x_i / x_i}\right) = \frac{\partial \mathrm{HI}}{\partial x_i}\frac{x_i}{\mathrm{HI}}, \tag{9}$$

where x_i is meteorological parameter i, and $S(x_i)$ is the sensitivity coefficient of the parameter and indicates the percentage change in HI caused by a percentage change in the parameter x_i. A positive or negative value of $S(x_i)$ indicates whether HI will increase or decrease as the parameter x_i increases.

4. Results

4.1. Temporal Variation of Five Meteorological Variables and Reference Evapotranspiration (ET_o). T_a has shown significant increasing trends ($p < 0.05$) in Jiangsu province over the past 60 years, with lower and similar rates of change in Huaibei and Jianghuai (Table 1). T_a in southern Sunan had the greatest rate of increase, and T_a increased for the entire province at respective rates of 0.041, 0.015, 0.026, 0.036, 0.022, and 0.29°C/y for spring, summer, fall, winter, growing season, and annually (Table 1). The rates of change of T_a were ordered as spring > winter > annual > autumn > growing season > summer (Table 1).

In contrast to T_a, WS for the three regions and the entire province showed significant decreasing trends ($p < 0.05$) over the period 1960–2019 (Table 1). One notable difference from the behavior of T_a was that the greatest variation in WS was found in northern Huaibei with a decreasing rate of change ranging from −0.027 m/s/y in spring to −0.02 m/s/y in autumn (Table 1). The rates of change in WS in Jianghuai and Sunan were both less than that in northern Huaibei and resulted in the decreasing trends ranged from −0.022 m/s/y in spring to −0.015 m/s/y in summer for the entire province (Table 1).

In the period 1960–2019, RH also showed decreasing trends in Jiangsu province (Table 1). The negative rates of change in RH were least in Huaibei, in the range of −0.106%/y ($p < 0.05$) in spring to −0.013%/y in winter. In southern Sunan, RH had the greatest rate of decrease, between −0.223%/y ($p < 0.001$) in spring and −0.076%/y ($p < 0.05$) in winter (Table 1). RH for the entire province decreased significantly ($p < 0.01$) in all seasons except for winter, when RH decreased at an insignificant rate −0.047%/y. The significantly decreasing trends of the entire province ranged from −0.161%/y in spring to −0.078%/y in both summer and autumn (Table 1).

There was little change in SSH in spring for all three regions during 1960–2019 (Table 1). In the other three seasons and annually, SSH showed significant decreasing trends ($p < 0.05$) for the three regions (Table 1). Similar to

Table 1: Seasonal and annual trends of daily mean air temperature (T_a), wind speed (WS), relative humidity (RH), sunshine hours (SSH), precipitation (Pre), and reference evaporation (ET_o) in Huaibei (HB), Jianghuai (JH), Sunan (SN), and the entire Jiangsu province during 1960–2019 using the Mann–Kendall and Theil–Sen estimator tests.

Variable	Zone	Spring	Summer	Autumn	Winter	Growing season	Annual
T_a (°C/y)	HB	0.039***	0.011*	0.023***	0.04***	0.019***	0.028***
	JH	0.038***	0.013*	0.024***	0.036***	0.02***	0.027***
	SN	0.049***	0.022**	0.035***	0.048***	0.03***	0.037***
	Whole	0.041***	0.015*	0.026***	0.036***	0.022***	0.029***
WS (m/s/y)	HB	−0.027***	−0.021***	−0.02***	−0.022***	−0.021***	−0.023***
	JH	−0.021***	−0.013***	−0.016***	−0.019***	−0.014***	−0.017***
	SN	−0.019***	−0.014***	−0.015***	−0.022***	−0.014***	−0.017***
	Whole	−0.022***	−0.015***	−0.017***	−0.021***	−0.016***	−0.019***
RH (%/y)	HB	−0.106*	−0.035+	−0.046	−0.013	−0.035*	−0.044*
	JH	−0.164***	−0.077**	−0.073+	−0.049***	−0.088***	−0.089***
	SN	−0.223***	−0.121***	−0.139***	−0.076*	−0.159***	−0.135***
	Whole	−0.161***	−0.078***	−0.078**	−0.047	−0.09***	−0.084***
SSH (h/y)	HB	−0.007	−0.036***	−0.021***	−0.027***	−0.028***	−0.023***
	JH	0.005	−0.031***	−0.013*	−0.017***	−0.02***	−0.015***
	SN	0.005	−0.036***	−0.015**	−0.02**	−0.023***	−0.018***
	Whole	0.003	−0.033***	−0.015**	−0.02***	−0.024***	−0.018***
Pre (mm/y)	HB	−0.33	−1.2*	−0.17	0.64**	−1.78*	−1.55*
	JH	−0.53	0.55	0.13	1.02**	0.06	0.59
	SN	−0.52	1.84*	0.19	1.7***	1.68*	2.8*
	Whole	−0.45	0.44	0.09	1.07***	0.23	0.53
ET_o (mm/y)	HB	0.04	−0.82***	−0.28*	−0.19*	−0.19***	−1.37***
	JH	0.65***	−0.25	0.06	−0.01	0.003	0.33
	SN	1.04***	−0.02	0.36**	0.01	0.75*	1.32**
	Whole	0.56**	−0.38*	0.01	−0.04	−0.14	0.05

Note: ***, **, *, and + indicate significance levels of 0.001, 0.01, 0.05, and 0.1, respectively.

WS, these SSH trends were the greatest in northern Huaibei and were least in Jianghuai and Sunan; rates of decrease were between −0.036 h/y in summer for Huaibei and −0.013 h/y in autumn for Jianghuai (Table 1). For the entire province, SSH showed a significant decreasing trend ($p < 0.01$) for three seasons (there was almost no change in spring) and annually. The negative rates of change ranged from −0.033 h/y in summer to −0.015 h/y in autumn (Table 1).

Table 1 shows that in the period 1960–2019, Pre showed a significant positive increasing trend ($p < 0.01$) in winter with a rate of change of 0.64 mm/y in Huaibei. However, Pre decreased significantly in Huaibei at a rate of −1.2 mm/y ($p < 0.05$) in summer and showed significant decreasing trends ($p < 0.05$) in the growing season (−1.78 mm/y) and annually (−1.55 mm/y) (Table 1). In southern Sunan, Pre increased significantly ($p < 0.05$) in summer and winter and annually with rates of change between 1.68 mm/y for the growing season and 2.8 mm/y at the annual scale; changes for spring and autumn were not significant (Table 1). Pre showed significant ($p < 0.01$) increasing trends, with rates of change of 1.02 and 1.07 mm/y, for Jianghuai and the entire province only for winter (Table 1), respectively.

Similar to Pre, ET_o showed both positive and negative trends for different seasons and annually in each region and the entire province (Table 1). In northern Huaibei, this variable showed significant decreasing trends ($p < 0.05$) in all seasons except spring, with little change (Table 1). The negative rates of change in Huaibei ranged from −0.082 mm/y for summer to −0.019 mm/y for both winter and the growing season. Jianghuai differed from Huaibei, showing significant changes in ET_o only for spring ($p < 0.001$) with a rate of change of 0.65 mm/y (Table 1). Southern Sunan showed significant increasing trends ($p < 0.01$) for spring and winter with rates of change of 1.04 and 0.36 mm/y, which resulted in ET_o increasing significantly ($p < 0.05$) by 0.75 and 1.32 mm/y, respectively, in the growing season and annually (Table 1). The changes in seasonal and annual ET_o in each region eventually caused ET_o for the entire province to increase in spring and autumn and annually and decrease in the other seasons. However, these changes were only significant in spring and summer, with rates of change of 0.56 and −0.38 mm/y (Table 1).

4.2. Analysis of Seasonal and Annual Humidity Index (HI)

4.2.1. Characteristics of Monthly Humidity Index (HI), Reference Evapotranspiration (ET_o), and Precipitation (Pre). Variations in monthly Pre, ET_o, and HI are shown in Figure 2. It is clear that both Pre and ET_o first increased and then decreased, with maximum values of 233 and 137 mm in July. These two variables had minimum values of almost 25 mm for winter (December–February). Maximum HI was 1.7, also for July, with the greatest positive difference between Pre and ET_o. The greatest negative difference between Pre and ET_o was for May, with minimum HI close to 0.7.

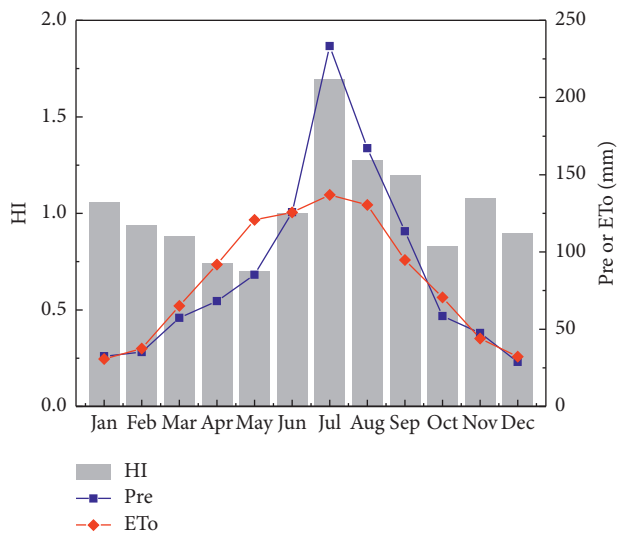


Figure 2: Variations in monthly humidity index (HI), reference evapotranspiration (ET_o), and precipitation (Pre).

4.2.2. Spatiotemporal Characteristics of Seasonal and Annual Humidity Index (HI) in Jiangsu Province. During 1961–2019, the increasing HI was only found in two sites of the Huaibei region in spring. For the other two regions in this season, we observed decreasing HI in all 17 sites (Figure 3(a)). In summer, the HI increased in 57.1%, 83.3%, and 100% of the sites of Huaibei, Jianghuai, and Sunan regions, respectively (Figure 3(b)).

In autumn, the HI generally showed increasing trends in Huaibei and Jianghuai regions with 85.7% and 58.3% of the sites with positive trends in each region (Figure 3(c)), respectively. For the southern Sunan region, the decreasing HI was found in all sites during 1961–2019 (Figure 3(c)). It was observed that the HI showed increasing trends in winter in all sites of Jiangsu province (Figure 3(d)). The decreasing HI was observed in 71.4% of the sites of the Huaibei region in both the growing season and annually (Figures 3(e) and 3(f)). Also, in these two periods, the HI showed increasing trends in most sites of the Jianghuai region (66.6% for the growing season and 75% annually) in 1961–2019. Furthermore, the increasing HI was found in all sites of the southern Sunan region in the growing season and annually (Figures 3(e) and 3(f)).

In addition to each site, we also used the Mann–Kendall test to examine seasonal and annual trends of HI in the whole Huaibei and Jianghuai Sunan regions, and Jiangsu province (Table 2), which were generally consistent with those shown in Figure 3. The Mann–Kendall test showed increasing trends in summer (only significant in Sunan) and winter (significant in all three regions and the entire province) during 1960–2019. In spring, the HI decreased in all three regions and the entire Jiangsu province but the decrease was only significant in Jianghuai and Sunan regions (Table 2). In autumn, the HI showed insignificant positive trends in Huaibei and Jianghuai regions and insignificant negative trends in Sunan and the entire province. Table 2 also exhibits that the growing season and annual HI only decreased in the Huaibei region and showed increasing trends in the other two regions and the entire province.

4.3. Sensitivity of Humidity Index (HI) to Key Meteorological Variables. Figure 4 shows the sensitivity coefficients relating HI to the five meteorological variables in each region and the entire province at both seasonal and annual scales for the period 1960–2019. The HI in Jiangsu province was clearly negatively sensitive to T_a, WS, and SSH, and positively sensitive to RH and Pre (Figure 4). The HI was least sensitive to T_a in winter; the absolute values of the sensitivity coefficients were <0.1 for all three regions and the entire province (Figure 4(a)). The HI was more sensitive to T_a in the other seasons and annually, with sensitivity coefficients ranging from −0.5 for Huaibei in spring to −0.23 for Sunan in autumn (Figure 4(b)). The HI was most sensitive to WS in Jiangsu province in winter and showed less sensitivity in the other seasons and annually (Figure 4(b)). The sensitivity coefficients were less than for T_a and ranged between −0.015 for Huaibei in summer and −0.02 for Sunan in winter (Figure 4(b)). The HI was more sensitive to SSH than to T_a and WS in Jiangsu province (Figure 4(c)). The HI was least sensitive to SSH in winter; sensitivity coefficients ranged from −0.33 in Huaibei to −0.18 in Sunan. For other seasons and annually, the absolute values of sensitivity coefficients of SSH were all >0.3, with maximum values in Huaibei −0.97 in spring, −0.93 in summer, −0.92 in the growing season, and −0.69 both in autumn and annually (Figure 3(c)).

The HI in Jiangsu province was positively sensitive to RH and Pre. It can be seen that the sensitivity of HI to RH was small to negligible, being <0.05 (Figure 4(d)). HI showed the greatest sensitivity to Pre in each region and the entire province, with sensitivity coefficients all >1 (Figure 4(e)). The sensitivity of HI to Pre at an annual scale was the greatest, with coefficients ranging from 1.47 in Jianghuai to 1.52 in Sunan. The sensitivity of HI to Pre was least in summer, with coefficients in the range of 1.09 in Sunan to —1.13 in Jianghuai (Figure 4(e)).

In Jiangsu province, the HI was most sensitive to Pre, followed by SSH, T_a, WS, and RH, in all seasons and annually except for winter (Figure 4). In the cold and dry winter, the sensitivity of HI to all five meteorological parameters was ranked Pre > SSH > WS > T_a > RH (Figure 4).

4.4. Contributions of Five Key Meteorological Parameters to Humidity Index (HI) Trends. We used the indicator R_{HI}, calculated by linear detrending, to evaluate the effects of changes in each meteorological variable on HI trends for different seasons and different regions (Figure 5). In spring, the decreases in RH and Pre were the main reasons for decreases in HI in all regions and the entire province. RH contributed more in Sunan (Figure 5(c)), and Pre contributed more in the other regions and the entire province (Figure 5(a), 5(b), and 5(d)). In summer, the negative effects of the decrease in Pre countered the positive effects of the changes in SSH and WS, resulting in little change in HI in Huaibei (Figure 5(a)). For the other two regions and the entire province, the HI increases in summer were mainly

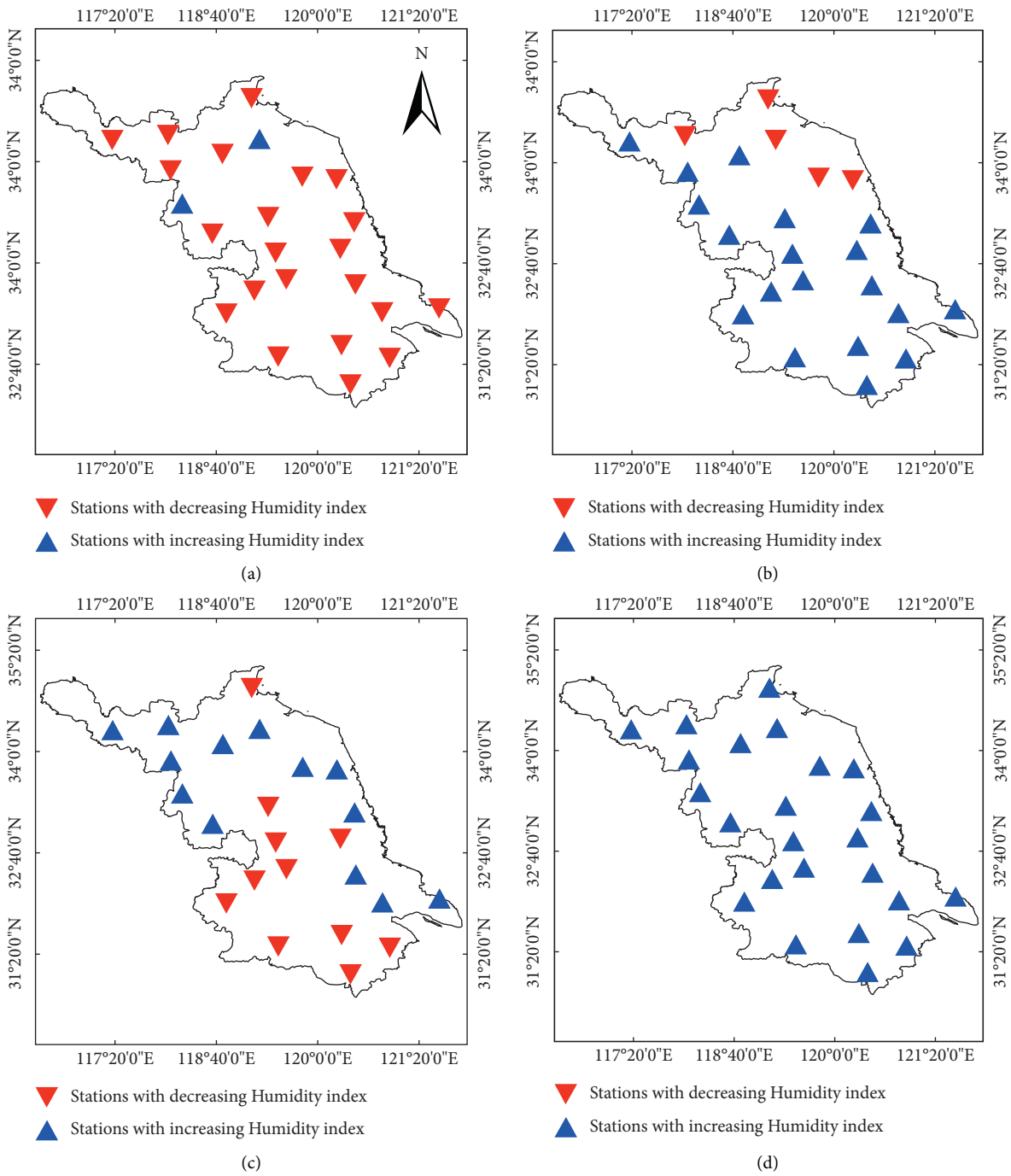


Figure 3: Continued.

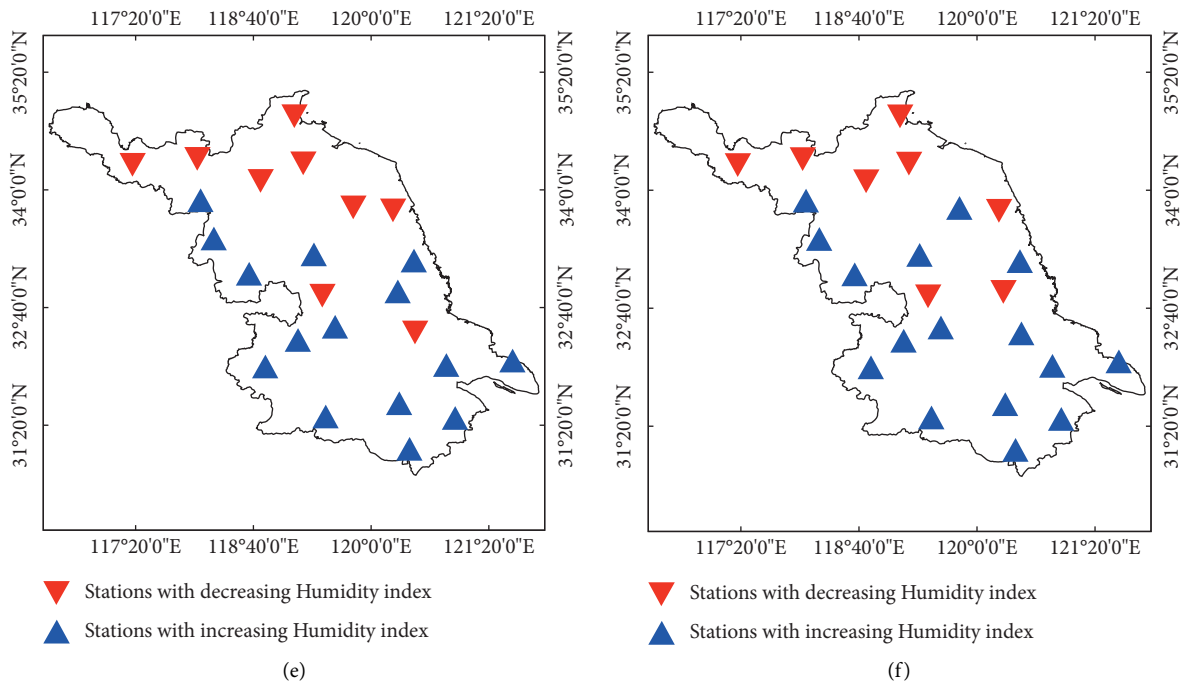


Figure 3: Spatial distribution of the sites with decreasing or increasing humidity index (HI) at seasonal and annual scales during 1961–2019 throughout Jiangsu province; the solid red inverted and blue regular triangles indicate the negative and positive trends of HI, respectively. (a) Spring, (b) summer, (c) autumn, (d) winter, (e) growing season, (f) annual.

Table 2: Trends of seasonal and annual humidity index (HI) in Huaibei, Jianghuai, Sunan, and the entire province during 1960–2019 using the Mann–Kendall test.

Season	Huaibei	Jianghuai	Sunan	Entire province
Spring	↓	↓$^{+}$	↓*	↓
Summer	↑	↑	↑$^{+}$	↑
Autumn	↑	↑	↓	↓
Winter	↑**	↑**	↑***	↑**
Growing season	↓	↑	↑	↑
Annual	↓	↑	↑	↑

Note: ***, **, *, and $^{+}$ indicate significance levels of 0.001, 0.01, 0.05, and 0.1, respectively.

caused by changes in WS, SSH, and Pre, with Pre having a greater effect than the other parameters in Sunan (Figure 5(c)), the decrease in SSH had a greater effect in both Jianghuai and the entire province (Figure 5(b) and 5(d)).

In autumn, similar to spring, the decreases in RH and Pre had the most effect on the negative trend of HI in Huaibei (Figure 5(a)). However, the combined effects of different parameters caused insignificant changes in the HI in the other two regions and the entire province in autumn (Figure 5). It is clear that the significant increase in Pre was responsible for the positive HI in Jianghuai, Sunan, and the entire province in winter (Figures 5(b)–5(d)). In northern Huaibei in winter, the decrease in WS had the most influence on the increase in HI (Figure 5(a)). In the growing season May–October, the large negative effect of a decrease in Pre overwhelmed the positive effects of decreases in WS and SSH, which resulted in the slightly decreasing trend of HI in Huaibei (Figure 5(a)). Also in the growing season, Pre showed an increasing trend and was dominant in the increase in HI in southern Sunan (Figure 5(c)). For Jianghuai and the entire province, decreases in WS and SSH (especially in WS) were the main reasons for the increase in HI in the growing season (Figures 5(b) and 5(d)). At the annual scale, the effects of SHH, WS, and Pre offset each other and resulted in little change in HI in Huaibei (Figure 5(a)). Similar to the growing season, the increase in Pre dominated the positive trend of HI in Sunan (Figure 5(c)), and the decreases in WS and SSH together were mainly responsible for the increase in HI in both Jianghuai and the entire province (Figures 5(b) and 5(d)).

Decreases in Pre and RH in Jiangsu province were generally responsible for the negative trends of HI in spring and autumn; the positive trends of HI in summer and winter were mainly due to decreases in WS and SSH and increases in Pre (Figure 5). In the growing season and annually in Jianghuai, Sunan, and the entire province, the decreases in WS and SSH and increases in Pre were the main causes of the increase in HI. The slight decreasing trend in HI in the growing season and annually in northern Huaibei was mainly caused by the decrease in Pre which counterbalanced the positive effects of changes in WS and SSH (Figure 5).

5. Discussion

5.1. Dry and Wet Trends in Jiangsu Province. Similar to some previous studies in China [6, 21, 32, 33, 48, 64, 65]. We found that the climate of Jiangsu province also exhibited different dry-wet trends in different seasons and annually during 1960–2019 (Figure 3). We found a dry trend in Jiangsu

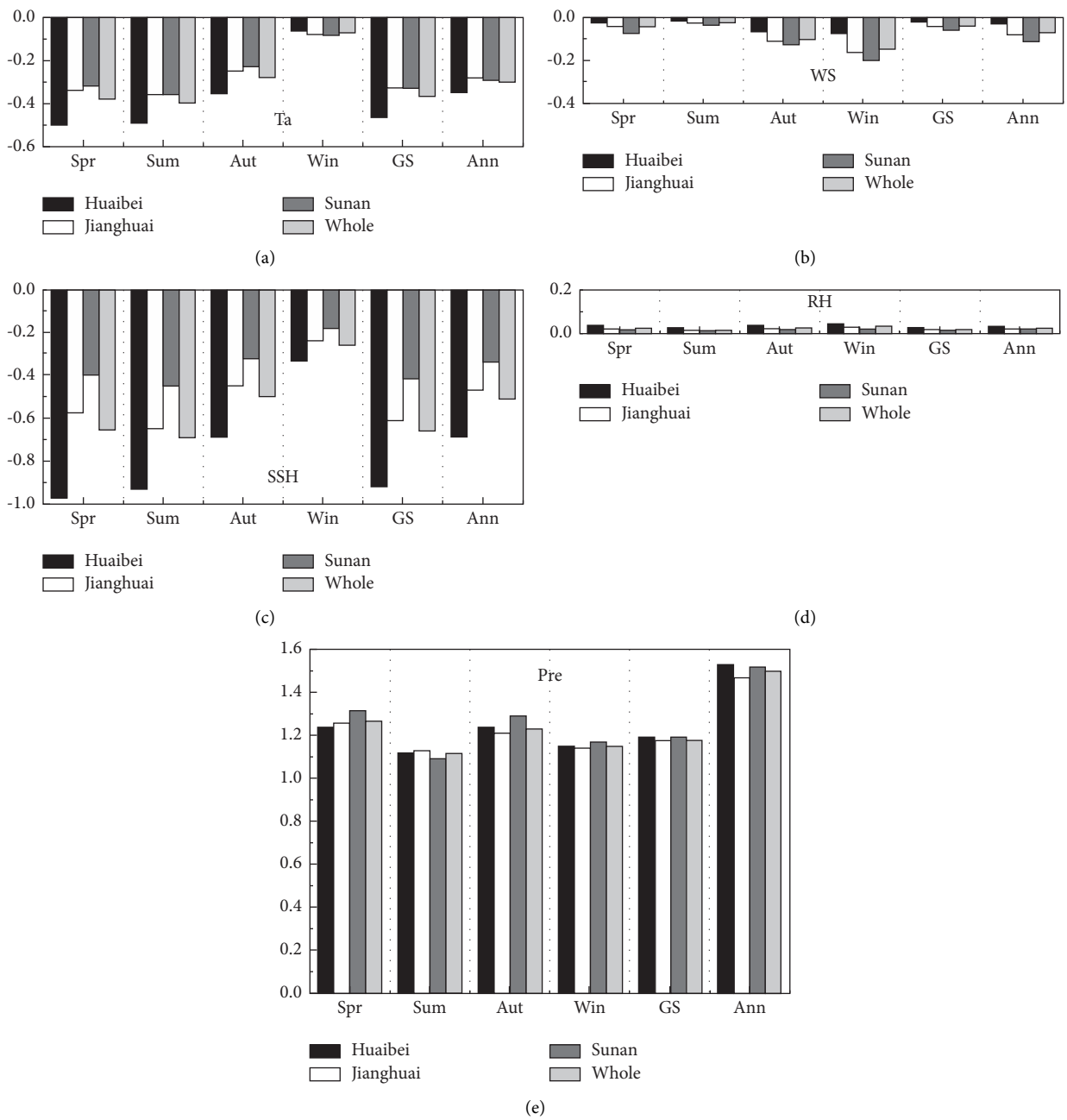


Figure 4: Sensitivity coefficients of seasonal and annual humidity index (HI) to five principal meteorological parameters in each region and the entire province; T_a, WS, SSH, RH, and Pre are mean daily air temperature, wind speed, sunshine hours, relative humidity, and precipitation, respectively.

Province in spring and autumn in the period 1960–2019 (Figures 3(a) and 3(c)). This was consistent with findings for some regions within China, such as the Zoige Wetland in southwestern China for 1961–2016 [21], Zhejiang province in southeastern China for 1961–2016 [33], Bosten Lake basin, and the Loess Plateau in northern China for 1980–2016 and 1961–2014 [15, 22]. Wang et al. [6] investigated dry-wet trends in China and found a drought tendency in the semi-arid or semi-humid and humid areas of southwestern China for a recent 58-year period. Wu et al. [7] found that the overall trend in China had changed from wet to dry since 1954. Worldwide, a drying climate has been observed in many countries, such as Finland [66], South Africa [67], Spain [17], India [27], Iraq [20], Iran [16], Rwanda [26], and Middle East and adjacent areas [28]. Except for spring and autumn, we detected a wet trend in Jiangsu province during 1960–2019 for the other seasons and annually (Figure 3). In reviewing previous studies, we noticed that areas with an increasingly wet climate were distributed mainly in western and northern China: Qinghai province [30], north Xinjiang [31], the Yellow River basin [49], and the Tibetan Plateau [15, 68]. Researchers have also

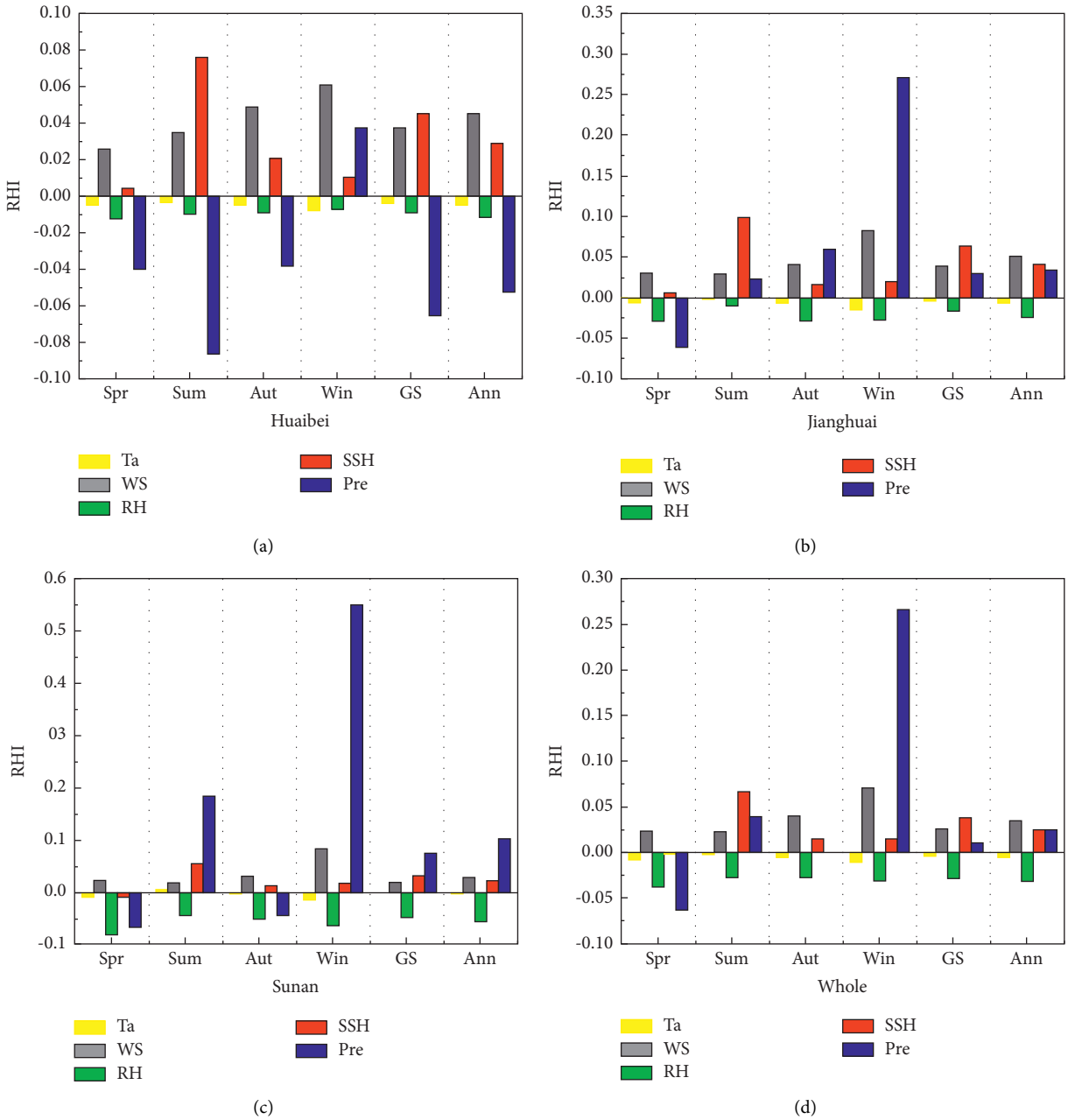


Figure 5: Contributions of five meteorological variables to humidity index (HI) trends at both seasonal and annual scales in (a) Huaibei, (b) Jianghuai, (c) Sunan, and (d) the entire province; T_a, WS, SSH, RH, and Pre are daily mean air temperature, wind speed, sunshine hours, relative humidity, and precipitation, respectively.

found an increasingly wet trend since 1970 for northwestern China and northern China west of longitude 100oE [34, 64]. However, the trend for an increasingly wet climate was found only for August in Zhejiang province in southeastern China for the period 1971–2015 and in summer, autumn, and annually for the humid regions of southern China for the period 1960–2017 [6, 33].

5.2. Contributions of Meteorological Variables to Humidity Index (HI) Trends. This study differs from most previous studies of trends of the humidity index, which considered only the effects of Pre, ET_o, and T_a [6, 7, 15, 20, 27–29, 48, 68]. We examined the effects of five meteorological variables, WS, RH, SSH, T_a, and Pre (Figure 5). We found that HI trends for each region and the entire province were primarily affected by changes in Pre in most seasons and annually during 1960–2019 (Figure 5). This result is consistent with the results of most previous studies [6, 7, 15, 20, 27, 32, 48]. However, for all three regions and the entire province, the decrease in RH had a negative effect on both seasonal and annual HI and was predominant in decreasing HI in spring for Sunan (Figure 5(c)). Changes in RH have been found to affect the humidity index in the Yellow River basin in

northern China [49] and the Zoige Wetland in southwestern China [21]. We also analyzed the effects of SSH and WS, which were included in only a few previous studies, and found that both contributed positively to the HI trend in Jiangsu province (Figure 5). Our previous study found that, for Jiangsu province in 1960–2019, the decrease in RH and SSH decreased ET_o [69], which increased HI.

5.3. Sensitivity of Humidity Index (HI) to Meteorological Variables. The HI was most sensitive to Pre in Jiangsu province for the period 1960–2019 (Figure 4(e)). This is consistent with results for both the entire northwestern and southwestern parts of China [34, 65]. Wang et al. [6] also found that dry-wet conditions were most sensitive to Pre in the humid regions of China that had annual precipitation >800 mm. Our results differ from previous studies, which found HI was secondarily sensitive to vapor pressure or ET_o [6, 34], as we found that in ranking parameters to which HI was sensitive, SSH ranked second (Figure 4(c)). This difference may be due to the exclusion of SSH from sensitivity analysis in previous studies [6, 34]. However, measured SSH indicated that a limited amount of energy was available at the surface [40, 51, 70] and it was therefore necessary to include it in the sensitivity analysis. The sensitivity of HI to T_a and WS was much less than to Pre and SSH, with absolute values of sensitivity coefficients in Jiangsu province all <0.5 (Figures 4(a), 4(b), and 4(d)). HI in the Gansu province of northwestern China showed the least sensitivity to WS and to minimum air temperature [63]. Liu et al. [34] also found that HI for northwest China in 1960–2010 was least sensitive to T_a and WS. Studies that include sensitivity analysis have mostly concentrated on arid areas within northern China, and further similar work must be conducted for humid southern China, which has experienced a dry trend [6, 32, 33].

5.4. Uncertainties and Future Research. In this research, we adopted the FAO 56 PM method to estimate ET_o and many mathematical methods to determine the dominant meteorological variable. However, the selected five meteorological variables were not totally independent, and the influences between them would impact the determination of dominant factors [34]. Another uncertainty was caused by these mathematical methods without considering the physical mechanism. These uncertainties forced researchers to search a hydrometeorological model, which can clarify the interaction between different meteorological variables with the help of underlying climatic mechanisms. Moreover, our research was conducted in Jiangsu province on the southeast coast of China. This is helpful to explore the dry-wet trends for the whole country by combining with the existing research in arid areas of northwest China. In addition to this, future research is still needed to focus on the potential impacts of dry-wet changes on human society.

6. Conclusion

We investigated the contributions of five meteorological variables (T_a, WS, RH, SSH, and Pre) to both seasonal and annual HI trends for three regions of Jiangsu province and the province as a whole in humid southeastern China for the 60-year period 1960–2019. The results showed the following: (1) the decreased Pre and RH were generally responsible for the negative trends of HI in spring and autumn. (2) The wet trend in summer and winter was mainly caused by decreases in WS and SSH and increases in Pre (Figure 5). (3) For the growing season and annually, the decreased WS and SSH and increased Pre were the main causes for positive HI in Jianghuai, Sunan, and the entire province. The slight decreasing HI in the growing season and annually in northern Huaibei was mainly caused by the decrease in Pre (Figure 5). Sensitivity analysis indicated that the HI in Jiangsu province was positively sensitive to Pre and RH and negatively sensitive to T_a, WS, and SSH over the 60-year period (Figure 4). The climate of Jiangsu province was most sensitive to changes in Pre and then, in the decreasing order of sensitivity, to SSH, T_a, WS, and RH in all seasons (except for winter) and annually. For winter, sensitivity was ranked Pre > SSH > WS > T_a > RH (Figure 4).

This study has shown that the regional climate has experienced a dry trend in some seasons during the period 1960–2019 despite the humidity in southeastern China. This will draw attention to the evolution of dry-wet conditions in humid southeastern China rather than the arid regions in western China. Furthermore, in addition to the Pre that was directly related to dry-wet conditions, the other climatic factors also impacted the dry-wet conditions and need to be considered in humid regions. As an important agricultural area in China, the observed dry trend may threaten existing irrigation systems and the security of water resources, which in turn affects crop yield in Jiangsu province. The contribution of each meteorological variable to HI trends, and the sensitivity of HI to each variable, can provide a basis for structural adjustment of farming systems in Jiangsu province at different spatiotemporal scales. Further work should extend this research to the entire humid area of southwestern China and clarify the mechanisms that lead to different trends of humidity indexes in different regions or over different seasons.

Data Availability

The raw data are shared in the China Meteorological Data Service Center: http://data.cma.cn/.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

This study was financially supported by the National Natural Science Foundation of China (Grant no. 42075051), Drought Meteorological Science Research Fund (Grant no. IAM202108), and Jiangsu Meteorological Bureau Science Foundation for Young Scholars (nos. KQ202126 and KQ202226).

References

[1] Y. Wang, Y. Xu, C. Lei et al., "Spatio-temporal characteristics of precipitation and dryness/wetness in Yangtze River Delta, eastern China, during 1960–2012," *Atmospheric Research*, vol. 172-173, pp. 196–205, 2016.

[2] N. Nicholls and L. Alexander, "Has the climate become more variable or extreme? Progress 1992-2006," *Progress in Physical Geography: Earth and Environment*, vol. 31, no. 1, pp. 77–87, 2007.

[3] M. A. Asadi Zarch, B. Sivakumar, and A. Sharma, "Droughts in a warming climate: a global assessment of standardized precipitation index (SPI) and reconnais-sance drought index (RDI)," *Journal of Hydrology*, vol. 526, pp. 183–195, 2015.

[4] S. M. Quiring and T. N. Papakryiakou, "An evaluation of agricultural drought indices for the Canadian prairies," *Agricultural and Forest Meteorology*, vol. 118, no. 1–2, pp. 49–62, 2003.

[5] A. K. Mishra and V. P. Singh, "Drought modeling—a review," *Journal of Hydrology*, vol. 403, no. 1–2, pp. 157–175, 2011.

[6] S. Wang, Q. Zhang, P. Yue, and J. S. Wang, "Effects of evapotranspiration and precipitation on dryness/wetness changes in China," *Theoretical and Applied Climatology*, vol. 142, no. 3-4, pp. 1027–1038, 2020.

[7] S. Wu and X. D. Yan, "Variations in droughts and wet spells and their influences in China: 1924–2013," *Theoretical and Applied Climatology*, vol. 135, no. 1-2, pp. 623–631, 2018.

[8] W. C. Palmer, *Meteorological Drought*, US Department of Commerce. US Weather Bureau, Washington, DC, USA, 1965.

[9] T. B. Mckee, N. J. Doesken, and J. Kleist, "The relationship of drought frequency and duration to time scales,"vol. 17, pp. 179–183, in *Proceedings of the 8th Conference on Applied Climatology*, vol. 17, pp. 179–183, American Meteorological Society, Boston, MA, USA, 1993.

[10] C. W. Thornthwaite, "An approach toward a rational classification of climate," *Geographical Review*, vol. 38, no. 1, pp. 55–94, 1948.

[11] S. M. Vicente-Serrano, S. Beguería, J. I. López-Moreno, M. Angulo, and A. El Kenawy, "A new global 0.5° gridded dataset (1901–2006) of a multiscalar drought index: comparison with current drought index datasets based on the palmer drought severity index," *Journal of Hydrometeorology*, vol. 11, no. 4, pp. 1033–1043, 2010.

[12] S. M. Vicente-Serrano, S. Beguería, J. I. Lorenzo-Lacruz et al., "Performance of drought indices for ecological, agricultural, and hydrological applications," *Earth Interactions*, vol. 16, no. 10, pp. 1–27, 2012.

[13] Q. Yang, M. X. Li, Z. Y. Zheng, and Z. G. Ma, "Regional applicability of seven meteorological drought indices in China," *Science China Earth Sciences*, vol. 47, no. 3, pp. 337–353, 2017.

[14] Z. G. Ma and C. B. Fu, "The basic facts of northern China from 1951 to 2004 drought," *Chinese Science Bulletin*, vol. 51, no. 20, pp. 2429–2439, 2006.

[15] M. Wen, D. Cheng, J. Song, G. Zhang, W. Lai, and W. Jiang, "Impacts of climate change on aridity index and its spatio-temporal variation in the Loess Plateau of China, from 1961 to 2014," *Environmental Earth Sciences*, vol. 77, no. 4, p. 137, 2018.

[16] H. Tabari and M. B. Aghajanloo, "Temporal pattern of aridity index in Iran with considering precipitation and evapotranspiration trends," *International Journal of Climatology*, vol. 33, no. 2, pp. 396–409, 2012.

[17] F. J. Moral, F. J. Rebollo, L. L. Paniagua, A. García-Martín, and F. Honorio, "Spatial distribution and comparison of aridity indices in extremadura, southwestern Spain," *Theoretical and Applied Climatology*, vol. 126, no. 3-4, pp. 801–814, 2016.

[18] Y. Wand, P. X. Liu, L. G. Cao, Y. Gao, and G. Z. Yong, "Characteristics of southwestern China dry-wet condition based on wetness index in 1960-2011," *Natural Resources Journal*, vol. 5, pp. 830–838, 2014.

[19] Y. F. Liu, Z. H. Yuan, J. M. Feng, L. Zhang, and L. Wu, "Dry-wet condition of shaanxi province in recent 56 Years based on surface humidity index," *Arid Land Geography*, vol. 39, no. 6, pp. 1186–1196, 2016.

[20] N. Şarlak and O. M. A. Mahmood Agha, "Spatial and temporal variations of aridity indices in Iraq," *Theoretical and Applied Climatology*, vol. 133, no. 1-2, pp. 89–99, 2018.

[21] X. Jin, H. Qiang, L. Zhao et al., "Spei-based analysis of spatio-temporal variation characteristics for annual and seasonal drought in the Zoige Wetland, southwest China from 1961–2016," *Theoretical and Applied Climatology*, vol. 139, no. 1-2, pp. 711–725, 2020.

[22] H. Yang, J. Xu, Y. Chen et al., "Has the Bosten Lake basin been dry or wet during the climate transition in northwest China in the past 30 years?" *Theoretical and Applied Climatology*, vol. 141, no. 1-2, pp. 627–644, 2020.

[23] Z. Zhang, Q. Zhang, X. Chen, J. Zhang, and J. Zhou, "Statistical properties of moisture transport in east asia and their impacts on wetness/dryness variations in north China," *Theoretical and Applied Climatology*, vol. 104, no. 3-4, pp. 337–347, 2011.

[24] S. K. MIn, W. T. Kwon, E. H. Park, and Y. Choi, "Spatial and temporal comparisons of droughts over Korea with east asia," *International Journal of Climatology*, vol. 23, no. 2, pp. 223–233, 2003.

[25] A. D. Mueller, G. A. Islebe, M. B. Hillesheim et al., "Climate drying and associated forest decline in the lowlands of northern Guatemala during the late holocene," *Quaternary Research*, vol. 71, no. 2, pp. 133–141, 2009.

[26] I. Muhire and F. Ahmed, "Spatiotemporal trends in mean temperatures and aridity index over Rwanda," *Theoretical and Applied Climatology*, vol. 123, no. 1-2, pp. 399–414, 2015.

[27] M. V. S. Ramarao, J. Sanjay, R. Krishnan, M. Mujumdar, A. Bazaz, and A. Revi, "On observed aridity changes over the semiarid regions of India in a warming climate," *Theoretical and Applied Climatology*, vol. 136, no. 1-2, pp. 693–702, 2018.

[28] H. Sahour, M. Vazifedan, and F. Alshehri, "Aridity trends in the Middle East and adjacent areas," *Theoretical and Applied Climatology*, vol. 142, no. 3-4, pp. 1039–1054, 2020.

[29] L. L. Paniagua, A. García-Martín, F. J. Moral, and F. J. Rebollo, "Aridity in the iberian Peninsula (1960–2017): distribution, tendencies, and changes," *Theoretical and Applied Climatology*, vol. 138, no. 1-2, pp. 811–830, 2019.

[30] Z. Liu, P. Zhou, F. Zhang, X. Liu, and G. Chen, "Spatio-temporal characteristics of dryness/wetness conditions across Qinghai province, northwest China," *Agricultural and Forest Meteorology*, vol. 182–183, pp. 101–108, 2013.

[31] Y. F. Wu, G. X. Zhang, H. Shen, Y. J. Xu, and B. Batur, "Attribute analysis of aridity variability in north Xinjiang, China," *Advances in Meteorology*, vol. 2016, Article ID 9610960, 11 pages, 2016.

[32] L. Liu, Y. Wang, N. You, Z. Liang, D. H. Qin, and S. Li, "Changes in aridity and its driving factors in China during 1961–2016," *International Journal of Climatology*, vol. 39, no. 1, pp. 50–60, 2019.

[33] W. Lou, L. Wu, Y. Mao, and K. Sun, "Precipitation and temperature trends and dryness/wetness pattern during 1971–2015 in Zhejiang province, southeastern China," *Theoretical and Applied Climatology*, vol. 133, no. 1-2, pp. 47–57, 2018.

[34] X. Liu, D. Zhang, Y. Luo, and C. Liu, "Spatial and temporal changes in aridity index in northwest China: 1960–2010," *Theoretical and Applied Climatology*, vol. 112, no. 1-2, pp. 307–316, 2013.

[35] L. Hao, X. Huang, M. S. Qin, Y. Q. Liu, W. H. Li, and G. Sun, "Ecohydrological processes explain urban dry island effects in a wet region, southern China," *Water Resources Research*, vol. 54, no. 9, pp. 6757–6771, 2018.

[36] C. Huang, M. Zhang, J. Zou et al., "Changes in land use, climate and the environment during a period of rapid economic development in Jiangsu province, China," *Science of the Total Environment*, vol. 536, pp. 173–181, 2015.

[37] M. Mai, S. P. Zha, and B. Zhu, "Agro-climatic resource zoning in Jiangsu province," *Meteorological and Environmental Research*, vol. 1, no. 11, pp. 59–63, 2010.

[38] L. Hao, G. Sun, Y. Q. Liu et al., "Urbanization dramatically altered the water balances of a paddy field-dominated basin in southern China," *Hydrology and Earth System Sciences*, vol. 19, no. 7, pp. 3319–3331, 2015.

[39] N. Pye, N. J. Middleton, and D. S. G. Thomas, "World atlas of desertification," *The Geographical Journal*, vol. 160, no. 2, p. 210, 1994.

[40] R. Allen, L. Pereira, D. Raes, M. Smith, and R. G. Allen, *Crop Evapotranspiration: Guidelines for Computing Crop Water Requirements*, Food and Agriculture Organization, Rome, Italy, 1998.

[41] Z. Wu, N. E. Huang, S. R. Long, and C. K. Peng, "On the trend, detrending, and variability of nonlinear and nonstationary time series," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 38, pp. 14889–14894, 2007.

[42] H. B. Mann, "Nonparametric tests against trend," *Econometrica*, vol. 13, no. 3, pp. 245–259, 1945.

[43] M. G. Kendall, "A new measure of rank correlation," *Biometrika*, vol. 30, no. 1-2, pp. 81–93, 1938.

[44] P. K. Sen, "Estimates of the regression coefficient based on kendall's tau," *Journal of the American Statistical Association*, vol. 63, no. 324, pp. 1379–1389, 1968.

[45] R. M. Hirsch, J. R. Slack, and R. A. Smith, "Techniques of trend analysis for monthly water quality data," *Water Resources Research*, vol. 18, no. 1, pp. 107–121, 1982.

[46] M. Nouri and M. Bannayan, "Spatiotemporal changes in aridity index and reference evapotranspiration over semi-arid and humid regions of Iran: trend, cause, and sensitivity analyses," *Theoretical and Applied Climatology*, vol. 136, no. 3-4, pp. 1073–1084, 2018.

[47] H. Ahani, M. Kherad, M. R. Kousari, L. van Roosmalen, R. Aryanfar, and S. M. Hosseini, "Non-parametric trend analysis of the aridity index for three large arid and semi-arid basins in Iran," *Theoretical and Applied Climatology*, vol. 112, no. 3-4, pp. 553–564, 2013.

[48] Y. Y. Zhou and C. X. Lu, "Drought/wetting variations in a semiarid and sub-humid region of China," *Theoretical and Applied Climatology*, vol. 140, no. 3-4, pp. 1537–1548, 2020.

[49] K. X. Zhang, S. M. Pan, W. Zhang et al., "Influence of climate change on reference evapotranspiration and aridity index and their temporal-spatial variations in the Yellow River basin, China, from 1961–2012," *Quaternary International*, vol. 380-381, pp. 75–82, 2015.

[50] C. Y. Xu, L. B. Gong, T. Jiang, D. L. Chen, and V. P. Singh, "Analysis of spatial distribution and temporal trend of reference evapotranspiration and Pan evaporation in changjiang (Yangtze River) catchment," *Journal of Hydrology*, vol. 327, no. 1–2, pp. 81–93, 2006.

[51] M. Qin, L. Hao, L. Sun, Y. Q. Liu, and G. Sun, "Climatic controls on watershed reference evapotranspiration varied during 1961–2012 in southern China," *JAWRA Journal of the American Water Resources Association*, vol. 55, no. 1, pp. 189–208, 2019.

[52] Q. Liu, Z. Yang, B. Cui, and T. Sun, "The temporal trends of reference evapotranspiration and its sensitivity to key meteorological variables in the Yellow River basin, China," *Hydrological Processes*, vol. 24, no. 15, pp. 2171–2181, 2010.

[53] Z. L. Huo, X. Q. Dai, S. Y. Feng, S. Z. Kang, and G. H. Huang, "Effect of climate change on reference evapotranspiration and aridity index in arid region of China," *Journal of Hydrology*, vol. 492, pp. 24–34, 2013.

[54] Z. J. Li, Z. Li, Z. Xu, and X. Zhou, "Temporal variations of reference evapotranspiration in heihe River basin of China," *Hydrology Research*, vol. 44, no. 5, pp. 904–916, 2013.

[55] M. Li, R. H. Chu, A. R. M. T. Islam, Y. Jiang, and S. S. Shen, "Attribution analysis of long-term trends of aridity index in the Huai River basin, eastern China," *Sustainability*, vol. 12, no. 5, p. 1743, 2020.

[56] R. H. Mccuen, "A sensitivity and error analysis of procedures used for estimating evaporation," *Journal of the American Water Resources Association*, vol. 10, no. 3, pp. 486–497, 1974.

[57] Y. H. Yin, S. H. Wu, and E. F. Dai, "Determining factors in potential evapotranspiration changes over China in the period 1971–2008," *Chinese Science Bulletin*, vol. 55, no. 29, pp. 3329–3337, 2010.

[58] L. Q. Liang, L. J. Li, and Q. Liu, "Temporal variation of reference evapotranspiration during 1961-2005 in the taoer River basin of northeast China," *Agricultural and Forest Meteorology*, vol. 150, no. 2, pp. 298–306, 2010.

[59] D. Zhang, X. M. Liu, and H. Y. Hong, "Assessing the effect of climate change on reference evapotranspiration in China," *Stochastic Environmental Research and Risk Assessment*, vol. 27, no. 8, pp. 1871–1881, 2013.

[60] Z. L. Wang, P. W. Xie, C. G. Lai et al., "Spatiotemporal variability of reference evapotranspiration and contributing climatic factors in China during 1961–2013," *Journal of Hydrology*, vol. 544, pp. 97–108, 2017.

[61] J. Du, D. Biao, J. Hu, B. La, and K. S. Zhou, "Responses of climatic change on surface humid index in northern tibet during the period of 1961–2006," *Acta Ecologica Sinca*, vol. 05, pp. 2437–2444, 2009.

[62] Y. Li, A. Feng, W. Liu, X. Ma, and G. Dong, "Variation of aridity index and the role of climate variables in the southwest China," *Water*, vol. 9, no. 10, p. 743, 2017.

[63] S. P. Wang, Y. H. Li, J. Y. Feng, J. S. Wang, and J. Wang, "Changes and driving factor of surface wetness in Gansu, China from 1961 to 2012," *Journal of Desert Research*, vol. 34, no. 6, pp. 1624–1632, 2014.

[64] K. Zhang, X. Qian, P. Liu et al., "Variation characteristics and influences of climate factors on aridity index and its association with ao and enso in northern China from 1961 to 2012," *Theoretical and Applied Climatology*, vol. 130, no. 1-2, pp. 523–533, 2017.

[65] Y. Zhao, X. Zou, L. Cao, Y. Yao, and G. Fu, "Spatiotemporal variations of potential evapotranspiration and aridity index in relation to influencing factors over southwest China during

1960–2013," *Theoretical and Applied Climatology*, vol. 133, no. 3-4, pp. 711–726, 2018.

[66] M. Irannezhad, A. Torabi Haghighi, D. Chen, and B. Kløve, "Variability in dryness and wetness in Central Finland and the role of teleconnection patterns," *Theoretical and Applied Climatology*, vol. 122, no. 3-4, pp. 471–486, 2015.

[67] M. Mbiriri, G. Mukwada, and D. Manatsa, "Spatiotemporal characteristics of severe dry and wet conditions in the free state province, South Africa," *Theoretical and Applied Climatology*, vol. 135, no. 1-2, pp. 693–706, 2018.

[68] Y. Zhao, X. Zou, J. Zhang et al., "Spatio-temporal variation of reference evapotranspiration and aridity index in the Loess Plateau region of China, during 1961–2012," *Quaternary International*, vol. 349, pp. 196–206, 2014.

[69] M. S. Qin, Y. Zhang, S. Q. Wan, Y. Yue, Y. Cheng, and B. Zhang, "Impact of climate change on "evaporation paradox" in province of Jiangsu in southeastern China," *PLoS One*, vol. 16, no. 2, Article ID e0247278, 2021.

[70] Z. D. Gao, J. S. He, K. B. Dong, and X. Li, "Trends in reference evapotranspiration and their causative factors in the west liao River basin, China," *Agricultural and Forest Meteorology*, vol. 232, pp. 106–117, 2017.

Hindawi
Advances in Meteorology
Volume 2022, Article ID 3594641, 16 pages
https://doi.org/10.1155/2022/3594641

Research Article

Hydroclimatic Variability, Characterization, and Long Term Spacio-Temporal Trend Analysis of the Ghba River Subbasin, Ethiopia

Mehari Gebreyohannes Hiben,[1,2] **Admasu Gebeyehu Awoke**,[1] **and Abraha Adugna Ashenafi**[3]

[1]*School of Civil and Environmental Engineering, Addis Ababa University, Addis Ababa Institute of Technology (AAiT), Addis Ababa, Ethiopia*
[2]*MG Water Resources Consultancy Firm, Mekelle, Tigray, Ethiopia*
[3]*The Ministry of Water Irrigation and Energy (MoWE), Addis Ababa, Ethiopia*

Correspondence should be addressed to Mehari Gebreyohannes Hiben; hiben123@gmail.com

Received 22 May 2022; Revised 13 July 2022; Accepted 16 July 2022; Published 23 August 2022

Academic Editor: Upaka Rathnayake

Understanding hydroclimatic variability and trend for the past four decades in the Upper Tekeze River basin is significant for future sustainable water resource management as it indicates regime shifts in hydrology. Despite its importance for improved and sustainable water allocation for water supply-demand and food security, varying patterns of streamflow and their association with climate change are not well understood in the basin. The main objective of this study was to characterize, quantify, and validate the variability and trends of hydroclimatic variables in the Upper Tekeze River basin at Ghba subbasin using graphical and statistical methods for homogeneous stations for the time period from 1953 to 2017, not uniform at all stations. The rainfall, temperature, and streamflow trends and their relationships were evaluated using the regression method, Mann–Kendall (MK) test, Spearman's rho (SR) test, Sen's slope, and correlation analysis. The analysis focused on rainfall, temperature, and streamflow collected from 11 climate and six hydrostations. For simplicity to discuss the interannual and temporal variability the stations were categorized into two clusters according to their record length, category 1 (1983–2017) and category 2 (1953–2017). About 73% and 27% of the rainfall stations exhibited normal to moderate annual rainfall variability. The MK and SR test showed that most of the significant trends in annual rainfall were no change except in one station decreasing and the test also showed no significant change in temperature except in three stations showed an increasing trend. Overall, streamflow trends and change point timings were found to be consistent among the stations and all have shown a decreasing trend. Changes in streamflow without significant change in rainfall suggest factors other than rainfall drive the change. Most likely the observed changes in streamflow regimes could be due to changes in catchment characteristics of the subbasin. These research results offer critical signals on the characteristics, variability and trend of rainfall, temperature, and streamflow necessary to design improved and sustainable water allocation strategies.

1. Introduction

Climate change and human intervention combined with other driving factors have brought an apprehension in rainfall and streamflow variability over a basin on water resources management [1, 2]. Research by [3–9] identified how climate change mirrored in altering rainfall patterns and affecting hydrological regimes over river basins. Hence, understanding climate change and other driving forces on hydrological variability is important for improved water management and allocation.

Examining the streamflow and rainfall records and identifying their linkages and trend between them is fundamental to recognizing the impact of climate change

on hydrology and, enabling us to look for climate-resilient water management techniques. Such studies have received increased attention from the scientific community over the last two decades [10–12] due to the rising need to secure water for human uses such as domestic, nondomestic, and irrigation, as well as for environmental flow. In addition, rising concern about climate change and its impacts on streamflow has been an important driver of such studies [13, 14]. Several studies have made known that rainfall is the main atmospheric factor that is directly related to the streamflow patterns [15–18]. Paudel and Acharya [19] studied trends for 40 years of climate data of the Chitwan River Basin, Nepal, in annual and seasonal records. They found a decreasing trend in streamflow, which was strongly correlated to precipitation. They also tried to check changes in streamflow with land use changes (forest cover) but proved that changes in plant cover had an insignificant impact on the streamflow. Studies by Pellicciotti et al [20], Zhang et al. [21], Gebert and Krug [22], Birsan et al. [6, 23], and Setegn et al. [24] showed that changes in rainfall are not enough to describe the trends in the streamflow. Studies by [23, 25] showed streamflow trends in the mountain basins, as in the current study of this chapter, are the most vulnerable environments in the context of climate change.

Remarkable annual fluctuation of precipitation over the Blue Nile and Atbara basins resulted in a reduction in river flows between 1945 and 1984 [26]. On the contrary, recent studies agreed that rainfall over the Upper Blue Nile basin has had no significant trend for the last 45 years [27–31]. In the current study, anxieties around water resources and climate change have gained global significance in the study on hydroclimatic trends [32, 33]. By considering only one climate station it was tried to study the pattern of rainfall over the upper part of the Tekeze River basin [6, 34, 35]. Their output showed that the amount of rainfall remained constant for the past 40 years (1962–2002). Gebremicael et al. [6] studied that rainfall over the Nile basin headwaters has remained constant and a significant decreasing pattern of streamflow in the eastern and northern part of the basin was analyzed in the last four decades. Even with the reputation of streamflow to ensure sustainable water resource allocation and food security in the study area, long-term trends and change points of flow regimes and the relationship with climate change are not properly studied. Thus, it is vital to comprehend the linkages between rainfall and streamflow trends of the subbasin and establish whether hydrological variability is driven by changes in climate or any other driving forces. The main objective of this research is to identify, quantify and analyze recent trends in streamflow and climate data in the mountainous, semi-arid region of Ethiopia for improved water allocation planning in the subbasin. Hence, the 1st section contains the general introduction and literature review, and the rest of the study is organized as follows: Section 2 contains the methodology (method); Section 3 contains the results; Section 4 contains the conclusions and policy implications.

2. Data and Methods

2.1. Study Area. The Ghba subbasin is located in northern Ethiopia and covers from 38°38′ to 39°48′ Eastern longitudes and 13°14′ to 14°16′ northern latitudes as shown in Figure 1. The total area of the Ghba subbasin is about 5125 km^2 and comprises the Tigray regional state's capital city Mekelle. It forms the headwaters of the Upper Tekeze River basin, one of the major tributaries of the Nile River [36]. The landscape is characterized by highlands and hills in the north and northeastern, and highlands in the central part of the catchment [37]. The central highlands are divided by numerous rivers that flow towards the southwestern part of the subbasin and joins the main Tekeze River at Chemey [38]. As shown in Figure 1, the altitude varies from 3,300 meters above sea level (m.a.s.l.) at Mugulat Mountains near Adigrat town to 930 m.a.s.l. At the subbasin outlet [39]. The mean elevation of the catchment is 2144 m with a standard deviation of 361 m indicating that the topography is very rugged [38].

The Ghba subbasin is categorized under a semi-arid climate region where rainfall occurs generally from July to September and has a long period of the dry season. The high-intensity storm is falling between July and August [36]. The mean annual precipitation of the subbasin is between 450 mm in the eastern part to 850 mm in the northern and western parts of the subbasin [6]. The high variability of rainfall in the subbasin is mainly due to the complex topography nature and the seasonal migration of the intertropical convergence zone (ITCZ) [6, 40]. Hence, more than 85% of the total rainfall falls in the wet season between June to September with maximum effective rainfall of less than or to 60 days and a dry period extending up to 10 months [41]. The alterations are frequently associated with the seasonal relocation of the Intertropical convergence zone (ITCZ).

The LULC of the Ghba subbasin is characterized by severe land degradation through deforestation, overgrazing, and cultivation on the rugged topography. The dominant land use and land cover of the study area for the specified periods were rain-fed agriculture; the main crops included wheat, teff, sorghum, barley, maize, and pulses, followed by shrubs, bare land, wood, grassland, plantation, residential areas, forest, and water. Nevertheless, because of the governmental plan of millennium development the last 10 to 15 years irrigated agriculture of small-scale irrigation schemes have been increased significantly in the eastern, northern, and central parts of the subbasin [42–44].

Climate and hydrological spatiotemporal datasets are required to undergo analysis, quantify, characterize, and validate the variability and trends of hydroclimatic variables in the Ghba subbasin. The results of this study will assist in a better understanding of the system in the subbasin, including patterns, trends, and the extent of the changes in the hydrology regime. Thus, the study will help planners to propose climate-resilient water allocation strategies in the Ghba subbasin.

2.2. Data Preparation. The long-term hydroclimatic data were collected from the Ethiopian National Meteorology Services Agency [45], and the hydrological shape files were

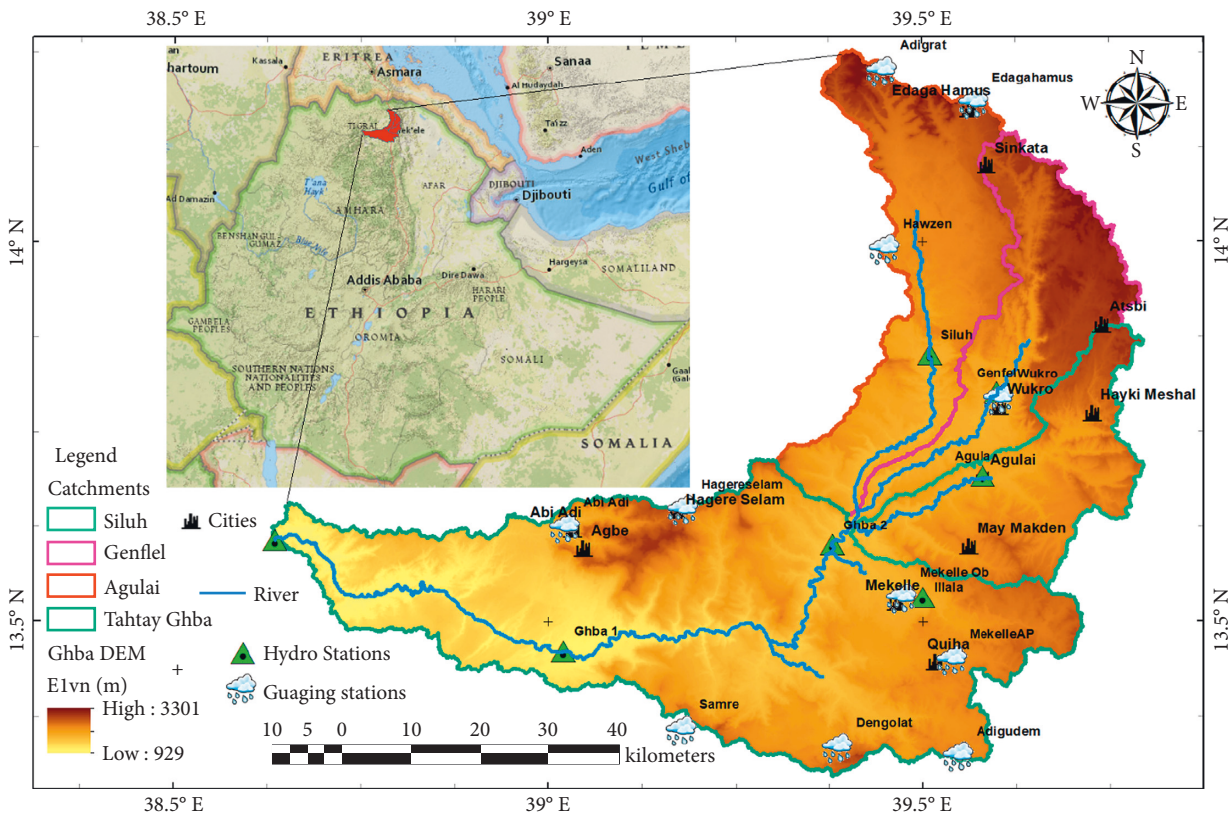


Figure 1: Location of the study area.

obtained from the Ministry of Water, Irrigation, and Energy (MoWE). The current analysis focused on the three hydroclimatic variables: rainfall, temperature, and streamflow of the study area. After screening independent, nonstationary, and homogeneous stations were selected for trend analyses. Furthermore, long-term data availability and the percentage of missing data were considered. The duration selected was more than 25 years and the length of data recorded in all rainfall stations are varying. However, all are more than 30 years, and although the recording of flow data over the subbasin started in the late 1960s, it was discontinued for most of the gauging stations during the civil war in the 1980s. However, the stations in the subbasin have an extended period of more than 25 years of data and these were used in the analyses. The location and general information of hydroclimatic data of the subbasin are shown in Figure 1. In this study, sites with less than 10% of the missing data were selected for both rainfall and streamflow stations. Missing gaps greater than or equal to 5% were completed using missing data estimation methods. Literature such as the studies by [46, 47]were reviewed and noted that the inverse distance weighting (IDW) method is a frequent and commonly used method for estimating missing data in the fields of hydrology. Studies by Teegavarapu and Chandramouli [48] and Teegavarapu et al [49] provided several alternatives to IDW methods. Teegavarapu et al [50] used connotation rules within weighting methods to improve estimates of missing precipitation data. Global estimation methods that use trend and regression analysis are also applicable for spatial interpolation. However, the selection of the appropriate estimation tool stances a major problem due to the large range of candidate functions [51, 52]. In the study area, four estimation methods were selected: arithmetic average (AA), inverse distance weightage (IDW), normal ratio weightage (NRW), and coefficient of correlation weightage (CCW). Besides, these methods were compared with their performance of estimations by introducing uncertainty techniques of sampling, i.e., creating different percentages of missing data (5%, 10%, 15%, 20%) from the available dataset later to be used for calibration purpose to the estimated data. Thus, the results with the minimum RMSE and MAE and highest positive correlation coefficients of the estimator were selected as the best method of the estimator. Hence, the proximity and coefficients of correlation in the IDW and NRW methods improved the rainfall and streamflow estimates and were chosen for filling the missing data where it best fits.

In the current study, annual, monthly, and seasonal daily rainfall and streamflow data were organized. The subbasin was subdivided into four catchments based on their location and physical factors as can be seen in Figure 1. Furthermore, the 11 homogeneous rainfall stations and their valued contribution to each catchment were computed using GIS by the Theisen polygon method as given in Table 1. Siluh catchment covering the northern and northwestern of the subbasin; Genfel catchment covers the central part of the subbasin; Agluai catchment covers the eastern part of the subbasin; and last, Tahtay Ghba covers the southern and

Table 1: Results of rainfall contribution of stations to each catchment by the Theisen polygon method.

S/N	Catchments	Stations										
		AbiAdi	Adigrat	Adigudem	Dengolat	E/Hamus	Hawzen	H/Selam	MekelleAP	Mekelle Ob	Samre	Wukro
1	Siluh		13			17	33	5		5		27
2	Genfel					22	3			5		70
3	Agulai								6	30		64
4	Tahtay Ghba	31		3	10			16	12	14	13	

southwestern part of the subbasin. A time series data with similar time and length was required to evaluate the relation of interannual and temporal variability over different locations [32]. One of the major challenges in the Ghba subbasin was less data availability with similar time and length (varying duration). Thus, the rainfall and streamflow data have different periods with different start and end dates. In order to address this challenge and to get representative stations within the subbasin, stations with a similar period were clustered. Datasets of 35 years were categorized in one group and, datasets from 43 years up to 65 years were categorized in the second group.

Hydroclimatic data quality checks and analyses were performed to characterize, quantify and validate their variability and trend in the subbasin. While dealing with hydroclimatic extremes and trend analysis, it is always affected by data inhomogeneity [53]. In this study, the coefficient of variation (CV), standardized anomaly index (SAI), and graphical methods were used to understand the system. First, the normality test of the annual and seasonal discharge and rainfall time series data was performed. Second, the hydroclimatic variability and trend were tested using different statistical and graphical methods. Thus, the rainfall and streamflow trends and their relationships were evaluated to the selected homogeneous time series of annual and seasonal rainfall, temperature, and discharge using the regression method, Mann–Kendall (MK) test, SR test, Sen's slope estimator test, and correlation analysis to show whether increasing or decreasing trends in the data series. As briefly discussed in Sections 2.3 and 2.4, the method of selecting annual and three seasonal periods, rainy (June–September), dry (October–February), and short rainy (March–May), helps to provide a seasonal comparison of changes in the hydroclimatic variables.

2.3. Variability Analyses. The CV is a statistical measure that expressed the difference of data about the mean; it is the ratio of standard deviation to the mean (equation (1)). In the current study, a CV of <20, ≥20 ≤ 30, and >30 was taken as normal, moderate, and highly variable, respectively. In many studies, the CV (equation (1)) was used to characterize the variability of rainfall [32, 54, 55].

$$CV = \frac{\sigma}{\overline{X}}, \tag{1}$$

where σ and $\overline{X}$ denote the standard deviation and mean of rainfall or streamflow, respectively. The SAI indicates a measure of distance between the data and its mean (equation.(2)). It shows the probability of being of an observed rainfall or streamflow quantity related to the long-term mean rainfall or streamflow. Many studies were undertaken to illustrate SAI (equation (2)) in examining rainfall and streamflow variability [56–58].

$$SAI = \frac{(x - \overline{X})}{\sigma}, \tag{2}$$

where x represents the rainfall, and $\overline{X}$ and σ denote the mean standard deviation of the rainfall or streamflow data.

2.4. Mk Test and Sen's Slope. In this study, the modified Mann–Kendall trend is used [59, 60]. This test is primarily innovated by Mann–Kendall [61, 62]. Mann–Kendall (MK) and Sen's slope tests are helpful for efficient water resource management [63]. Furthermore, these tests were used for the selected homogeneous hydrological and meteorological variables. Trend analysis is an effective method of noticing alterations in climatic and hydrological variables [63–68]. Hence, the Mann–Kendall test has been commonly applied to identify trends in hydroclimatic variables such as reference evapotranspiration, temperature, rainfall, and streamflow time series in different regions of the world [60, 62, 68–71]. In the current study, a modified Mann–Kendall trend test was used to notice the alteration in precipitation, streamflow, and average temperature. Mann–Kendall trend test and Sen's slope were applied to evaluate the trend of rainfall, streamflow, and mean annual temperature in the study area. The Mann–Kendall statistic output of the time series data was analyzed to compare whether the trend of the hydroclimatic variables has been noticed to the critical values to test or not. Input data arrangement must be in time sequential while performing the Mann–Kendall analysis. Thus, the first step is to govern the sign of the difference between consecutive sample results. Sgn $(X_j - X_k)$ is an indicator function that results in the values 1, 0, or −1 referring to the sign of $X_j - X_k$ where $j > k$. A positive value is an indicator of an increasing (upward) trend and a negative value is an indicator of decreasing (downward) trend. This MK statistic (S) is given by the following equations (3) and (4):

$$\left\{ s = \sum_{i=1}^{n=1} \sum_{j=1+1}^{n} \operatorname{sgn}\left(x_j - x_i\right) \text{where} \theta = \left(x_j - x_i\right). \right. \tag{3}$$

$$\operatorname{sgn}(\theta) = \begin{cases} +1 \\ 0 \quad \text{if} \\ -1 \end{cases} \begin{bmatrix} \theta > 0 \\ \theta = 0 \\ \theta < 0 \end{bmatrix}, \tag{4}$$

where x_j and x_i represent the data points in period j and I, while the size of the data series is larger than or equivalent to ten ($n \geq 10$); since $n \geq 10$, the MK test is then categorized by a standard distribution with the mean E (S) = 0 and variance Var (S) is given as

$$V(S) = \frac{1}{18}\left[n(n-1)(2n+5) - \sum_{i=1}^{m} t_i(t_i-1)(2t_i+5)\right], \quad (5)$$

where S and V (S) are Kendall's statistics and variance (equation (6)), respectively.

$$Zs = \begin{cases} \dfrac{s-1}{\sqrt{Var(s)}} & \\ 0 & \text{if} \begin{bmatrix} S>0 \\ S=0 \\ S<0 \end{bmatrix}. \\ \dfrac{s+1}{\sqrt{Var(S)}} & \end{cases} \quad (6)$$

In a Z test, the null hypothesis (H_o) implies no trend, and the alternative hypothesis (H_a) means a significant change in time series. The positive Z value indicates an increasing trend, whereas the negative z value indicates a decreasing trend. A significant trend at 0.1, 0.05, and 0.01 significance levels exist when the $|Z| > 1.645$, $|Z| > 1.96$, and $|Z| > 2.576$, respectively [59, 72]. Thus, in the current study, 90, 95, and 99% confidence levels were applied to analyze the hypothesis.

Sen's slope test [73] calculates the extent of the identified trends. Numerous studies [63, 71, 74, 75] used Sen's slope estimator for trend detection. This test calculates both the slope (i.e., the linear rate of alteration) and intercept. Thus, according to Sen's method of analysis first, a set of linear slopes is calculated as

$$Q_i = \frac{x_j - x_k}{j-k}, \quad \text{for } i = 1 \ldots, N, \quad (7)$$

where x_j are the data values at j and x_k give the data values at k ($j > k$). When there is one data point in each time period, then $N = (n(n-1)/2)$, where n is the number of time periods. However, if there are more data points then, $N < (n(n-1)/2)$, where n is the total number of observations. The values of N are arranged from minimum to maximum values. Then, the median of slope or Sen's slope estimator (Q_{mid}) was computed as

$$Q_m = \begin{cases} \dfrac{Q_{N+1}}{2}, & \text{if } N \text{ odd}, \\ \dfrac{Q_{(N/2)} + Q_{(N+1/2)}}{2}, & \text{if } N \text{ odd}. \end{cases} \quad (8)$$

When the value $Q_{mid} > 0$, it indicates an increasing trend, whereas $Q_{mid} < 0$ represents a decreasing trend.

2.5. Spearman's Rho (SR) Test. The SR test [76, 77] is similar to the MK method and is a nonparametric test. Spearman's rho (SR), the test is a simple method with uniform power for linear and nonlinear trends and is frequently used to confirm the absence of trends [78, 79]. In this test, the null hypothesis (H_0) is that all the data in the time series are independent and identically distributed, while the alternative hypothesis (H_1) is that increasing or decreasing trends exist [80, 81]. The SR test statistic D and the standardized test statistic Z_{SR} is computed as follows:

$$D = 1 - \frac{6\sum_{i=1}^{n}(R_i - i)^2}{n(n^2-1)}, \quad (9)$$

$$Z_{SR} = D\sqrt{\frac{n-2}{1-D^2}}, \quad (10)$$

where R_i is the rank of ith observation X_i in the time series and n is the length of the time series. Positive values of Z_{SR} indicate upward trends, while negative Z_{SR} indicates downward trends in the time series.

3. Results and Discussion

3.1. Monthly and Seasonal Rainfall Characteristics. The mean monthly rainfall of the stations in the Ghba subbasin varied from 1 to 252 mm in the period 1952–2017, as shown in Figure 2. Relatively, Figure 2 shows the monthly rainfall was low from October to February but started to increase in June Besides, comparatively heavy rainfall was received between June and August, with the maximum mean monthly rainfall received in August at the Abi-Adi station. The minimum monthly rainfall was recorded in all stations and the lowest rainfall occurred in the dray season (October through January). As shown in Figure 2, the mean monthly rainfall in the Siluh, Genfel, and Agulai catchments varied almost similar from 2.4 to 196, 1.2 to 195, and 1.2 to 194 mm, respectively. However, the mean monthly rainfall in the Tahtay Ghba catchment varied from 1 mm to 252 mm. Thus, according to the data series in the study area, the minimum rainfall was documented in December and the maximum in August.

3.2. Interannual and Temporal Rainfall Variability. The chart in Figure 3 displays the annual rainfall of the 11 rainfall stations in the Ghba subbasin; the annual rainfall varied from 280 to 1191 mm in the analysis period. In the current study, the Hagere Selam and Wukro stations recorded the highest and lowest annual rainfall, respectively, as shown in Figure 3. For simplicity to discuss the interannual and temporal variability the stations were categorized into two clusters according to their record length.

The chart in Figure 4 displays the annual and seasonal rainfall variabilities of the first categorized group based on the length of observation, i.e., 35 years (1983–2017). In this category, annual rainfall varied from 380 to 1191 mm at Adigudem and Hagereselam stations, respectively. The variability of catagory1 stations showed a similar increasing and decreasing pattern except for a few years with a tremendously high amount of rainfall. Thus, high rainfall amount was recorded at Hagere Selam and Abi-Adi in 1969

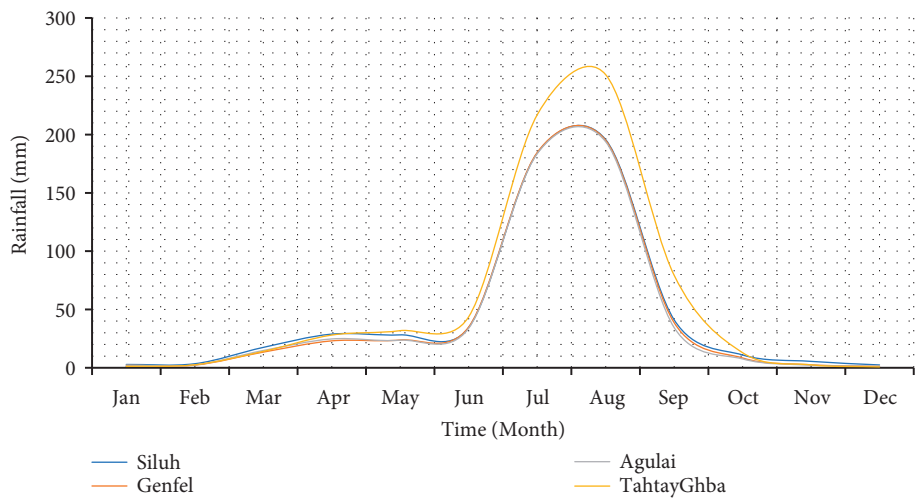


Figure 2: The mean monthly rainfall (mm) at Siluh, Genfel, Agulai, and Tahtay Ghba catchments.

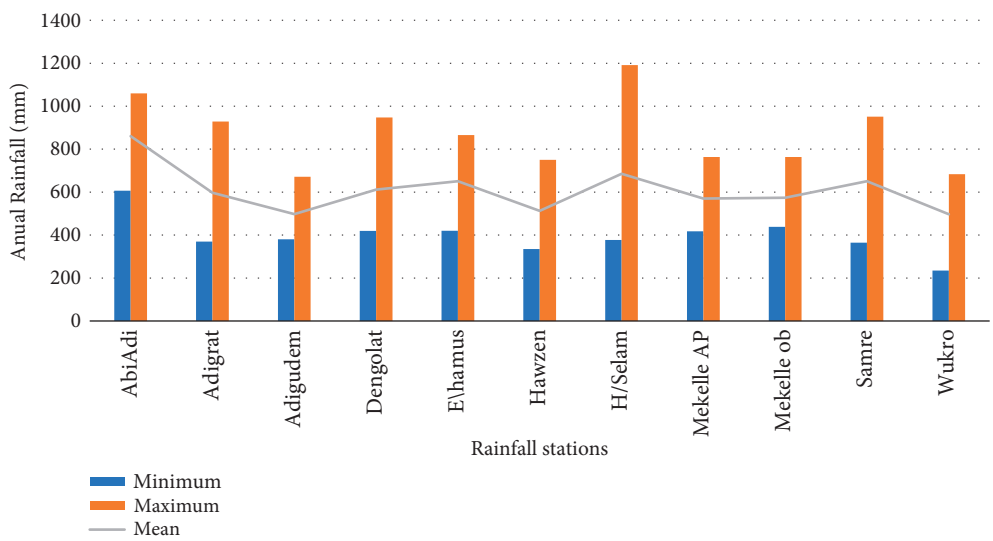


Figure 3: The annual rainfall for the 11 stations of the study area.

and 2016, respectively. In category-one stations; the annual rainfall pattern indicated a similar variability or no significant change as shown in Figure 4.

The chart in Figure 5 displays the annual and seasonal rainfall variability of the second categorized group based on the length of observation, i.e., from 43 up to 65 years, in total (1953–2017). The variability of rainfall in category two is shown in Figure 5; the annual rainfall varied from 280 mm at Wukro to 978 mm at Dongelat, while the annual temporal rainfall variability indicated a similar pattern of rise and fall in a majority of stations. However, there was no significant decreasing pattern in recent years in all stations.

The summary of descriptive statistical analysis for the 11 rainfall stations located in Ghba subbasin is given in Table 2. The annual skewness presented that 64% of the stations were approximately symmetrical while 26% were moderately symmetrical. Besides, all stations indicated a Z-score of skewness and kurtosis closer to zero, telling no significant difference from normality [82, 83]. The general indication of coefficient of variance (CV) is as follows: a CV of <20 has small or no variability, a CV of 20–30% has moderate variability, and a CV of >30% has high variability [84, 85]. In this study, some annual rainfall variability stations were detected in three stations: Hagere Selam, Samre, and Wukro. The seasonal variability was also tested in order to evaluate the full rainfall characteristics that occurred in the period 1952–2017. The two rainfall seasons selected were June–September (rainy season) and March–May (short rainy). The rainy season rainfall showed a slightly similar result as that of the annual rainfall variability. Out of eleven rainfall stations, around 73% of the stations showed the mean annual rainfall had a CV of less than 20%, while 27% of stations indicated a CV between 20 and 26%, as given in Table 2. Hence, this reviled in the annual time series eight stations showed no rainfall variability, and three stations showed a moderate variability. In the rainy seasonal period, except Abi-Adi station all stations are characterized by moderate variability. However, as given in Table 2, the period of dry season rainfall showed that 100% of the stations had a CV > 30 which indicates high variability.

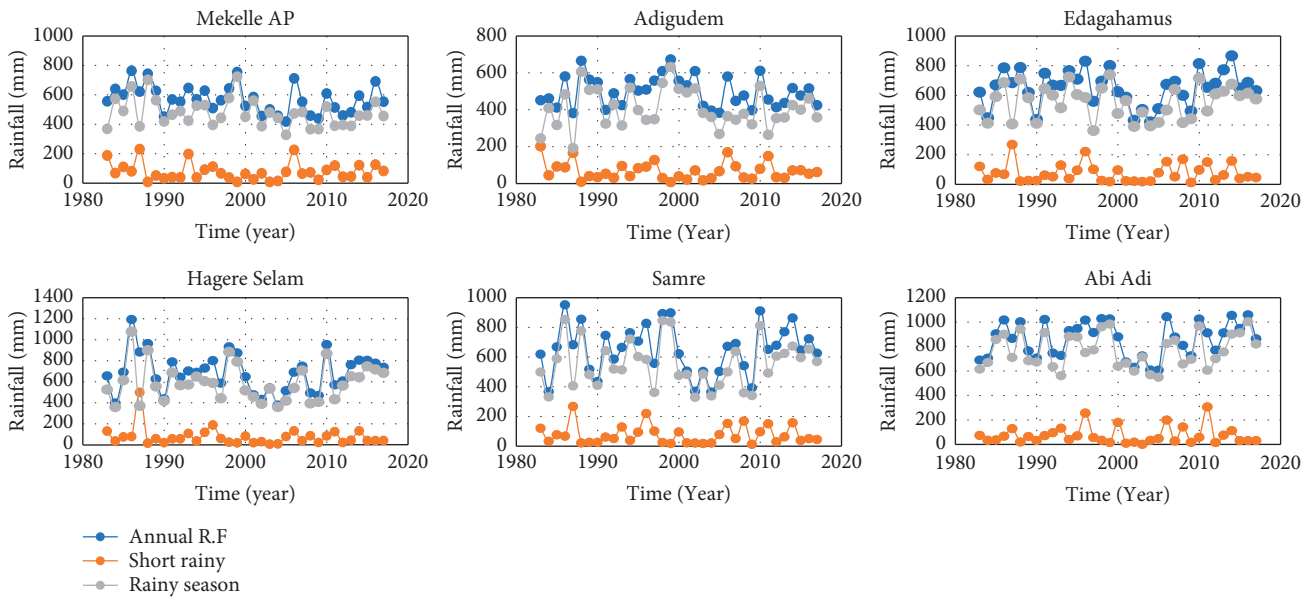


Figure 4: Seasonal temporal variability of rainfall (mm) in stations of 35 years of observation.

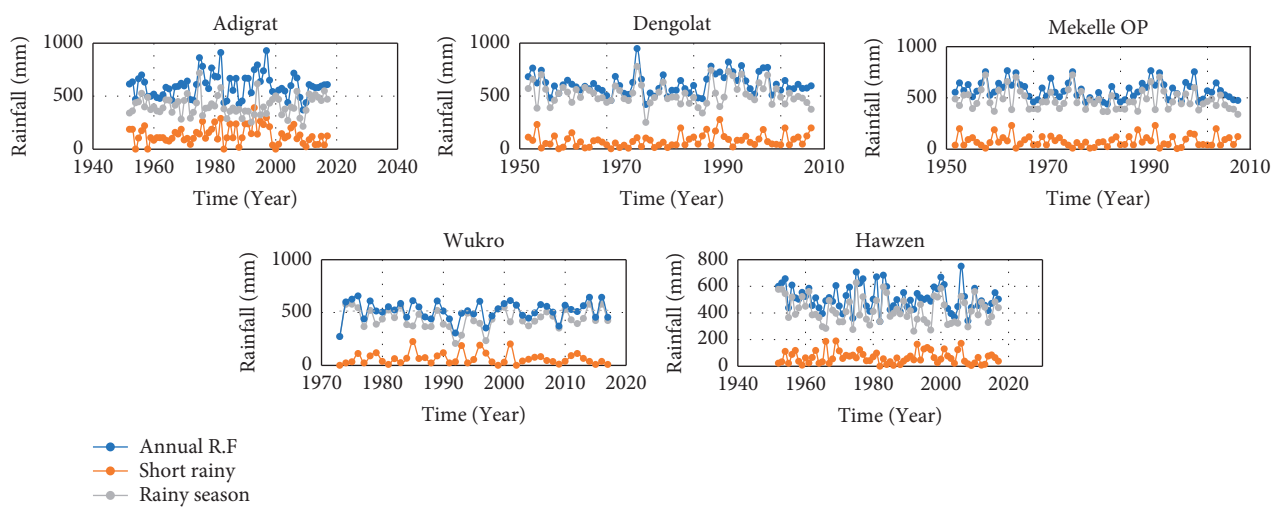


Figure 5: Seasonal temporal variability of rainfall (mm) in stations of 43–65 years of observation.

The chart in Figure 6 displays the SAI of the annual rainfall for sample rainfall stations in the Ghba subbasin. This figure shows that the negative anomaly varied from 42% to 71% of the total number of observations at Edaga Hamus and Mekelle AP, respectively. On seasonal time series, short rainy (MAM) and the rainy season (JJAS) showed a negative anomaly between 45% to 67% and 45% to 65% of the observations, respectively. The short rainy (MAM) rainfall displayed a higher negative anomaly at Abi-Adi, whereas the rainy season (JJAS) showed a higher negative anomaly at Mekelle AP. In some stations (e.g., Hagere Selam and Samre) a negative anomaly was observed in consecutive years.

3.3. Monthly and Seasonal Streamflow Characteristics. The mean monthly streamflow of the four catchments of the Ghba subbasin is shown in Figure 7 where Ghba 1 and Ghba 2 hydro stations are located in the Tahtay Ghba catchment. The Ghba River discharge was low from mid-October to mid-February but started to increase in the short rainy season (May) as shown in Figure 7. The highest flow was recorded in July at all stations except in Ghba 1 where the highest flow was recorded in August. The maximum streamflow was 53.6 cubic meters per second (m^3/s) in August at Ghba 1, while the minimum was 0.04 m^3/s in the dry season in October at Genfel station. The results from all of the stations in the subbasin showed a low flow during the dry season (October–February) but the increase started in the short rainy season (March) and slightly declines from May to June and starts to increase in the end of June, as shown in Figure 7.

3.4. Interannual and Temporal Streamflow Variability. The annual streamflow of 6 stations in the Ghba subbasin is shown in Figure 8. The annual flow varied from a minimum of 0.5 m^3/s at Ilala station to a maximum of 19.14 m^3/s at Ghba 1 station.

Table 2: Summary of descriptive statistical analysis of rainfall stations.

S/N	Station	Analyses period	Annual					Rainy season		Short rainy	
			Mean (mm)	SD	CV (%)	Skewness	Kurtosis	Mean (mm)	CV (%)	Mean (mm)	CV (%)
1	Abi-Adi	1983–2017	860	142	16	−0.236	−1.257	766	18	70	96
2	Adigrat	1970–2017	597	115	19	0.646	0.506	409	22	134	61
3	Adigudem	1983–2017	498	81	16	0.366	−0.888	406	25	66	70
4	Dengolat	1975–2017	612	98	16	0.779	0.813	516	20	77	77
5	E/hamus	1983–2017	651	118	18	−0.339	−0.616	553	20	76	81
6	Hawzen	1971–2017	512	92	18	0.355	−0.415	422	22	65	70
7	H/Selam	1983–2017	685	180	26	0.412	0.129	589	29	74	113
8	Mekelle AP	1983–2017	570	92	16	0.372	−0.605	476	20	77	75
9	Mekelle Ob	1953–2017	574	87	15	0.573	−0.381	479	20	78	70
10	Samre	1983–2017	650	161	24	−0.053	−0.749	553	27	61	81
11	Wukro	1973–2017	497	98	20	−0.592	0.080	428	23	49	88

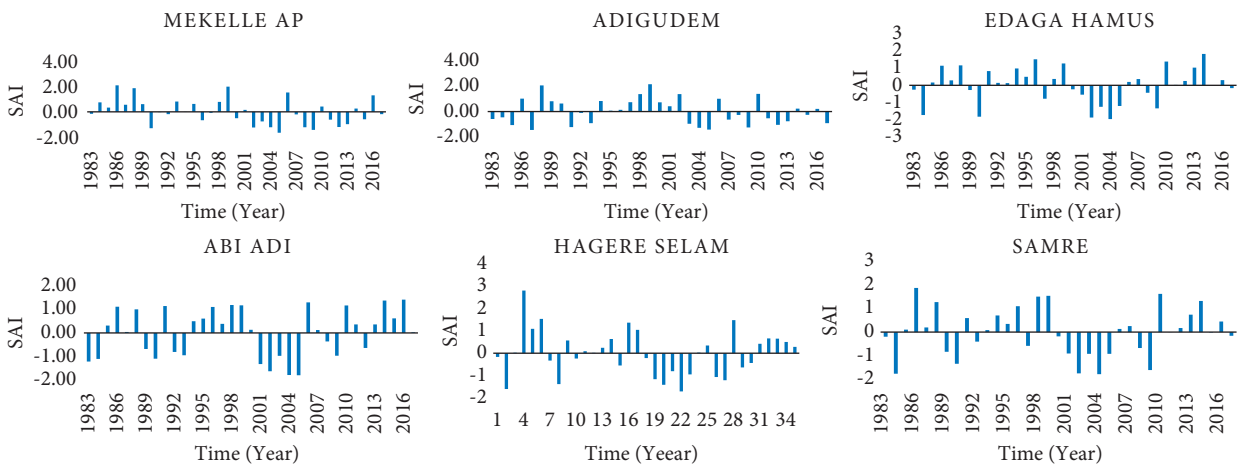


Figure 6: The annual standardized anomaly index (SAI) of rainfall stations in the subbasin.

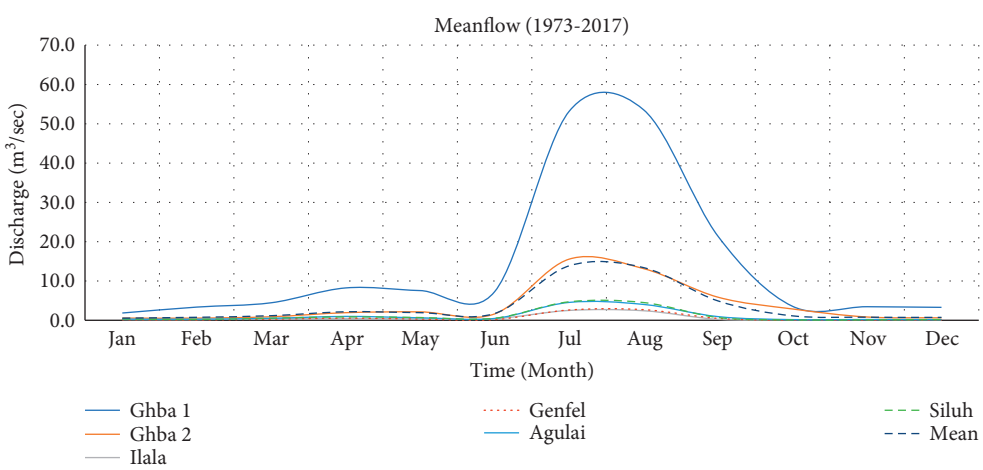


Figure 7: The mean monthly river discharge at six stations in the Ghba subbasin.

The chart in Figure 9 displays the annual streamflow variability of the subbasin at six stations. The hydrographs had a similar pattern on annual, rainy (JJAS) and short rainy (MAM) seasons river discharge. There were a few years when the high flow was recorded. Extreme streamflow was recorded at the Ghba 1, Ghba 2, and Ilala stations in 2001, 1996, and 1984, respectively. As identified by the graphical and linear trend on annual streamflow shown in Figure 9, a hydrology regime change in flow reduction was detected in all stations except in Agulai and Siluh stations. However, the temporal streamflow variability with decreasing pattern shown in Figure 9 was exhibited in all stations.

The summary of descriptive statistics of the six streamflow stations in Table 3 showed the annual streamflow in all the stations indicated a Z-score of skewness and kurtosis not closer to zero, indicating a significant difference

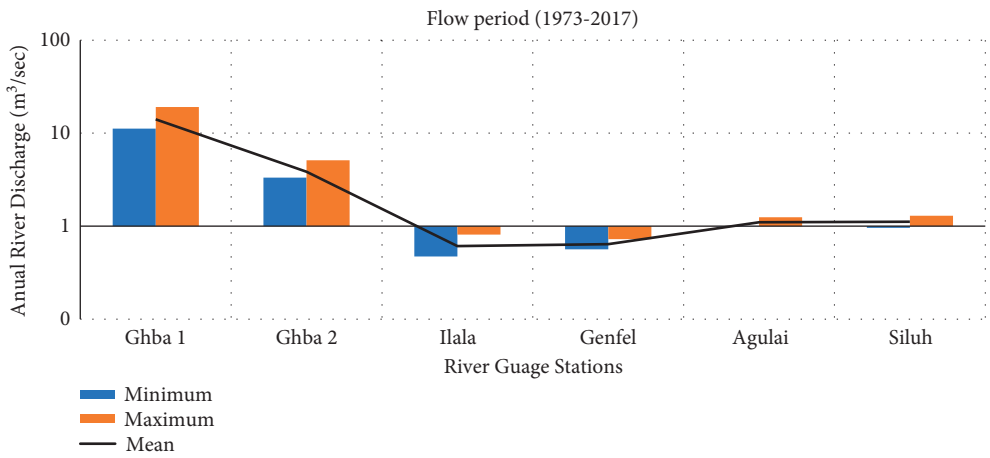


Figure 8: The annual river discharge of Ghba subbasin at six stations.

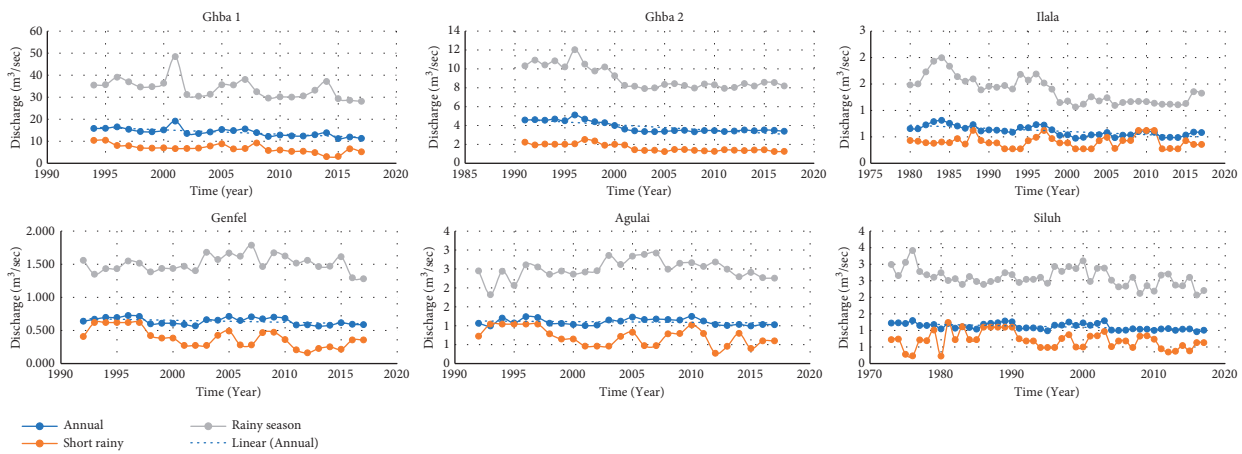


Figure 9: The temporal variability in the annual and seasonal streamflow in the subbasin.

from normality, i.e., positively skewed showed us the decreasing trend of the data series or somehow a change in the hydrology regime.

In this study, as can be seen in Table 3, all the stations in the annual and rainy season (JJAS) variability test showed normal (CV < 20), whereas the streamflow variability test in the short rainy season (MAM) showed a high CV from 23.2% to 36.5% in all the stations. Only two stations showed moderate variability, with a CV of 23.2% and 27% at Ghba 1 and at Ghba 2, respectively.

The chart in Figure 10 displays the SAI of the annual streamflow for all hydro stations in the Ghba subbasin. The figure shows that the negative anomaly varied from 51% to 69% of the total number of observations. Most of the negative anomalies are in a consequent year giving signals of a possible change in the hydrology regime.

3.5. Independence, Stationarity, and Homogeneity Tests. The results of the von-Neumann independence, Wald–Wolfowitz stationarity, and Mann–Whitney homogeneity tests applied on the daily rainfalls (in millimeters) and flow discharge (in m^3/sec) series recorded were analyzed using standardized variates, Skewness, Kurtosis, and Z-value. The critical values of the Z-score were ±1.645, ±1.96, and ±2.576 at 0.1, 0.05, and 0.01% significance level, respectively. For a probability of 0.1, 0.05, and 0.01% significance level, the series was independent at 11 rainfall and six hydro stations because the von-Neumann's Q statistic test results were greater than the critical values. For the same significance level, the Wald–Wolfowitz stationarity test shows not stationary for 11 rainfall and six hydro stations because the standardized variates test results were less than their tail probability. Finally, the Mann–Whitney homogeneity test was conducted and the tail probability of the series for 11 rainfall and six hydro stations was homogeneous because the Mann–Whitney homogeneity test statistic result was greater than the critical values. Thus, out of 16 rainfall and six hydro stations, the trend analyses were performed for the homogeneous, nonstationary, and independent 11 rainfall and six hydro stations.

3.6. Hydroclimatic Trend Analysis

3.6.1. Rainfall Data Series. Annual and seasonal trend analysis was performed for the selected homogeneous rainfall data series. A summary result of the Mann–Kendall

Table 3: Summary of descriptive statistical analysis of streamflow stations.

S/N	Station	No. of observations	Annual					Rainy season		Short rainy	
			Mean (m^3/sec)	Sd	CV (%)	Skewness	Kurtosis	Mean (m^3/sec)	CV (%)	Mean (m^3/sec)	CV (%)
1	Ghba 1	24	14.10	1.81	12.9	0.579	0.459	33.9	13.0	6.8	27.0
2	Ghba 2	27	3.84	0.56	14.6	0.765	−1.025	9.0	12.9	1.7	23.2
3	Ilala	26	0.64	0.05	8.0	0.087	−1.478	1.5	8.1	0.4	36.5
4	Genfel	26	0.64	0.05	8.0	0.087	−1.478	1.5	8.1	0.4	36.5
5	Agulai	26	1.10	0.08	7.5	0.341	−1.308	2.5	9.8	0.7	32.6
6	Siluh	45	1.12	0.10	8.6	0.267	−1.246	2.6	10.3	0.7	36.3

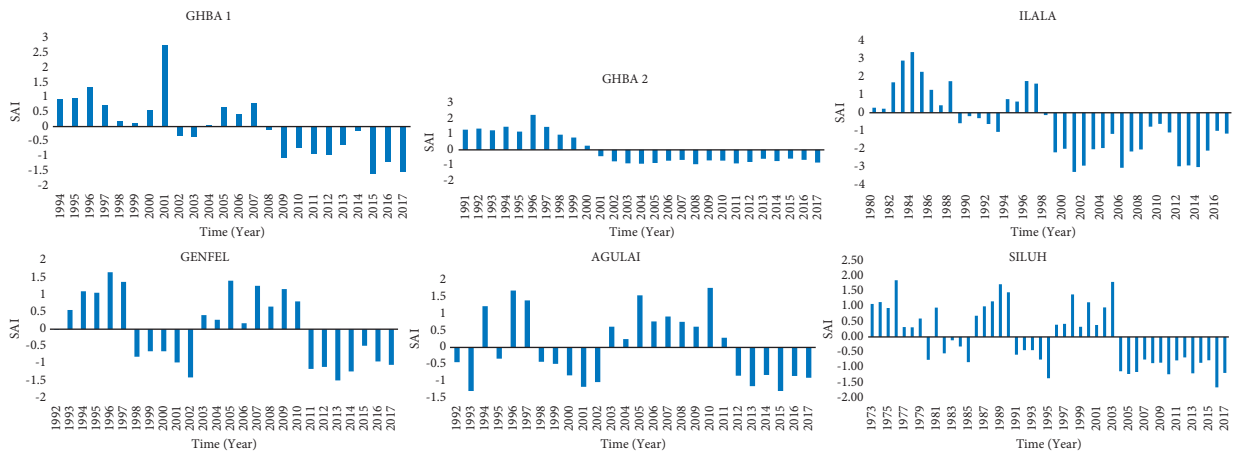


Figure 10: The annual standardized anomaly index (SAI) of streamflow stations in the subbasin.

(MK), Spearman's rho (SR), and Sen's slope (Q) tests of seasonal, and annual rainfall of 11 stations is given in Table 4. The analysis was checked against 1%, 5%, and 10% significant levels and is marked with bold and its P values (in brackets). The annual and seasonal rainfall trend at all stations has not shown a significant trend except at Dengolat (5% significant level) in the short rainy season and Mekelle AP (10% significant level) at annual rainfall, while five of the stations showed a decreasing (negative) trend and six an increasing trend at annual and rainy season. In the short rainy season, seven stations showed a decreasing trend while four stations showed a positive increasing trend. The dry season indicated no significant trend at all stations, while 10 of the stations showed a positive increasing trend and one decreasing trend. Moreover, a negative Zs and SR value was detected in about 45%, 63.63%, 45%, and 9.1% of the stations in annual, short rainy (MAM), rainy (JJAS), and dry seasons, respectively.

3.6.2. Temperature Data Series. Annual and seasonal trend analysis was performed for the selected homogeneous temperature data series. A summary result of the Mann–Kendall (MK), Spearman's rho (SR), and Sen's slope (Q) tests of seasonal, and annual temperature of 11 stations is given in Table 5. The analysis was checked against 1%, 5%, and 10% significant levels and is marked with bold and its P values (in brackets). The annual and seasonal temperature trend at eight stations has not shown a significant trend, whereas at Abi-Adi and Mekelle AP (5% significant level) in all seasons and at Adigudem (10% significant level) in all seasons except in dry season has shown a significant trend. Moreover, only one of the stations showed a decreasing (negative) trend and 10 stations showed an increasing trend during annual and seasonal periods. Moreover, a negative Zs and SR value was detected in about 9.1%, of the stations in annual, short rainy (MAM), rainy (JJAS), and dry seasons.

3.6.3. Discharge Data Series. Annual and seasonal trend analysis was performed for the selected homogeneous streamflow data series. A summary result of the Mann–Kendall (MK), Spearman's rho (SR), and Sen's slope (Q) tests of seasonal, and annual streamflow of 6 stations is given in Table 6. The analysis was checked against 1%, 5%, and 10% significant levels and is marked with bold and its P -values were all at a 5% significant level. Moreover, a negative Zs and SR value was detected in all stations either at annual, short rainy (MAM), rainy (JJAS), and dry seasons or all as given in Table 6.

As can be seen in Figure 11, the overall significant trend is on hydroclimatic data with increasing (+), decreasing (-), and no change (0) trends on annual rainfall, temperature, and streamflow.

3.7. Relationship between Rainfall and Streamflow. As shown in Figure 12 , the annual rainfall and streamflow at six hydro gauging stations were evaluated and induced a positive correlation coefficient which varied from 0.18 to 0.67. The

Table 4: Summary of the rainfall trend test using Mann–Kendall, SR, and Sen's slope.

N	Station	Annual			Rainy season			Short rainy			Dry season		
		Zs	SR	Sen's slope	Zs	SR	Sen's slope	Zs	SR	Sen's slope	Zs	SR	Sen's slope
1	Abi-Adi	0.12	0.11	1.75	0.09	0.08	1.76	−0.09	−0.08	−0.25	−0.01	−0.01	0.00
2	Adigrat	0.02	0.02	0.16	0.03	0.03	0.21	−0.02	−0.02	−0.06	0.02	0.01	0.03
3	Adigudem	−0.06	−0.05	−0.78	−0.01	−0.01	−0.18	−0.08	−0.07	−0.39	0.05	0.04	0.11
4	Dengolat	0.05	0.04	0.34	−0.08	−0.07	−0.54	0.18 (0.03)	0.17	0.68	0.02	0.02	0.00
5	E/hamus	0.02	0.02	0.31	0.06	0.07	0.62	−0.03	−0.02	−0.11	0.12	0.11	0.25
6	Hawzen	−0.12	−0.11	−0.80	−0.10	−0.10	−0.78	−0.02	−0.02	−0.06	0.02	0.02	0.00
7	H/Selam	−0.44	−0.45	−0.01	0.07	0.06	1.99	−0.04	−0.05	−0.59	0.09	0.08	0.23
8	Mekelle AP	−0.23 (0.05)	0.22	−3.42	−0.17	−0.18	−1.83	0.01	0.01	0.05	0.11	0.12	0.21
9	Mekelle Ob	−0.13	−0.12	−0.86	−0.13	−0.14	−0.65	0.01	0.01	0.00	0.03	0.02	0.00
10	Samre	0.03	0.04	0.40	0.06	0.05	1.07	−0.03	−0.02	−0.11	0.11	0.12	0.25
11	Wukro	0.13	0.12	0.88	0.07	0.06	0.58	0.08	0.07	0.22	0.07	0.06	0.03

Table 5: Summary of temperature trend test using Mann–Kendall and Sen's slope.

S/N	Station	Annual			Rainy season			Short rainy			Dry season		
		Zs	SR	Sen's slope	Zs	SR	Sen's slope	Zs	SR	Sen's slope	Zs	SR	Sen's slope
1	Abi-Adi	0.46 (0.02)	0.53	0.02	0.36	0.35	0.01	0.40	0.41	0.03	0.29	0.28	0.02
2	Adigrat	−0.09	−0.10	0.16	−0.07	0.06	0.00	−0.04	0.05	0.00	−0.04	0.03	0.00
3	Adigudem	0.20 (0.09)	0.19	0.02	0.26	0.24	0.02	0.35	0.36	0.03	0.08	0.09	0.01
4	Dengolat	0.03	0.02	0.00	0.01	0.02	0.00	0.00	0.00	0.68	0.01	0.01	0.00
5	E/hamus	0.06	0.07	0.00	0.01	0.01	0.00	0.04	0.05	0.00	0.06	0.05	0.00
6	Hawzen	0.06	0.05	0.00	0.04	0.06	0.00	0.04	0.04	0.00	0.06	0.07	0.00
7	H/Selam	0.4	0.06	0.00	0.04	0.05	0.00	0.01	0.01	0.00	0.01	0.01	0.00
8	Mekelle AP	0.49 (0.02)	0.51	0.03	0.42	0.44	0.02	0.38	0.37	0.03	0.36	0.35	0.02
9	Mekelle Ob	0.02	0.02	0.00	0.00	0.00	0.00	0.02	0.03	0.00	0.00	0.00	0.00
10	Samre	0.02	0.02	0.00	0.00	0.00	0.00	0.02	0.03	0.00	0.00	0.00	0.00
11	Wukro	0.08	0.07	0.00	0.02	0.02	0.01	0.02	0.00	0.22	0.07	0.06	0.00

Table 6: Summary of the streamflow trend test using Mann–Kendall and Sen's slope.

Station	Annual			Rainy season			Short rainy			Dry season		
	Zs	SR	Sen's slope	Zs	SR	Sen's slope	Zs	SR	Sen's slope	Zs	SR	Sen's slope
Ghba 1	−0.62	−0.61	−0.19	−0.45	−0.44	−0.32	−0.58	−0.53	−0.08	−0.65	−0.64	−0.17
Ghba 2	−0.48	−0.51	−0.05	−0.44	−0.42	−0.10	−0.59	−0.55	−0.04	−0.58	−0.56	−0.02
Ilala	−0.51	−0.49	−0.01	−0.57	−0.53	−0.02	−0.07	−0.08	0.00	0.02	0.02	0.00
Genfel	−0.28	−0.27	0.00	0.09	0.07	0.00	−0.47	−0.48	−0.02	−0.15	−0.13	0.00
Agulai	−0.17	−0.15	0.00	0.09	0.08	0.00	−0.32	−0.33	−0.02	−0.15	−0.14	0.00
Siluh	−0.39	−0.37	0.00	−0.30	−0.29	−0.01	−0.20	−0.19	−0.01	0.08	0.07	0.00

correlation coefficient was 0.18, 0.27, and 0.67 at Agulai, Siluh, and Ghba 1, respectively. By comparison, the correlation was very low in one station at Agulai (0.18), and low in four stations: Siluh (0.27), Ilala (0.29), Ghba 2 (0.30), and Genfel (0.32). While one station at Ghba 1 showed relatively higher a positive correlation of 0.67. Hence, the current finding revealed the existence of a positive correlation relationship in a different time series of the tested stations.

4. Conclusion and Recommendations

The current study presents a full statistical analysis of the presence of trends and point changes in rainfall, temperate, and streamflow in the Ghba basin. The analyses were done for 11 rainfall and temperature stations as well as six streamflow monitoring stations. Linkages between the trends in rainfall, temperature, and streamflow across the whole subbasin were cautiously surveyed at different scales. Annual and seasonal trend analysis for rainfall, temperature, and discharge was performed at each station. To meet the main objective of this study, i.e., to understand the hydroclimatic variables trend, characterization, the Mann–Kendall test, SR, and Sen's slope test estimator were employed. Following these analyses, the main driving force for streamflow reduction over the subbasin was assessed.

The finding of this study indicated the occurrence of temporal rainfall variability. Out of eleven rainfall stations, around 73% of the stations showed the mean annual rainfall had a CV of less than 20%, while 27% of stations indicated a CV between 20 and 26%. Hence, this reviled in the annual time series eight stations showed no rainfall variability and three stations showed a moderate variability. In the rainy

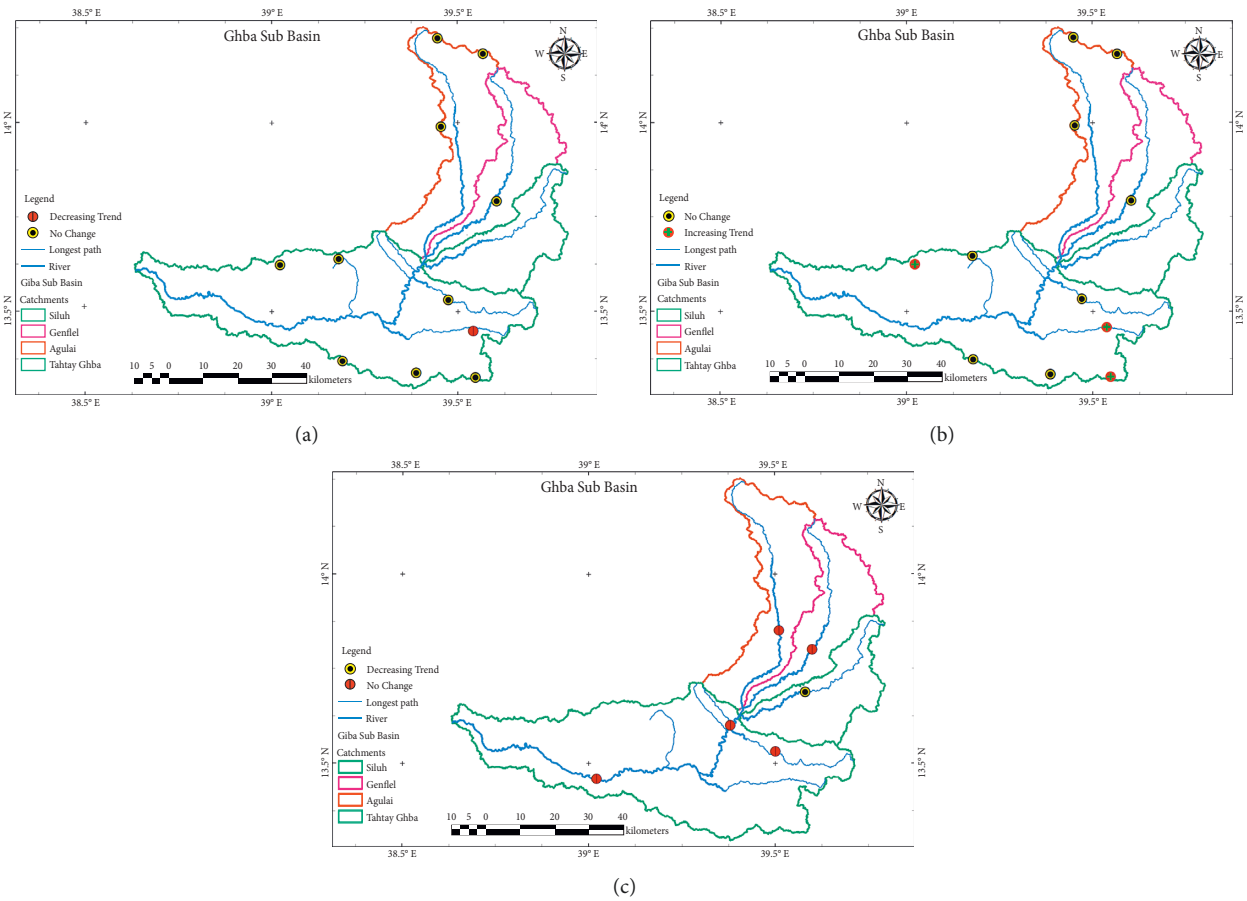


Figure 11: Location of hydroclimatic stations with increasing (+), decreasing (−), and no change (0) trends on annual rainfall (a), temperature (b), and streamflow (c).

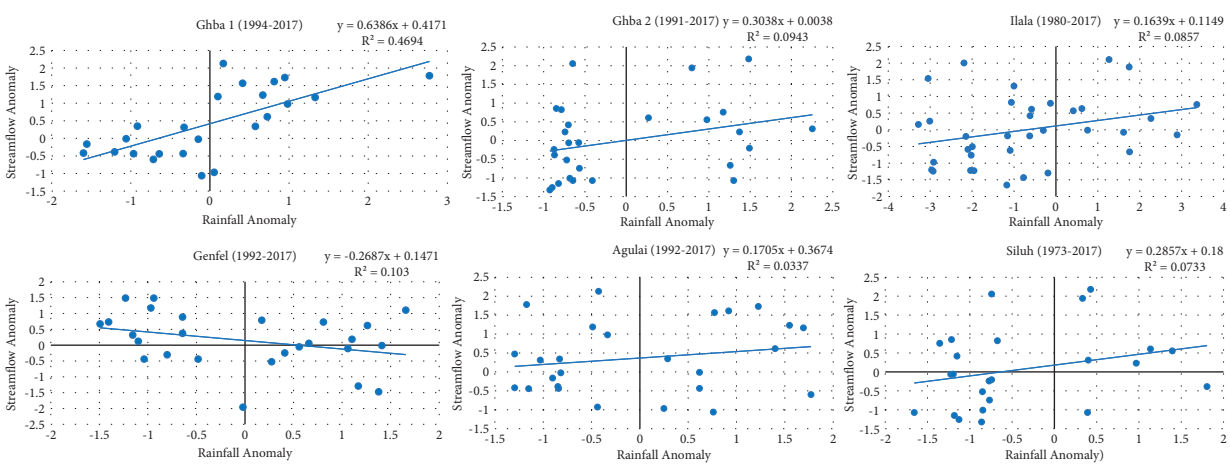


Figure 12: Relationship between rainfall and streamflow trend in the Ghba subbasin.

seasonal period, except Abi-Adi station all stations were characterized by moderate variability. However, during the period of the dry season rainfall showed that 100% of the stations had a CV > 30 which indicates high variability. Understanding this variability trend is key to the water allocation plan in the subbasin.

In order to understand the seasonal comparison of changes in hydroclimatic variables, the analyses were done for three seasons: rainy, short rainy, and dry.

The SAI in annual and seasonal time series indicated a negative anomaly. On seasonal time series, short rainy (MAM) and the rainy season (JJAS) showed a negative

anomaly between 45% to 67% and 45% to 65% of the observations, respectively. Furthermore, the SAI result showed that most of the years exhibited negative anomalies in discharge, indicating a drier period. Moreover, a lot of large and small-scale irrigation schemes are found in Ghba subbasin [86, 87], especially in the Agulai and Genfel catchments, which could be directly affected by low water availability during short rainy (MAM). The occurrence of high streamflow variability combined with the expansion of irrigation in the subbasin influences agriculture and crop productivity, which aggravates the prevalence of food insecurity and drought in the subbasin.

The results of MK, Spearman's rho (SR), and Sen's slope tests showed a statistically significant decreasing annual discharge trends for all the stations, while no significant decreasing trend in annual rainfall was observed. Similar studies found the same result [6, 27]. The majority of rainfall stations exhibited no change trend in the mean annual rainfall for rainy and short rainy seasons. But, only one station at Mekelle AP showed a decreasing trend. The study found also a positive significant trend in the temperature of three stations. In general, the trends exhibited higher variability of discharge during the annual, rainy, short rainy, and dry seasons opposite to the rainfall trends. A weak relationship between rainfall and streamflow leads to assessing other factors of catchment characteristics change, catchment management interventions, and water abstractions upstream.

Based on the prior results and discussion, it can be decided that decreasing and increasing levels of rainfall, discharge, and temperature throughout the stations showed the change in hydroclimatic variables. This would propose that any climate and hydrological studies need to understand the variability and trends of a given dataset before applying for any future impact assessment.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The study described in this manuscript has been funded by MG Water Resource Consultancy Firm. The authors are extremely grateful to the Ethiopian National Meteorological Services Agency for providing climate data and the Ministry of Water and Energy (Ethiopia) for providing water resource data.

References

[1] S. Tesfaye, E. Birhane, T. Leijnse, and S. J. S.T. E. van der Zee, "Climatic controls of ecohydrological responses in the highlands of northern Ethiopia," *Science of the Total Environment*, vol. 609, pp. 77–91, 2017.

[2] J. A. Gore, J. Banning, A. F. Casper et al., *River Resource Management and the Effects of Changing Landscapes and Climate*, pp. 295–312, Wiley, Hoboken, NJ, USA, 2016.

[3] P. H. Gleick, "Climate change, hydrology, and water resources," *Reviews of Geophysics*, vol. 27, no. 3, pp. 329–344, 1989.

[4] L. Cuo, Y. Zhang, Y. Gao, Z. Hao, and L. Cairang, "The impacts of climate change and land cover/use transition on the hydrology in the upper Yellow River Basin, China," *Journal of Hydrology*, vol. 502, pp. 37–52, 2013.

[5] Q. Zhang, C.-Y. Xu, and T. J. W. Yang, "Variability of water resource in the yellow river basin of past 50 Years, China," *Water Resources Management*, vol. 23, no. 6, pp. 1157–1170, 2009.

[6] T. G. Gebremicael, Y. A. Mohamed, E. Y. J. H. Hagos, and E. S. Sciences, "Temporal and spatial changes of rainfall and streamflow in the Upper Tekezē-Atbara river basin," *Ethiopia*, vol. 21, pp. 2127–2142, 2017.

[7] D. J. G. E. C. Conway, "From headwater tributaries to international river: observing and adapting to climate variability and change in the Nile basin," *Global Environmental Change*, vol. 15, no. 2, pp. 99–114, 2005.

[8] P. Dahal, M. L. Shrestha, J. Panthi, and D. J. E. R. Pradhananga, "Modeling the future impacts of climate change on water availability in the Karnali River Basin of Nepal Himalaya," *Environmental Research*, vol. 185, Article ID 109430, 2020.

[9] T. Estrela, M. Pérez-Martin, and E. Vargas, "Impacts of climate change on water resources in Spain," *Hydrological Sciences Journal*, vol. 57, no. 6, pp. 1154–1167, 2012.

[10] S. Huang, Q. Huang, J. Chang, and G. Leng, "Linkages between hydrological drought, climate indices and human activities: a case study in the Columbia River basin," *International Journal of Climatology*, vol. 36, no. 1, pp. 280–290, 2016.

[11] M. D. Tomer and K. E. Schilling, "A simple approach to distinguish land-use and climate-change effects on watershed hydrology," *Journal of Hydrology*, vol. 376, pp. 24–33, 2009.

[12] P. Dey and A. J. Mishra, "Separating the impacts of climate change and human activities on streamflow: a review of methodologies and critical assumptions," *Journal of Hydrology*, vol. 548, pp. 278–290, 2017.

[13] F. E. Dunn, S. E. Darby, R. J. Nicholls, S. Cohen, C. Zarfl, and B. M. Fekete, "Projections of declining fluvial sediment delivery to major deltas worldwide in response to climate change and anthropogenic stress," *Environmental Research Letters*, vol. 14, no. 8, Article ID 084034, 2019.

[14] J. A. Elliott, "Is the future blue-green? A review of the current model predictions of how climate change could affect pelagic freshwater cyanobacteria," *Water Research*, vol. 46, no. 5, pp. 1364–1371, 2012.

[15] M. H. I. Dore, "Climate change and changes in global precipitation patterns: what do we know?" *Environment International*, vol. 31, no. 8, pp. 1167–1181, 2005.

[16] N. Dahal, U. B. Shrestha, A. Tuitui, and H. R. Ojha, "Temporal changes in precipitation and temperature and their implications on the streamflow of rosi river, Central Nepal," *Climate*, vol. 7, no. 1, p. 3, 2018.

[17] J. J. Barnes, *Scale and Agency: Climate Change and the Future of Egypt's Water*, pp. 127–145, Yale University Press, New Haven, CT, USA, 2015.

[18] J. Lorenzo-Lacruz, S. M. Vicente-Serrano, J. I. López-Moreno, and E. Morán-Tejeda, "Recent trends in Iberian streamflows," *Journal of Hydrology*, vol. 414, pp. 463–475, 2012.

[19] B. Paudel, B. S. Acharya, R. Ghimire, K. R. Dahal, and P. Bista, "Adapting agriculture to climate change and variability in

Chitwan: long-term trends and farmers' perceptions," *Agricultural Research*, vol. 3, no. 2, pp. 165–174, 2014.

[20] F. Pellicciotti, A. Bauder, and M. Parola, "Effect of glaciers on streamflow trends in the Swiss Alps," *Water Resources Research*, vol. 46, 2010.

[21] X. Zhang, K. D. Harvey, W. D. Hogg, and T. R. Yuzyk, "Trends in Canadian streamflow," *Water Resources Research*, vol. 37, no. 4, pp. 987–998, 2001.

[22] W. A. Gebert and W. R. Krug, "Streamflow trends in wisconsin's driftless area," *Journal of the American Water Resources Association*, vol. 32, no. 4, pp. 733–744, 1996.

[23] M.-V. Birsan, P. Molnar, P. Burlando, and M. Pfaundler, "Streamflow trends in switzerland," *Journal of Hydrology*, vol. 314, pp. 312–329, 2005.

[24] S. G. Setegn, D. Rayner, A. M. Melesse, B. Dargahi, and R. Srinivasan, "Impact of climate change on the hydroclimatology of Lake Tana Basin," *Ethiopia*, vol. 47, 2011.

[25] M. Bhutiyani, "Spatial and temporal variability of climate change in high-altitude regions of NW Himalaya," *Climate Change, Glacier Response, and Vegetation Dynamics in the Himalaya*, pp. 87–101, Springer, Berlin, Germany, 2016.

[26] D. Conway and M. J. C. Hulme, "Recent fluctuations in precipitation and runoff over the Nile sub-basins and their impact on main Nile discharge," *Climatic Change*, vol. 25, no. 2, pp. 127–151, 1993.

[27] T. Gebremicael, Y. Mohamed, G. Betrie, P. Van der Zaag, and E. Teferi, "Trend analysis of runoff and sediment fluxes in the upper blue nile basin: a combined analysis of statistical tests, physically-based models and landuse maps," *Journal of Hydrology*, vol. 482, pp. 57–68, 2013.

[28] Z. K. Tesemma, Y. A. Mohamed, and T. S. Steenhuis, "Trends in rainfall and runoff in the blue Nile basin: 1964-2003," *Hydrological Processes*, vol. 24, no. 25, pp. 3747–3758, 2010.

[29] D. F. Mekonnen, Z. Duan, T. Rientjes, and M. J. H. Disse, "Analysis of combined and isolated effects of land-use and land-cover changes and climate change on the upper Blue Nile River basin's streamflow," *Hydrology and Earth System Sciences*, vol. 22, no. 12, pp. 6187–6207, 2018.

[30] M. T. Taye, P. Willems, and P. J. Block, "Implications of climate change on hydrological extremes in the Blue Nile basin: a review," *Journal of Hydrology: Regional Studies*, vol. 4, pp. 280–293, 2015.

[31] M. L. Berihun, A. Tsunekawa, N. Haregeweyn et al., "Hydrological responses to land use/land cover change and climate variability in contrasting agro-ecological environments of the Upper Blue Nile basin, Ethiopia," *Science of the Total Environment*, vol. 689, pp. 347–365, 2019.

[32] M. T. Tadese, L. Kumar, R. Koech, and B. J. H. Zemadim, "Hydro-climatic variability: a characterisation and trend study of the awash river basin, Ethiopia," *Hydrology*, vol. 6, no. 2, 2019.

[33] L. L. Mahato, M. Kumar, S. Suryavanshi, S. K. Singh, and D. J. E. Lal, "Statistical investigation of long-term meteorological data to understand the variability in climate: a case study of Jharkhand, India," *Environment, Development and Sustainability*, vol. 23, no. 11, pp. 16981–17002, India, 2021.

[34] B. A. Abebe, B. Grum, A. M. Degu, and H. J. M. A. Goitom, "Spatio temporal rainfall variability and trend analysis in the Tekeze Atbara river basin, northwestern Ethiopia," *Meteorological Applications*, vol. 29, no. 2, Article ID e2059, 2022.

[35] Y. Seleshi and U. Zanke, "Recent changes in rainfall and rainy days in Ethiopia," *International Journal of Climatology*, vol. 24, no. 8, pp. 973–983, 2004.

[36] T. Gebremicael, Y. Mohamed, and P. Van der Zaag, "Attributing the hydrological impact of different land use types and their long-term dynamics through combining parsimonious hydrological modelling, alteration analysis and PLSR analysis," *Science of the Total Environment*, vol. 660, pp. 1155–1167, 2019.

[37] J. Nyssen, A. Frankl, A. Zenebe, J. Deckers, and J. Poesen, "Land management in the northern Ethiopian highlands: local and global perspectives; past, present and future," *Land Degradation & Development*, vol. 26, no. 7, pp. 759–764, 2015.

[38] J. Nyssen, S. Tielens, T. Gebreyohannes et al., "Understanding spatial patterns of soils for sustainable agriculture in northern Ethiopia's tropical mountains," *Plos One*, vol. 14, no. 10, Article ID e0224041, 2019.

[39] T. G. Gebremicael, Y. A. Mohamed, P. v Zaag, and E. Y. Hagos, "Temporal and spatial changes of rainfall and streamflow in the Upper Tekezē-Atbara river basin, Ethiopia," *Hydrology and Earth System Sciences*, vol. 21, no. 4, pp. 2127–2142, 2017.

[40] A. Alhamshry, A. A. Fenta, H. Yasuda, R. Kimura, and K. J. W. Shimizu, "Seasonal rainfall variability in Ethiopia and its long-term link to global sea surface temperatures," *Water*, vol. 12, no. 1, 2019.

[41] E. Negash, T. Getachew, E. Birhane, and H. J. E. S. Gebrewahed, "Ecosystem Service value distribution along the agroecological gradient in north-central Ethiopia," *Earth Systems and Environment*, vol. 4, no. 1, pp. 107–116, 2020.

[42] M. Behailu, N. Tadesse, A. Legesse, D. J. M. Teklu, and E. Ilri, *Community Based Irrigation Management in the Tekeze Basin: Performance Evaluation of Small Scale Irrigation Schemes*, 2004.

[43] F. Hagos, *Land Degradation and Strategies for Sustainable Land Management in the Ethiopian Highlands: Tigray Region: ILRI (Aka ILCA and ILRAD)*, 2002.

[44] S. K. Weldegebriel and K. J. S. Yeshitela, "Measuring the semi-century ecosystem-service value variation in Mekelle city region, northern Ethiopia," *Sustainability*, vol. 13, no. 18, Article ID 10015, 2021.

[45] ENMSA, *Climate Data*, ENMSA, Addis Ababa, Ethiopia, 2021.

[46] F.-W. Chen and C.-W. Liu, "Estimation of the spatial rainfall distribution using inverse distance weighting (IDW) in the middle of Taiwan," *Paddy and Water Environment*, vol. 10, no. 3, pp. 209–222, 2012.

[47] A. S. Dhevi, "Imputing missing values using Inverse Distance Weighted Interpolation for time series data," *Sixth International Conference on Advanced Computing (ICoAC)*, pp. 255–259, IEEE, Piscataway, NJ, USA, 2014.

[48] R. S. Teegavarapu and V. Chandramouli, "Improved weighting methods, deterministic and stochastic data-driven models for estimation of missing precipitation records," *Journal of Hydrology*, vol. 312, pp. 191–206, 2005.

[49] R. S. Teegavarapu, M. Tufail, and L. Ormsbee, "Optimal functional forms for estimation of missing precipitation data," *Journal of Hydrology*, vol. 374, no. 1-2, pp. 106–115, 2009.

[50] R. S. V. Teegavarapu, A. Aly, C. S. Pathak, J. Ahlquist, H. Fuelberg, and J. J. I. JoC. Hood, "Infilling missing precipitation records using variants of spatial interpolation and data driven methods: use of optimal weighting parameters and nearest neighbour-based corrections," *International Journal of Climatology*, vol. 38, no. 2, pp. 776–793, 2018.

[51] D. O'sullivan and D. Unwin, *Geographic Information Analysis*, John Wiley & Sons, Hoboken, NJ, USA, 2003.

[52] R. S. V. Teegavarapu, "Missing precipitation data estimation using optimal proximity metric-based imputation, nearest-neighbour classification and cluster-based interpolation methods," *Hydrological Sciences Journal*, vol. 59, no. 11, pp. 2009–2026, 2014.

[53] M. H. Daba, G. T. Ayele, and S. J. You, "Long-term homogeneity and trends of hydroclimatic variables in upper awash river basin, Ethiopia," *Advances in Meteorology*, vol. 202021 pages, Article ID 8861959, 2020.

[54] A. Asfaw, B. Simane, A. Hassen, and A. J. W. Bantider, "Variability and time series trend analysis of rainfall and temperature in northcentral Ethiopia: a case study in Woleka sub-basin," *Weather and Climate Extremes*, vol. 19, pp. 29–41, 2018.

[55] M. O. Kisaka, M. Mucheru-Muna, F. Ngetich, J. Mugwe, D. Mugendi, and F. J. AiM. Mairura, "rainfall variability, drought characterization, and efficacy of rainfall data reconstruction: case of eastern Kenya," *Advances in Meteorology*, vol. 2015, Article ID 380404, 16 pages, 2015.

[56] A. Alemayehu, W. J. G. A. S. A. Bewket, and P. Geography, "Local spatiotemporal variability and trends in rainfall and temperature in the central highlands of Ethiopia," *Geografiska Annaler-Series A: Physical Geography*, vol. 99, no. 2, pp. 85–101, 2017.

[57] M. M. Alemu, G. T. Bawoke, and C. Change, "Analysis of spatial variability and temporal trends of rainfall in Amhara region, Ethiopia," *Journal of Water and Climate Change*, vol. 11, no. 4, pp. 1505–1520, 2020.

[58] G. Bayable, G. Amare, G. Alemu, and T. Gashaw, "Spatio-temporal variability and trends of rainfall and its association with pacific ocean sea surface temperature in west harerge zone, eastern Ethiopia," *Environmental Systems Research*, vol. 10, pp. 7–21, 2021.

[59] K. H. Hamed and A. R. J. J. Rao, "A modified Mann-Kendall trend test for autocorrelated data," *Journal of Hydrology*, vol. 204, pp. 182–196, 1998.

[60] Z. Hu, S. Liu, G. Zhong, H. Lin, and Z. Zhou, "Modified Mann-Kendall trend test for hydrological time series under the scaling hypothesis and its application," *Hydrological Sciences Journal*, vol. 65, no. 14, pp. 2419–2438, 2020.

[61] H. Mann, "Nonparametric tests against trend," *Econometrica*, vol. 13, pp. 245–259, 1945.

[62] V. Nourani, A. Danandeh Mehr, and N. J. E. E. S. Azad, "Trend analysis of hydroclimatological variables in Urmia lake basin using hybrid wavelet," *Mann-Kendall and Şen tests*, vol. 77, pp. 1–18, 2018.

[63] A. Mondal, S. Kundu, A. Mukhopadhyay, and S. E. Earth, "Rainfall trend analysis by Mann-Kendall test: a case study of north-eastern part of Cuttack district, Orissa," vol. 2, pp. 70–78, 2012.

[64] R. Ali, A. Kuriqi, S. Abubaker, and O. J. W. Kisi, "Long-term trends and seasonality detection of the observed flow in Yangtze River using Mann-Kendall and Sen's innovative trend method," *Water*, vol. 11, no. 9, 2019.

[65] F. Wang, W. Shao, H. Yu et al., "Re-evaluation of the power of the mann-kendall test for detecting monotonic trends in hydrometeorological time series," *Frontiers in Earth Science*, vol. 14, 2020.

[66] S. Yue, P. Pilon, and G. Cavadias, "Power of the Mann-Kendall and Spearman's rho tests for detecting monotonic trends in hydrological series," *Journal of Hydrology*, vol. 259, pp. 254–271, 2002.

[67] G. C. Okafor, O. D. Jimoh, K. I. J. A. Larbi, and C. Sciences, "Detecting changes in hydro-climatic variables during the last four decades (1975–2014) on downstream kaduna river catchment, Nigeria," *Atmospheric and Climate Sciences*, vol. 7, pp. 161–175, 2017.

[68] J. Mallick, S. Talukdar, M. Alsubih et al., "Analysing the trend of rainfall in Asir region of Saudi Arabia using the family of Mann-Kendall tests, innovative trend analysis, and detrended fluctuation analysis," *Theoretical and Applied Climatology*, vol. 143, pp. 823–841, 2021.

[69] F. Fathian, Z. Dehghan, M. H. Bazrkar, and S. Eslamian, "Trends in hydrological and climatic variables affected by four variations of the Mann-Kendall approach in Urmia Lake basin," *Iran*, vol. 61, pp. 892–904, 2016.

[70] E. Nikzad Tehrani, H. Sahour, and M. J. Booij, "Trend analysis of hydro-climatic variables in the north of Iran," *Theoretical and Applied Climatology*, vol. 136, pp. 85–97, 2019.

[71] R. M. Da Silva, C. A. G. Santos, M. Moreira, J. Corte-Real, V. C. L. Silva, and I. Medeiros, "Rainfall and river flow trends using Mann–Kendall and Sen's slope estimator statistical tests in the Cobres River basin," *Natural Hazards*, vol. 77, no. 2, pp. 1205–1221, 2015.

[72] H. Wu and H. Qian, "Innovative trend analysis of annual and seasonal rainfall and extreme values in Shaanxi, China, since the 1950s," *International Journal of Climatology*, vol. 37, no. 5, pp. 2582–2592, 2017.

[73] P. K. Sen, "Estimates of the regression coefficient based on kendall's tau," *Journal of the American Statistical Association*, vol. 63, no. 324, pp. 1379–1389, 1968.

[74] B. Anand, D. Karunanidhi, T. Subramani, K. Srinivasamoorthy, and M. J. E. Suresh, "Long-term trend detection and spatio-temporal analysis of groundwater levels using GIS techniques in Lower Bhavani River basin, Tamil Nadu, India," *Environment, Development and Sustainability*, vol. 22, no. 4, pp. 2779–2800, 2020.

[75] A. S. Yusuf, C. O. Edet, C. O. Oche, and E. Agbo, "Trend analysis of temperature in gombe state using mann kendall trend test," *Journal of Scientific Research and Reports*, vol. 20, 2018.

[76] C. Spearman, "The proof and measurement of association between two things," *American Journal of Psychology*, vol. 15, no. 1, pp. 72–101, 1904.

[77] E. L. Lehmann and H. J. D'Abrera, *Nonparametrics: Statistical Methods Based on Ranks*, Springer, Berlin, Germany, 1975.

[78] M. Rodrigues, J. San Miguel, S. Oliveira, F. Moreira, and A. Camia, "An insight into spatial-temporal trends of fire ignitions and burned areas in the European Mediterranean countries," *Fire Ecology and Post-Fire Restoration*, vol. 3, 2013.

[79] C. S. Sharma, S. N. Panda, R. P. Pradhan, A. Singh, and A. Kawamura, "Precipitation and temperature changes in eastern India by multiple trend detection methods," *Atmospheric Research*, vol. 180, pp. 211–225, 2016.

[80] M. A. Rahman, L. Yunsheng, N. J. M. Sultana, and A. Physics, "Analysis and prediction of rainfall trends over Bangladesh using Mann-Kendall, Spearman's rho tests and ARIMA model," *Meteorology and Atmospheric Physics*, vol. 129, no. 4, pp. 409–424, 2017.

[81] E. Aamir and I. Hassan, "Trend analysis in precipitation at individual and regional levels in Baluchistan, Pakistan," *IOP Conference Series: Materials Science and Engineering*, vol. 414, Article ID 012042, 2018.

[82] S. DEMİR, "Comparison of normality tests in terms of sample sizes under different skewness and Kurtosis coefficients," *International Journal of Assessment Tools in Education*, vol. 9, no. 2, pp. 397–409, 2022.

[83] K. Jöreskog,1999, https://www.ssicentral.com/lisrel.Formulas%20for%20skewness%20and%20kurtosis.

[84] G. R. Sandercock, P. D. Bromley, and D. A. Brodie, "The reliability of short-term measurements of heart rate variability," *International Journal of Cardiology*, vol. 103, no. 3, pp. 238–247, 2005.

[85] M. Doughty and A. Eye, "Importance of standardizing the number of cells measured for coefficient of variation (COV) estimates of corneal endothelial cell area values as relevant to contact lens wear," *Contact Lens and Anterior Eye*, vol. 44, no. 3, Article ID 101336, 2021.

[86] M. Hiben, T. Goitom, and T. Berhanu, "Sectoral water allocation in the ghba sub-basin," *Tigray Water Resource Bureau*, vol. 2, 2016.

[87] T. G. Gebremicael, *Understanding the Impact of Human Interventions on the Hydrology of Nile Basin Headwaters, the Case of Upper Tekeze Catchments*, CRC Press, Boca Raton, FL, USA, 2019.

Hindawi
Advances in Meteorology
Volume 2022, Article ID 7374193, 6 pages
https://doi.org/10.1155/2022/7374193

Research Article

Land-Atmosphere Energy Exchange Characteristics in Ali of Tibetan

Ge Wang,[1] **Lin Han**,[2] **and Xingying Tang**[1]

[1]*Institute of Plateau Meteorology, China Meteorological Administration, Chengdu 610072, China*
[2]*School of Atmospheric Sciences, Chengdu University of Information Technology, Chengdu 610225, China*

Correspondence should be addressed to Ge Wang; wg800110@aliyun.com

Received 7 May 2022; Revised 5 July 2022; Accepted 7 July 2022; Published 11 August 2022

Academic Editor: Upaka Rathnayake

Based on the comprehensive data from the land-atmosphere interaction observation station in Ali of Tibetan in 2019, the characteristics of land-atmosphere energy exchange processes in Ali were analyzed. The results indicated that the timing of the mean intraday net radiation peak in Ali over the past 20 years has been delayed, and the month when the maximum monthly mean net radiation occurred has been delayed by about 2 months; the maximum daily mean, maximum monthly mean, minimum monthly mean, and annual mean sensible heat were 99.63 w/m^2, 76.53 w/m^2, 17.47 w/m^2, and 46.74 w/m^2, respectively, and the maximum daily mean, maximum monthly mean, minimum monthly mean, and annual mean latent heat flux were 73.27 w/m^2, 36.13 w/m^2, 0.67 w/m^2, and 8.32 w/m^2, respectively; and the monthly mean sensible heat was greater than the latent heat in all months.

1. Introduction

The Tibet Plateau is a source of both heat and thermal convective dynamic disturbances that enter the free atmosphere through land-atmosphere energy exchanges, such as radiation and heat transport. This has an important thermal and dynamic effect on the formation and changes in weather and climate in China. Surface net radiation is the main energy source driving atmospheric movement [1, 2]. As an important component of land-air interaction, surface heat flux is an important index to evaluate the simulation ability of land surface models and has been the focus of research in boundary layer meteorology and atmospheric physics [3]. Observations and experimental studies of land-atmosphere exchange processes have been conducted since the 1950s. While within China, most of the observations and experimental studies of land-atmosphere exchange processes were carried out in Tibet and other areas such as the northwest arid region [4–6]. Research in China has come to fruition on energy budget [7, 8], radiation balance [9], heat source intensity [8–10], and physical processes in land-atmosphere exchange processes [10–12]. Previous studies have shown the distribution of sensible and latent heat in the Tibetan Plateau has obvious regional differences and seasonal variation characteristics [13]. The latent heat in the eastern plateau is greater than the sensible heat, while the sensible heat in the southern plateau is greater than the latent heat [14, 15]. However, as the Tibet Plateau has a harsh climate, difficult terrain, and few observation sites, most of the observations and experimental studies of land-atmosphere exchange processes over the Tibetan Plateau are based on short-term or single-point observations [16, 17], particularly in the hinterland of Tibet Plateau, where environmental conditions make it difficult to obtain the land-atmosphere interaction observation data. Most studies have focused on Naqu in the hinterland of the Tibetan Plateau [18–20]. So, more observations and experimental studies of the land-atmosphere exchange process need to be conducted in Ali in the hinterland of the Tibetan Plateau.

Based on the above reasons, we used the land-atmosphere interaction land-atmosphere exchange observation data in Ali to analyze the variation characteristics of intraday, daily and seasonal net radiation, latent heat, and heat sensing in 2019. The results

were used to provide data support in the parameterization of land-atmosphere exchange processes in the Tibetan Plateau.

2. Materials and Methods

2.1. Study Area. Ali is located in the hinterland of the Tibetan Plateau, and the land-atmosphere interaction observation station is located in the observation field of Ali Meteorological Bureau (32.49°N, 80.10°E).

The altitude of the station is 4120 meters, the annual average temperature of the station is 0.2°C, the annual mean precipitation is 60–70 mm, and the underlying surface is sandy soil [21]. There is no vegetation cover at any time of year, and the ground surface of the test site is flat and open. The gradient observation data of wind, temperature, humidity, fluxes, and radiation were collected in the observation station [22].

2.2. Data. The radiation data were obtained from a four-component radiometer; the radiation meter model is CM11, primarily used in measuring upwelling shortwave radiation, downwelling shortwave radiation, upwelling longwave radiation, and downwelling longwave radiation, and the net radiation could be calculated by the radiation data. Incident and reflected radiations are measured within 305–2800 nm, and the total measurement band including longwave radiation is 305–50000 nm. We averaged the observations conditionally for various weather conditions. The four-component radiometer was installed 1.5 meters above and facing the underlying surface. Beijing time was used. Spring was defined as from March to May; summer as from June to August; autumn as from September to November; and winter as January, February, and December. The observation period was the whole of 2019, and 30 min mean values were recorded. The mean values of the two preceding and two following days were used to fill up the missing data. Finally, the daily mean and the seasonal and annual variation of net radiation [23] were calculated. The flux data were obtained from the eddy correlation system (Campbell EC150) installed 5.45 meters above and facing the underlying surface. The EddyPro software (v5.1.1) developed by Li-COR (Lincoln, NE, USA) was used to check the observation data quality of the eddy correlation system [24]. If there were less than 15,000 10 Hz data involved in the 30 min flux calculation, the flux data at the corresponding time were eliminated. If the automatic gain control value was greater than a certain threshold, the flux data at the corresponding time were eliminated. The flux data calculated during precipitation were also eliminated. Outliers were eliminated by a variance test [25]. The rejected data were treated as missing data. The missing and rejected flux data were interpolated according to the mean daily variation method [26]. In the flux calculation, the mean time was 30 min.

3. Results

3.1. Intraday Variation Characteristics of Land-Atmosphere Energy Exchanges. Figure 1 shows the intraday variation characteristics of the sensible heat, latent heat, and net radiation fluxes in different seasons in 2019 in Ali. The net radiation in different seasons in Ali was negative at night. After sunrise, the net radiation started to change from negative to positive value and reached a maximum at 15:00, which then decreased rapidly, reaching a minimum value at about 22:00. Gong Yuanfa reported that the net radiation flux in Ali reached a maximum value at about 14-15:00 in the afternoon based on the observed radiation balance data from 1997 to 1998 [19]. Based on the observation data in 2019, it was found that the net radiation flux in Ali reached a maximum value at 15:00. This means that the timing of the mean intraday variation peak of net radiation in Ali in the past 20 years was delayed, although this needs to be verified by more observation data. The occurrence time of the daily maximum value of net radiation in the Ali area showed a very significant delay trend based on the observation data of the Ali Meteorological Bureau from 1993 to 2016 [27]. The maximum intraday variation in the net radiation flux was 529.91 w/m^2, 563.87 w/m^2, 433.60 w/m^2, and 313.64 w/m^2 in spring, summer, autumn, and winter in Ali, respectively. The strongest fluxes were observed in spring and summer, with slightly larger fluxes in summer. This was followed by autumn, with the weakest flux in winter. This indicates that the seasonal variation of net radiation was mainly controlled by the seasonal change of total solar radiation [28].

There was a clear intraday variation in sensible heat in Ali in all seasons, but the intraday variation in summer was significantly larger than in winter. The mean intraday variation in the sensible heat flux in all seasons fluctuated around 0 at night, even becoming negative on some occasions. This was due to the strong ground radiation cooling at night on the plateau when the ground temperature is lower than the air temperature and a downward heat transfer occurs [29]. The maximum intraday variation in the sensible heat flux occurred at 15:30, lagging behind that of net radiation by about 30 min. The maximum intraday variation in the surface heat flux was 210.83 w/m^2, 183.11 w/m^2, 158.32 w/m^2, and 102.05 w/m^2 in spring, summer, autumn, and winter in the region area, respectively. The maximum value occurred in spring and the lowest value occurred in winter, followed by summer and autumn, indicating that the seasonal variation in the sensible heat flux was controlled by the seasonal variation of net radiation and the atmospheric circulation [30].

The phase characteristics of the intraday variation in latent heat flux were the same as those of sensible heat, but the timing of the mean intraday variation peak of the latent heat flux occurred in spring, summer, and winter was earlier than that of the mean daily variation peak of net radiation. The latent heat flux was much smaller than the sensible heat flux at the same time; it indicated that the plateau soil was dry and the differences between ground and air were large. The sensible heat, therefore, played an important role in ground heating. The maximum values of the daily variation in the surface latent heat flux in spring, summer, autumn, and winter were 5.94, 49.41, 11.77, and 8.42 w/m^2, respectively. The maximum value occurred in summer. This was because summer is a rainy season in Ali, with relatively large amounts of precipitation and more net radiation available to drive the phase changes in water [31].

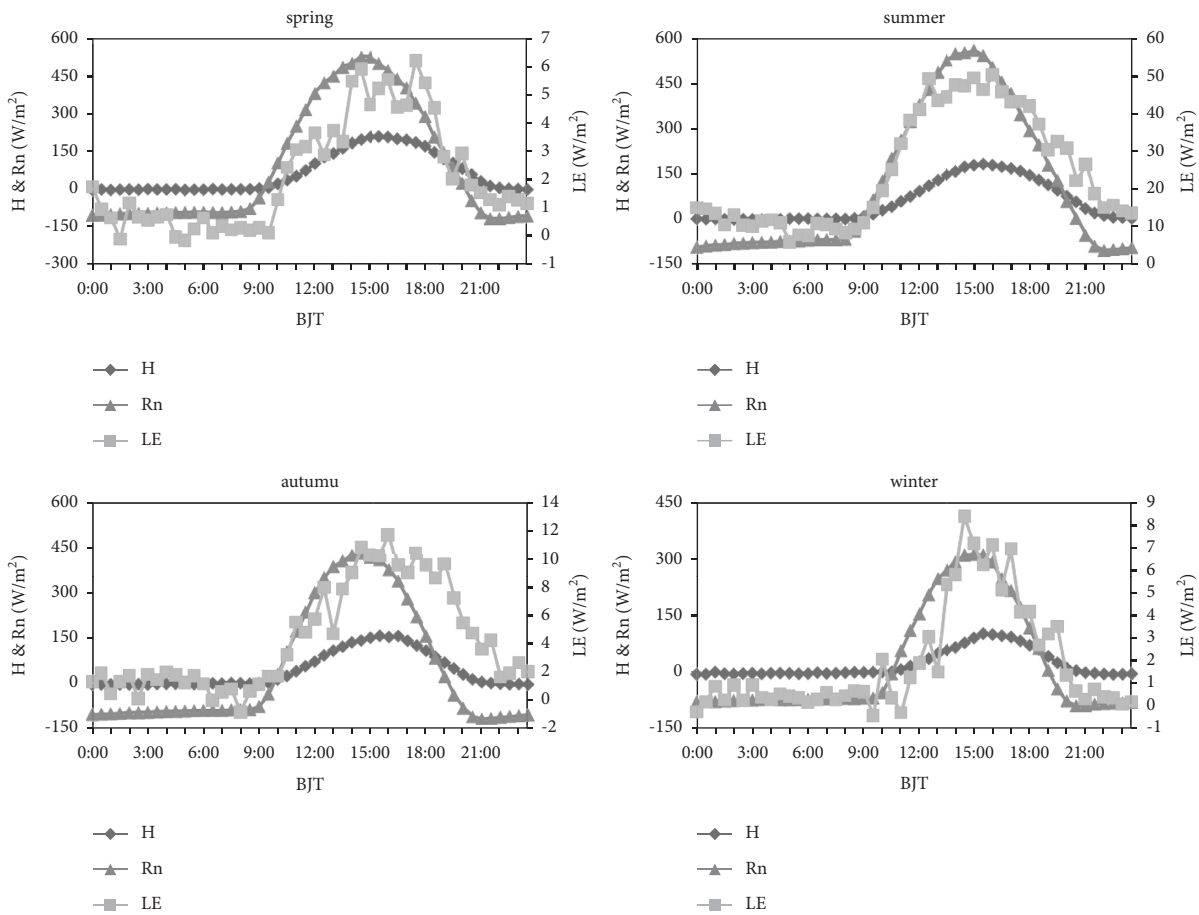


Figure 1: The intraday variation characteristics of the sensible heat, latent heat, and net radiation fluxes.

3.2. Daily Variation Characteristics of Land-Atmosphere Energy Exchanges. Figure 2 shows the daily variation characteristics of sensible heat, latent heat, and net radiation fluxes in Ali in 2019. It shows that the annual variation in net radiation was closely related to atmospheric circulation and climate conditions [9]. The net energy gained by the surface was the largest when the net radiation was the largest during midsummer, which played a decisive role in the formation of the ground heat source and heating to the atmosphere. The phase characteristics of the daily mean sensible heat flux were the same as those of the net radiation flux in the dry season but deviated in the main rainy season (July and August). Due to the extreme drought in Ali, the latent heat flux fluctuated around zero in the dry season, while the daily mean latent heat flux increased rapidly in the rainy season (July and August). The daily mean sensible heat flux was larger than that of the latent heat flux, indicating that the net radiation flux was mainly used for air heating. The daily mean latent heat flux increased rapidly in the rainy season, indicating that most of the net radiation flux was used for the phase transition during water evaporation. The maximum values of the daily mean sensible heat, latent heat, and net radiation fluxes in Ali in 2019 were 99.63, 73.27, and 174.98 w/m^2 on 27 June, 13 August, and 12 August, respectively. The annual mean sensible heat, latent heat, and net radiation fluxes were 46.74, 8.32, and 73.84 w/m^2, respectively. From the annual mean value, it was determined that 63.3% of the net radiation energy received in Ali was used to heat the atmosphere and 11.3% was used for the phase transition during water evaporation.

3.3. Seasonal Variation Characteristics of Land-Atmosphere Energy Exchanges. Figure 3 shows the monthly and seasonal mean variation characteristics of the sensible heat, latent heat, and net radiation fluxes in 2019 in Ali. The seasonal net radiation was 121.10 w/m^2 in summer, while in winter, it was 23.77 w/m^2, indicating that the effect of the seasonal conversion on net radiation was more prominent. The monthly mean net radiation flux in Ali was the largest in August, while Gong Yuanfa showed that the net radiation flux in the Ali area reached a maximum in June according to the radiation balance data from 1997 to 1998 [19]. Based on the observation data from August 2001 to September 2002, Ma Weiqiang found that the maximum net radiation flux in the northern Tibetan Plateau occurred in July [32]. Based on the latest data from 2019, the present study determined that the monthly mean net radiation flux in Ali reached a maximum in August. The maximum monthly mean net radiation flux in the Qiangtang Plateau had a time lag of about 2 months, although this needs to be verified by more observation data.

The monthly mean sensible heat flux in Ali was positive and increased gradually from January to May, reaching a

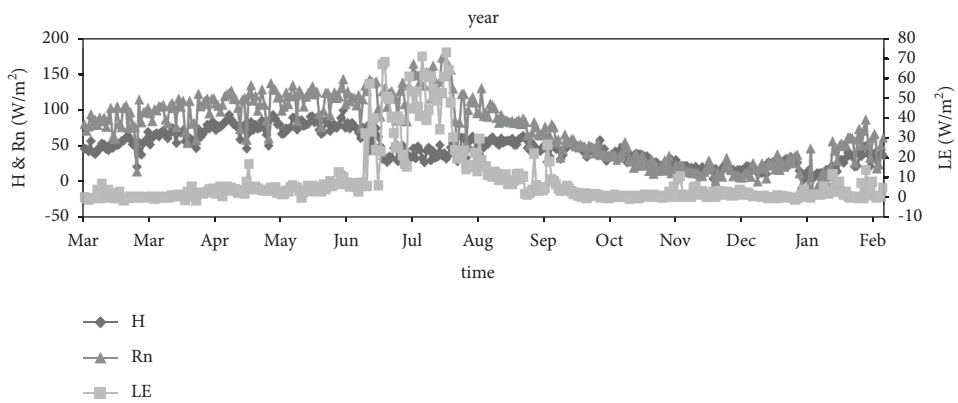


Figure 2: The daily variation characteristics of the sensible heat, latent heat, and net radiation fluxes.

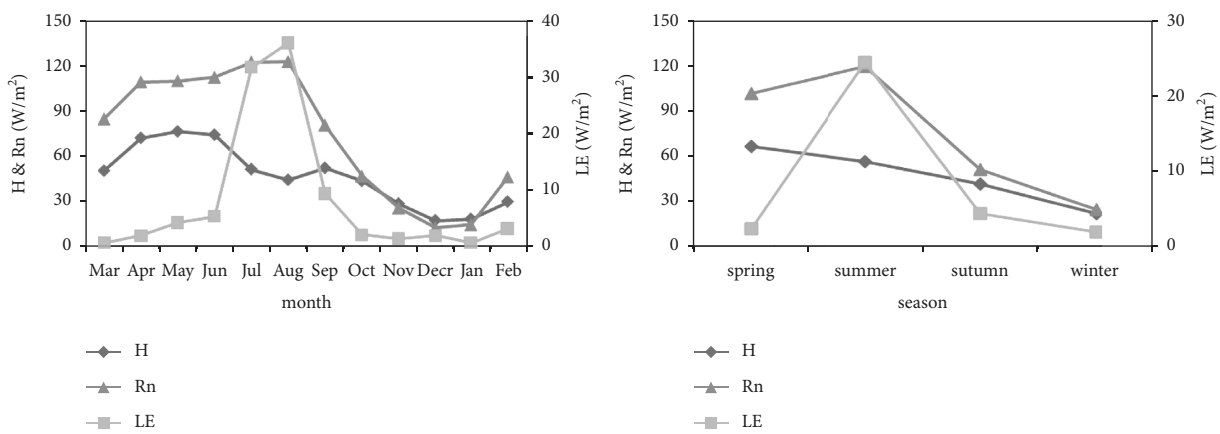


Figure 3: The monthly and seasonal mean variation characteristics of the sensible heat, latent heat, and net radiation fluxes.

maximum value of 76.53 w/m^2. It then decreased gradually until December when it reached a minimum value of 17.47 w/m^2. The maximum latent heat flux was 36.13 w/m^2 in August, while the minimum value was 0.67 w/m^2 in March. The monthly mean latent heat flux was larger than the sensible heat flux in all months. The annual mean sensible heat flux was 5.6 times the latent heat flux, indicating that the climate in the western region was dry and less water was used for evaporation. The heat exchange between the surface and the atmosphere was mainly sensible heat transfer. From the seasonal mean values, the sensible heat flux was the largest in spring and then decreased gradually, while the latent heat and net radiation were the largest in summer. The latent heat flux in the rainy season was affected by the precipitation and was one order of magnitude higher than in the dry season.

4. Discussion

Due to the limited representativeness of single-point observation results and the large range of Ali, the research results in this study can only represent the variation characteristics of the Shiquanhe region of Ali. Ali of Tibetan Plateau has a harsh climate, difficult terrain, and few observation sites. Due to these limitations, there was a lack of early observation data. We have made field observations for several years, but the missing rates of observation data were high in the early stage, and only the observation data in 2019 was relatively complete, so this study analyses land-atmosphere energy exchange characteristics in Ali in 2019. More observation data were needed to explain why the timing of the mean daily variation peak of net radiation in the Ali area region has decreased over the past 20 years and why the month when the maximum value of the monthly mean net radiation occurred has been delayed by about 2 months. It is also necessary to strengthen the interpolation and quality control of the missing data and comprehensively analyze the land and atmosphere energy exchange characteristics and their changing trends in Ali based on more observation data.

5. Conclusions

Based on comprehensive data of the land-atmosphere exchange observation station in Ali in 2019, the characteristics of intraday, daily, and seasonal variation in land-atmosphere energy exchange processes in Ali were analyzed. The followed conclusions were obtained.

(1) The maximum daily variation of net radiation flux in spring, summer, autumn, and winter in Ali was 529.91 w/m^2, 563.87 w/m^2, 433.60 w/m^2, and 313.64 w/m^2, respectively. The maximum daily mean net

radiation flux was 174.98 w/m^2, the mean summer flux was 121.10 w/m^2, the mean winter flux was 23.77 w/m^2, and the annual mean was 73.84 w/m^2.

(2) The maximum daily variation in the surface sensible heat flux in spring, summer, autumn, and winter was 210.83 w/m^2, 183.11 w/m^2, 158.32 w/m^2, and 102.05 w/m^2, respectively. The maximum daily mean sensible heat flux was 99.63 w/m^2, the maximum monthly mean sensible heat flux was 76.53 w/m^2, the minimum monthly mean sensible heat flux was 17.47 w/m^2, and the annual mean sensible heat flux was 46.74 w/m^2.

(3) The maximum daily variation in the latent heat flux in spring, summer, autumn, and winter was 5.94 w/m^2, 49.41 w/m^2, 11.77 w/m^2, and 8.42 w/m^2, respectively. The maximum daily variation in the latent heat flux was 73.27 w/m^2, the maximum monthly mean was 36.13 w/m^2, the minimum monthly mean was 0.67 w/m^2, and the mean annual latent heat flux was 8.32 w/m^2.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors would like to thank the staff of the Institute of Plateau Meteorology, China Meteorological Administration, Chengdu, for their contribution to fieldwork. This research was funded by The Second Tibetan Plateau Scientific Expedition and Research (STEP) Program (2019QZKK0103 and 2019QZKK0105), The Opening Foundation of Plateau Atmosphere and Environment Key Laboratory of Sichuan Province (PAEKL-2020-C7), and The Heavy Rain and Drought-Flood Disasters in Plateau and Basin Key Laboratory of Sichuan Province (SCQXKJYJXMS202116).

References

[1] J. F. Du, W. M. Chen, P. F. Wu, J. P. Zhang, and F. Sun, "GMS data-estimated surface net radiation over east China in summer," *Journal of Nanjing Institute of Meteorology*, vol. 27, no. 5, pp. 674–680, 2004.

[2] Y. Zuo, S. Jiang, S. Wu et al., "Terrestrial heat flow and lithospheric thermal structure in the chagan depression of the yingen-ejinaqi basin, north central China," *Basin Research*, vol. 32, no. 6, pp. 1328–1346, 2020.

[3] A. Z. Wang, J. M. Liu, D. X. Guan, and T. Pei, "Comparison of the measurement and estimate of sensible and latent fluxes over broadleaved Korean pine forest in Changbai Mountain," *Scientia Silvae Sinicae*, vol. 39, no. 6, pp. 20–22, 2003.

[4] H. Wang, Z. Y. Hu, D. L. Li, Y. Z. Zhao, and W. Q. Ma, "Comparative of climatologic characteristics of the surface radiation balance on Dingxin gobi and Zhangye oasis and desert underlaying surfaces in Heihe Watershed, Gansu," *Journal of Glaciology and Geocryology*, vol. 31, no. 3, pp. 464–473, 2009.

[5] P. Yue, S. J. Niu, Q. Zhang, and X. L. Liu, "Surface heat and radiation budget over semi-arid grassland in a clear day and a cloudy day during early spring," *Journal of Desert Research*, vol. 30, no. 6, pp. 1464–1468, 2010.

[6] W. F. Zhou, B. R. Zhou, X. D. Li, F. Li, and I. Wang, "Variation characteristics of radiation budget and its component in the eastern Qinghai-Xizang Plateau," *Plateau Meteorology*, vol. 32, no. 2, pp. 327–333, 2013.

[7] G. L. Ji, L. Z. Lu, and J. J. Zou, "The seasonal variation of surface radiation energy over the northern Tibetan Plateau," *Acta Energiae Solaris Sinica*, vol. 16, no. 4, pp. 340–346, 1995.

[8] Z. Y. Qian, Z. Y. Hu, P. Du, and Y. W. Zhang, "Energy transfer of near surface layer and micrometeorology characteristics in bailuhe area of Qinghai-Xizang Plateau," *Plateau Meteorology*, vol. 24, no. 1, pp. 43–48, 2005.

[9] Y. R. Lu and G. D. Gao, "The radiation balance on Tibetan plateau," *Journal of the Meteorological Sciences*, no. 2, pp. 59–65, 1983.

[10] Z. B. Shen, D. M. Weng, and S. W. Pan, *Survery of the heat source observation experiment over Qinghai-Xizang Tibet// Meteorology Scientific Experiment Treatises about Qinghai-Tibet Plateau*, Science Press, Beijing, China, 1987.

[11] X. D. Xu and L. S. Chen, "Advances of the study on Tibetan Plateau experiment of atmospheric sciences," *Journal of Applied Meteorological Science*, vol. 17, no. 6, pp. 756–772, 2006.

[12] L. Chao, K. Zhang, J. Wang, J. Feng, and M. Zhang, "A comprehensive evaluation of five evapotranspiration datasets based on ground and GRACE satellite observations: implications for improvement of evapotranspiration retrieval algorithm," *Remote Sensing*, vol. 13, no. 12, p. 2414, 2021.

[13] Y. M. Ma, "The observation of water-ice-air-ecosystem interactions and its application over the Tibetan Plateau area," *Strategic Study of CAE*, vol. 14, no. 9, pp. 28–34, 2012.

[14] J. J. Ji and M. Huang, "The estimation of the surface energy fluxes over Tibetan Plateau," *Advances in Earth Science*, vol. 21, no. 12, pp. 1268–1273, 2006.

[15] L. Qiu, L. He, H. Lu, and D. Liang, "Systematic potential analysis on renewable energy centralized co-development at high altitude: a case study in Qinghai-Tibet plateau," *Energy Conversion and Management*, vol. 267, Article ID 115879, 2022.

[16] P. B. C. Ren, F. Sigernes, and Y. Gjessing, "Ground-based measurements of solar ultraviolet radiation in Tibet: preliminary results," *Geophysical Research Letters*, vol. 24, no. 11, pp. 1359–1362, 1997.

[17] A. Dahlback, N. Gelsor, J. J. Stamnes, and Y. Gjessing, "UV measurements in the 3000-5000m altitude region in Tibet," *Journal of Geophysical Research*, vol. 112, no. D9, Article ID D09308, 2007.

[18] A. Numaguti, L. P. Liu, and L. D. Tian, "Proceedings of the 1st international workshop on GAME/Tibet, Chinese academy of sciences and Japanese national committee for GAME," *Xi'an*, vol. 69-81, pp. 99–102, 1999.

[19] Y. F. Gong, T. Y. Duan, L. X. Chen, and J. He, "The variation characteristics of radiation budget components of the western Tibetan Plateau in 1997/1998," *Acta Meteorologica Sinica*, vol. 63, no. 2, pp. 225–235, 2005.

[20] R. S. Wu and Y. M. Ma, "Comparative analyses on radiation characteristics in different areas over the Tibetan Plateau," *Plateau Meteorology*, vol. 29, no. 2, pp. 251–259, 2010.

[21] W. Ge, H. Lin, and T. Xinying, "Surface radiation characteristics of the Ali area, northern Tibetan plateau," *Environmental Research Communications*, vol. 3, no. 4, Article ID 041006, 2021.

[22] J. Sharif, M. K. Rafiq, M. T. Rafiq et al., "Climate change perceptions and adaptive actions by pastoral community on the Tibetan Plateau, China," *Applied Ecology and Environmental Research*, vol. 17, no. 4, pp. 7987–8009, 2019.

[23] W. Huang, Y. Zhao, C. Sun, H. Wang, and X. Wang, "Climate modulation of summer rainstorm activity in eastern China based on the Tibetan Plateau spring heating," *Arabian Journal of Geosciences*, vol. 13, no. 3, p. 126, 2020.

[24] T. Foken and B. Wichura, "Tools for quality assessment of surface based flux measurements," *Agricultural and Forest Meteorology*, vol. 78, no. 1-2, pp. 83–105, 1996.

[25] D. Vickers and L. Mahrt, "Quality control and Flux sampling problems for tower and aircraft data," *Journal of Atmospheric and Oceanic Technology*, vol. 14, no. 3, pp. 512–526, 1997.

[26] E. Falge, D. D. Baldocchi, R. J. Olson et al., "Gapfilling strategies for defensible annual sums of net ecosystem exchange," *Agricultural and Forest Meteorology*, vol. 107, no. 1, pp. 43–69, 2001.

[27] G. Wang, P. li, and S. J. Wang, "Variation characteristics of total radiation and net radiation in Ali in recent 30 years," *Plateau and Mountain Meteorology Research*, 2022.

[28] Y. M. Ma, J. M. Wang, and W. Liu, "The study of the characteristics of both the atmospheric turbulence structure and the transfer in the lower layer of the atmosphere above the nansha islands area," *Chinese Journal of Atmospheric Sciences*, vol. 21, no. 3, pp. 357–365, 1997.

[29] W. Q. Ma, Y. M. Ma, M. S. Li, Y. Z. Zhao, F. L. Sun, and M. H. Song, "Analyses on seasonal variation characteristics of surface energy in the Northern Tibetan Plateau and arid region of Northwest China," *Acta Energiae Solaris Sinica*, vol. 28, no. 8, pp. 922–928, 2007.

[30] S. Feng, H. Lu, P. Tian et al., "Analysis of microplastics in a remote region of the Tibetan Plateau: implications for natural environmental response to human activities," *Science of the Total Environment*, vol. 739, Article ID 140087, 2020.

[31] P. Tian, H. Lu, W. Feng, Y. Guan, and Y. Xue, "Large decrease in streamflow and sediment load of Qinghai-Tibetan Plateau driven by future climate change: a case study in Lhasa River Basin," *Catena*, vol. 187, Article ID 104340, 2020.

[32] W. Q. Ma, Y. M. Ma, and M. S. Li, "Seasonal variation on land surface energy budget and energy balance components in the Northern Tibetan Plateau," *Journal of Glaciology and Geocryology*, vol. 27, no. 5, pp. 673–679, 2005.

Hindawi
Advances in Meteorology
Volume 2022, Article ID 5775424, 8 pages
https://doi.org/10.1155/2022/5775424

Research Article

The Influence of Rainfall and Evaporization Wetting-Drying Cycles on the Slope Stability

Ya Zhao [1,2]

[1]*School of Civil Engineering, Changsha University of Science & Technology, Changsha 414000, China*
[2]*School of Management, Hunan University of Information Technology, Changsha 414000, China*

Correspondence should be addressed to Ya Zhao; zhaoya900803@163.com

Received 9 June 2022; Revised 24 June 2022; Accepted 25 June 2022; Published 9 July 2022

Academic Editor: Upaka Rathnayake

The decay of soil strength and the change of soil infiltration characteristics caused by the dry and wet cycle effect generated by the rainfall-evaporation process are important factors that induce slope instability. How to consider the effect of soil strength decay and water-soil characteristic curve hysteresis effect on transient stability change of slope is the key to solve this problem. In this paper, transient stability analysis of slopes considering soil strength decay and water-soil characteristic curve hysteresis is carried out based on Geo-Studio. The results of the study showed that the change of transient safety factor of the slope caused by rainfall-evaporation dry and wet cycle process has an overall decreasing trend and the safety factor decreased by 43% compared to the initial state. The seepage characteristics of the rainfall-evaporation dry-wet cycle have certain regularity. The location of slope measurement points has a greater influence on the magnitude of the pore pressure change: foot of slope > middle of slope > top of slope. Also, there is a significant response hysteresis in the change of pore pressure with increasing depth at the same location. The rainfall intensity has a certain influence on the change of slope safety factor, but its influence is not obvious when the rainfall intensity exceeds a certain amount.

1. Introduction

The soil strength decay and infiltration characteristics changes caused by the wet and dry cycles lead to the destabilization damage of slope works under the rainfall conditions, and the losses incurred in the actual projects are great [1–4]. Therefore, it is necessary to analyze the influence of the dry and wet cycle process formed by rainfall and evaporation on the stability of slopes.

The wet and dry cycle affects the properties of the soil in two main ways. First, dry and wet cycles will lead to the decay of soil strength [5–9]. Second, the wet and dry cycle will lead to changes in the soil and water characteristics curve [10–13]. For example, Yang and Xiao [9], Xu et al. [8], and Li et al. [7] have studied the strength decay laws of expansive soils and clays under the action of dry and wet cycles, respectively. They concluded that the cohesive force of the soil would decrease significantly with the increase of the number of wet and dry cycles. The reduction in the angle of internal friction is relatively insignificant. The cohesion and internal friction angle will be both stable after six wet and dry cycles. Zhang et al. [12]; on the other hand, investigated the changes in soil-water characteristic curves caused by dry and wet cycles. The calculation method for predicting soil and water characteristic curves is also proposed. In the study of slope stability, Rahardjo et al. [14] investigated the relationship between rainfalls; Satyanaga and Rahardjo [15] studied the unsaturated soil mechanical properties on slope stability; Rahardjo and Satyanaga [16] monitored the slope stability by slope instrumentation. Numerical simulation of slope stability considering the effect of wet and dry cycles has also been studied by many scholars. For example, Cui et al. [17] and Li [18] investigated the effect of dry and wet cyclic action on the stability of swelling soil slopes and red clay slopes by finite element, respectively. Gao [19], Zhou et al. [20], and Li and Tang [21] further explored the effect of wet and dry cyclic action on the stability of expansive soil slopes. However, the above finite element

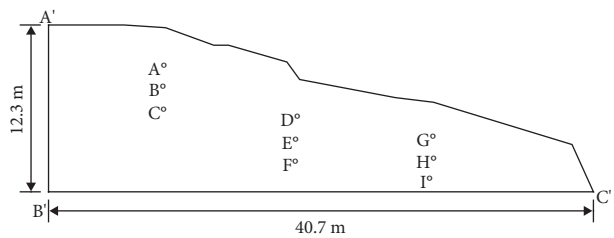


Figure 1: Computational model section.

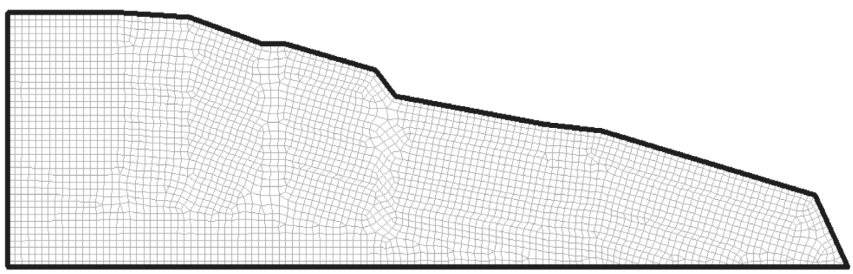


Figure 2: Model grid.

Table 1: Physical and mechanical parameters of soil.

Parameters	Dry density/(g·cm^{-3})	Poisson's ratio	Effective cohesion /(kPa)	Effective angle of friction/(°)	Permeability coefficient/(m·s^{-1})
Initial value	1.62	0.334	51	18.7	9.2×10^{-8}

Table 2: Material strength at different cycle times.

Number of cycles	0	1	2	3	4	5	6
Effective cohesion	51.0	40.1	34.9	31.5	29.2	27.4	26.0
Effective angle of internal friction	18.7	17.4	16.7	16.2	15.8	15.4	15.2

study only considered the effect of dry and wet cycles on the strength decay of the soil without considering the effect of dry and wet cycles on the infiltration characteristics of the soil. There are few reports on the effect of transient stability of slopes under the influence of dry and wet cyclic effects on soil strength and permeability characteristics in a comprehensive manner.

Therefore, in this paper, in order to comprehensively study the effect of wet and dry cycle effect on slope stability. Taking a slope project in Changsha as an example, numerical simulation analysis is carried out to consider the influence law of soil strength change and water-soil characteristic curve change on transient stability of the slope under the influence of rainfall and evaporation effect [22–24]. Subsequent experiments will be conducted using slope monitoring instruments for further research.

2. Finite Element Model Creation

2.1. Computational Models and Boundary Conditions. A typical section of a side slope project in Changsha is chosen as an example. The computational model is shown in Figure 1. The length of the side slope profile is 40.7 m. The slope height is 12.3 m.

The computational model meshing is shown in Figure 2. The mesh size is 0.1 m. The model is divided into 3341 points and 3217 units. The boundary conditions are set in the following manner. $A'B'$ and $B'C'$ are set to the impermeable boundary. The slope is set as the rainfall infiltration boundary and evaporation boundary. Also, the effect of the groundwater level is not considered. The calculation model used in this paper uses the SEEP/W and VADOSE/W modules of GEO-Studio to set the rainfall boundary and evaporation boundary parameters. After the boundary parameters are set, the slope stability coefficients of the rainfall-evaporation process can be obtained through the coupling calculation of the two modules. At the same time, different material parameters are set to simulate the changes of soil parameters for different cycle times. Thus, the dynamic change process of slope stability coefficient under the influence of dry and wet cycles is obtained.

2.2. Calculation Parameters. The strength parameters of the slope soil are determined by tests to determine the initial values. The specific parameter values are shown in Table 1. Numerous studies [9, 25, 26] have shown that the strength of the soil decays as a result of the wet and dry cycle and that the

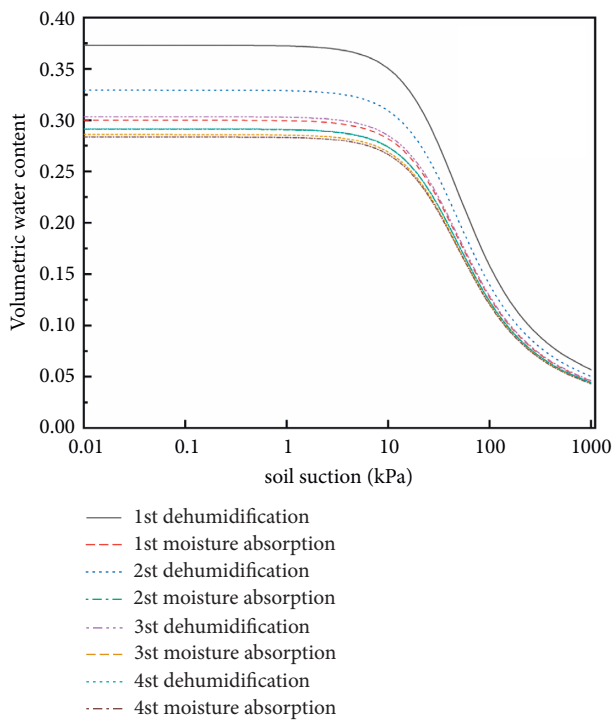


Figure 3: Water-soil characteristic curves for different number of cycles.

Table 3: Climate parameters.

Parameter name	Rainfall intensity /mm·d^{-1}	Duration of rainfall/d	Duration of Evaporation/d	Temperature/ °C	Relative humidity/%	Radiation/MJ/ d^{-1}·m^2
Rainfall	50	4	—	—	—	—
Evaporation	—	—	6	35	60	19

decay process is generally expressed as an exponential change. In this paper, the strength of the soil is fitted exponentially for different number of cycles according to the research results of Li et al. [7]. The soil strength values for different number of cycles were also calculated from the initial strengths in Table 1. The specific parameters are shown in Table 2. The unsaturated shear strength property used in this study in the analytical setup is the phi b angle.

The effect of wet and dry cycles on soil properties will have a hysteresis effect on the water-soil characteristic curve of the soil in addition to strength decay. This paper uses the results of Zhang et al.'s [12] research to predict the hysteresis changes in the soil and water characteristics curve after dry and wet cycles. The soil and water characteristic curves used for different number of cycles are shown in Figure 3. The change in the soil and water characteristics curve is predicted to stabilize after the 4th wet and dry cycle. Therefore, this paper only considers the effect of four wet and dry cycles on the soil and water characteristic curves [27–30]. Both cycles

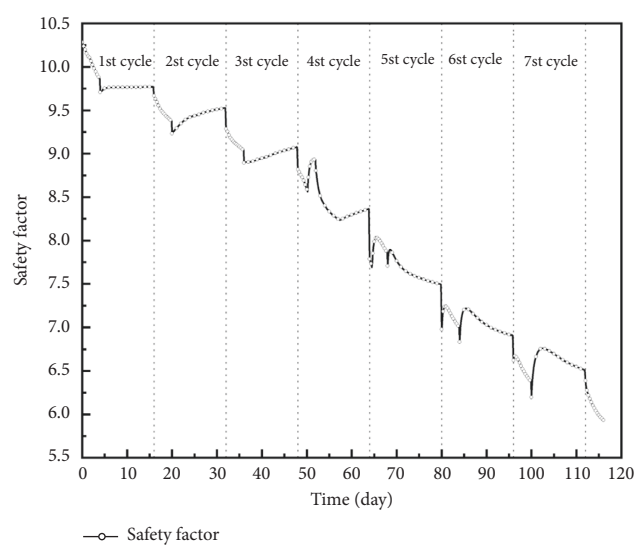


Figure 4: Slope safety coefficient change curve with time.

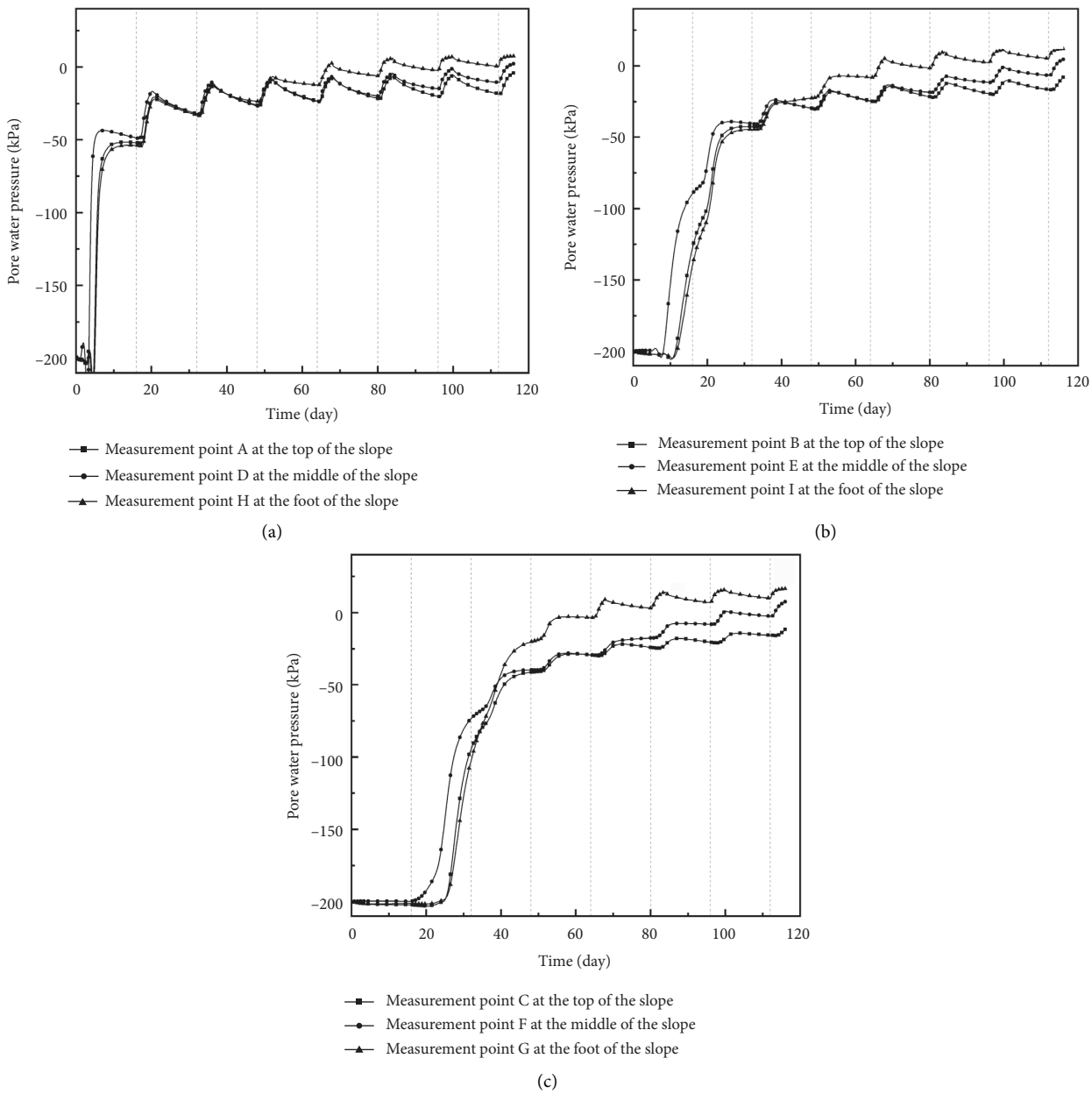


Figure 5: Variation curve of pore water pressure at different depths. (a) Pore water pressure variation curve at 3 m depth. (b) Pore water pressure variation curve at 4 m depth. (c) Pore water pressure variation curve at 6 m depth.

after the 4th are replaced by the soil and water characteristic curve of the 4th cycle. In this paper, the unsaturated permeability characteristics of the studied soils are described by means of water-soil characteristic curves. The relevant research results of Kristo et al. [31] are also referenced.

2.3. Analysis Settings. This paper uses the SEEP/W module of GEO-Studio coupled with the VADOSE/W module for seepage and evaporation analysis. According to meteorological data for Changsha, 4 days of rainfall and 6 days of evaporation are used as a dry and wet cycle. Specific climatic parameters are shown in Table 3. The analysis time step is set to 2 hours in one step. In order to consider the hysteresis variation of soil strength decay and soil-water characteristic curves, this paper simulates by replacing materials in layers and defining the properties of different materials. The depth of atmospheric influence is assumed to be 6 m [20] and is considered to vary exponentially with the number of wet and dry cycles.

3. Calculation Results and Analysis

3.1. Stability Analyses. The effects of different numbers of wet and dry cycles on the stability coefficients of slopes are shown in Figure 4. From Figure 4, it can be seen that the 1st

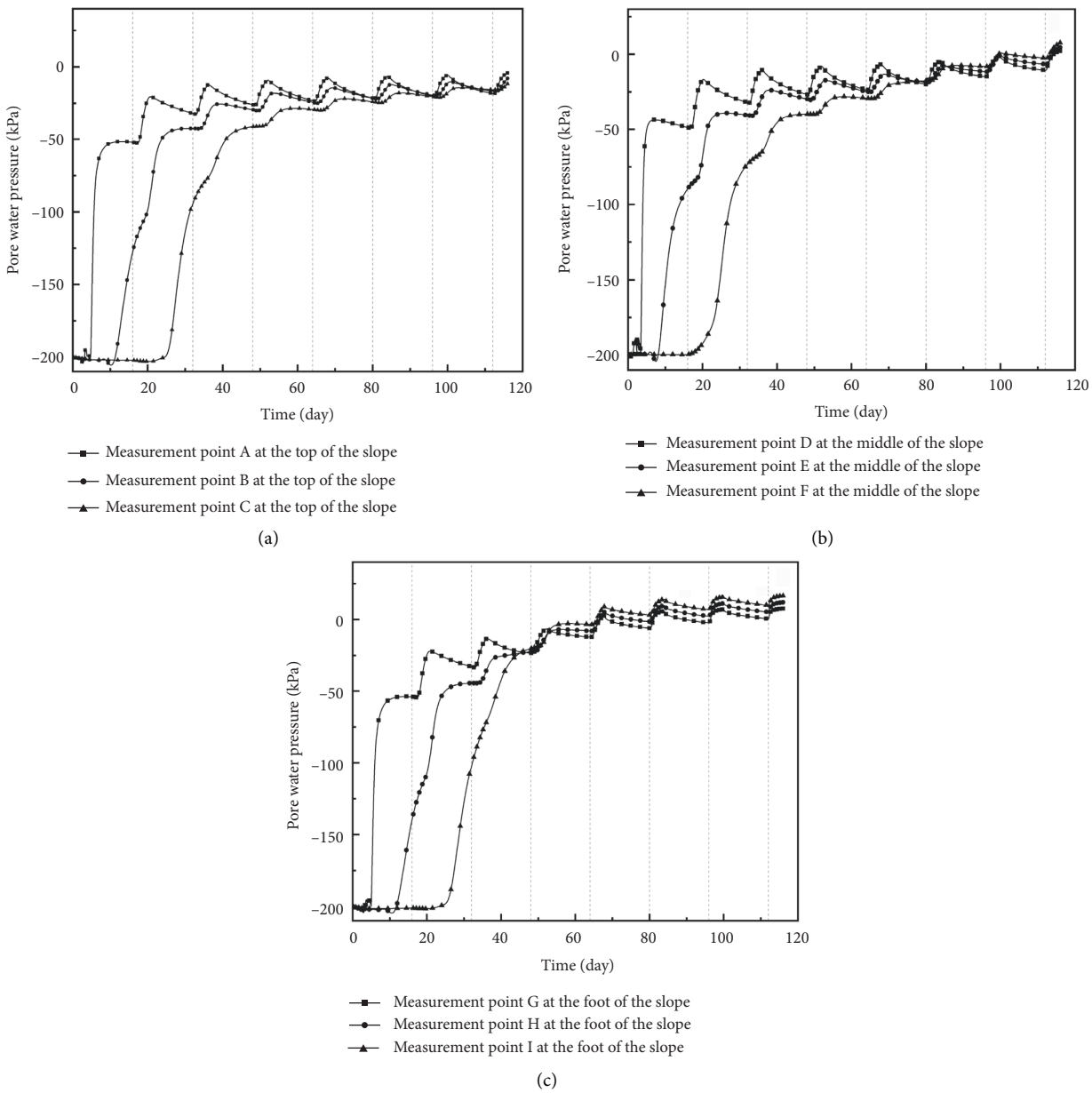


Figure 6: Variation curve of pore pressure at different positions. (a) Pore water pressure variation curve at the top of the slope. (b) Pore water pressure variation curve at the middle of the slope. (c) Pore water pressure variation curve at the foot of the slope.

cycle evaporation process has no effect on the safety factor. The 2nd–4th evaporation process has a clear tendency to increase its safety factor. Also, the safety factor in the evaporation phase of the 5th–7th cycle still shows a decreasing trend after the rebound. The reason for this result may be due to setting the initial pore pressure value of the slope to −200 kPa without considering the effect of groundwater level. Therefore, evaporation after the first rainfall infiltration is small and makes the change in the safety factor for the first evaporation phase relatively insignificant. The 2nd–4th cycles, on the other hand, are based on the first cycle, so the evaporation is no longer affected by the initial pore pressure. The change in pore pressure resulting from the evaporation phase increases the strength of the soil and thus the safety factor. The reason for the continuous decrease of the safety factor in the evaporation phase during the 5th–7th cycles after the rebound may be due to the increasing depth of rainfall infiltration as the number of cycles increases. The water content of the deep soil has a large increase from the initial state. However, the variation of water content in the evaporation phase is gradually weakened with the increase of depth. Therefore, in the evaporation phase instead, the safety factor will rebound a certain amount and then show a decreasing trend.

In general, the overall slope safety factor shows a decreasing trend with the increase in the number of wet and

dry cycles. Also, relative to the initial safety factor of 10.28, the safety factor after the 7th cycle was only 5.9 with reductions of nearly 43%.

3.2. Seepage Analyses. A total of nine measurement points were set up within the slope in order to monitor the changes of seepage process during the rainfall-evaporation cycle. The locations of the measurement points are shown in Figure 1. Figure 5 shows the variation curve of pore water pressure at the same depth at different locations. It can be seen from Figure 5(a) that the variation of pore water pressure at the top, middle, and foot of the slope in the 1st to 3rd cycles is almost the same. The pore water pressure at the foot of the slope has a large deviation from the top and foot of the slope at the beginning of the 4th cycle. The position in the middle of the slope, on the other hand, is deviated from the foot of the slope in the evaporation phase of the 6th cycle. Also, on the magnitude of the change in values, the foot of the slope > the middle of the slope > the top of the slope.

It can be seen from Figure 5(b) that the variation of pore water pressure at 4 m depth is significantly different from that at 3 m depth. In the 3rd cycle, the pore water pressure at the foot of the slope starts to deviate from that in the middle and foot of the slope. It can also be seen that the pore water pressure at depth 4 m stabilize only at the end of the 2nd cycle. Also, the 3 m depth is basically stabilized at the end of the 1st cycle. However, in terms of the magnitude of pore water pressure change, it is also consistent with the foot of slope > middle of slope > top of slope.

From Figure 5(c), it can be found that the pore water pressure variation at 6 m depth is basically unchanged after the completion of the 1st cycle. It is during the 2nd cycle phase that significant changes start to occur. This indicates that there is a significant hysteresis effect of pore water pressure change in the deep soil relative to the surface soil.

It is easy to see that the pore water pressure changes in different depths of the soil layer are not the same, but the rule of foot of the slope > middle of the slope > top of the slope is the same in terms of the magnitude of change. At the same time, the hysteresis effect of depth on the pore water pressure change gradually increases with depth.

To further investigate the effect of depth on the variation of pore water pressure, the variation curves of pore water pressure at different depths at the same location are given in Figure 6. The hysteresis effect of depth on the variation of pore water pressure can be seen more clearly in Figure 6. Also, this phenomenon was present in different locations. At the same time, it can be seen from the analysis of Figure 6 that the variation of pore water pressure at different depths at the top of the slope converges basically only in the 7th cycle. The middle of the slope, on the other hand, is largely convergent in the evaporation phase of the 5th cycle. The foot of the slope is converging at the end of the 3rd evaporation. This indicates that there is a significant difference in the time for the pore pressure change to stabilize at different locations. The change at the foot of the slope will stabilize at an earlier time, followed by the mid-slope position and finally the top-of-slope position.

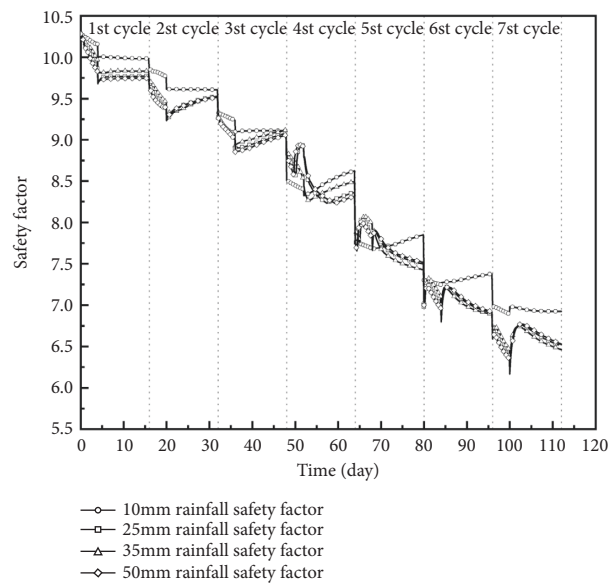


Figure 7: Variation curve of safety factor for different rainfall intensity.

3.3. Effect of Rain Intensity Variation on Stability. In order to explore the influence of rainfall intensity on slope stability coefficient, four different rainfall intensities (10 mm/d, 25 mm/d, 35 mm/d, 50 mm/d) are set for analysis in this paper. The calculation results are shown in Figure 7. According to Figure 7, it can be seen that 10 mm/d rain intensity has no significant change in the safety factor during the evaporation phase of the 1st–3rd cycles. This is mainly due to the low intensity rainfall resulting in a low impact of infiltration on the initial pore water pressure, which in turn allows the evaporation phase to be affected. Also, comparing the change of the safety factor for different intensity of rainfall shows that the overall safety factor for 10 mm/d rainfall intensity is higher than the rest of the rainfall intensity. Also, the difference in the variation of the safety factor for rainfall intensity bars above 25 mm/d is not significant. This indicates that although the effect of rainfall intensity on the change of the safety factor exists, there will be no significant difference in the change of the safety factor once the rainfall intensity exceeds a certain value. The reason for this phenomenon may be that the rainfall intensity is no longer controlled by the rainfall intensity but the infiltration capacity of the soil after exceeding a certain size, which leads to a convergence of the change in the safety factor.

4. Conclusion

In this paper, the transient stability change process of soil strength decay and water-soil characteristic curve hysteresis effect during rainfall-evaporation dry and wet cycles is considered by numerical simulation. Based on the results of the analysis the following conclusions were obtained.

(1) With the increase of the number of wet and dry cycles, the strength parameter of the slope soil will be decayed continuously. Also the water-soil

characteristic curve will have a hysteresis effect. The superposition of two factors will lead to an overall decreasing trend of slope safety factor and the safety factor decreased by nearly 43%.

(2) The seepage characteristics of rainfall-evaporation dry-wet cycle processes have obvious regularity. The location of the measurement point on the slope has a great influence on the magnitude of pore water pressure change: foot of the slope > middle of the slope > top of the slope. Also, there is a significant response hysteresis of pore pressure change with increasing depth at the same location.

(3) There is a certain influence of rain intensity on the slope stability during the wet and dry cycle. However, their effects gradually diminish and converge with the increase of rainfall intensity.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

The work described in this paper was fully supported by a grant from the Hunan Postgraduate Research Innovation Project "Experimental study on rainwater infiltration and stable evolution of steep slopes with expansive soil (No. CX20200841)" and Scientific Research Subject of Hunan Provincial Education Department "Experimental study on the change of mechanical properties of root-soil composite due to rainwater infiltration (No. 18C1578)."

References

[1] M. Liu, J. Liu, and F. Tan, "Field monitoring study of ditch slope treatment in Xinxiang swelling rock test section," *Journal of Yangtze River Scientific Research Institute*, vol. 28, no. 10, pp. 148–155, 2011.

[2] C. Nie, "The danger of swelling soil to highway and treatment," *Shanxi Science & Technology of Communications*, vol. 5, no. 5, pp. 6-7, 2002.

[3] L. Yin, L. Wang, W. Huang et al., "Haze grading using the convolutional neural networks," *Atmosphere*, vol. 13, no. 4, p. 522, 2022.

[4] L. Yin, L. Wang, B. D. Keim, K. Konsoer, and W. Zheng, "Wavelet analysis of dam injection and discharge in Three Gorges Dam and reservoir with precipitation and river discharge," *Water*, vol. 14, no. 4, p. 567, 2022.

[5] X. Chen, Q. Quan, K. Zhang, and J. Wei, "Spatiotemporal characteristics and attribution of dry/wet conditions in the Weihe River Basin within a typical monsoon transition zone of East Asia over the recent 547 years," *Environmental Modelling & Software: With Environment Data News*, vol. 143, Article ID 105116, 2021.

[6] Z. Chen, Z. Liu, L. Yin, and W. Zheng, "Statistical analysis of regional air temperature characteristics before and after dam construction," *Urban Climate*, vol. 41, Article ID 101085, 2022.

[7] X. Y. Li, H. B. Hu, and W. Guo, "Dry-wet circulation intensity attenuation law of subgrade under the condition of clay," *Highway Engineer*, vol. 164, no. 1, pp. 150–152, 2014.

[8] B. Xu, Z. Yin, and S. Liu, "Experimental study of the factors and laws affecting the strength of expansive soil," *Rock and Soil Mechanics*, vol. 32, no. 1, pp. 44–50, 2011.

[9] H. Yang and D. Xiao, "Effect of wet and dry cycle effect on shear strength of expansive soil," *Journal of Changsha University of Science and Technology (Natural Science)*, vol. 2, no. 2, pp. 1–5, 2005.

[10] X. Fang, Q. Wang, J. Wang, Y. Xiang, Y. Wu, and Y. Zhang, "Employing extreme value theory to establish nutrient criteria in bay waters: a case study of Xiangshan Bay," *Journal of Hydrology*, vol. 603, 2021.

[11] J. Xu, L. Y. Zhou, K. Hu, Y. Li, X. Zhou, and S. Wang, "Influence of wet-dry cycles on uniaxial compression behavior of fissured loess disturbed by vibratory loads," *KSCE Journal of Civil Engineering*, vol. 26, pp. 2139–2152, 2022.

[12] J. R. Zhang, X. Qiang, and D. Sun, "Simulation of soil-water characteristic curves during drying and wetting cycles," *Rock and Soil Mechanics*, vol. 35, no. 3, pp. 689–695, 2014.

[13] K. Zhang, M. H. Shalehy, G. T. Ezaz, A. Chakraborty, K. M. Mohib, and L. Liu, "An integrated flood risk assessment approach based on coupled hydrological-hydraulic modeling and bottom-up hazard vulnerability analysis," *Environmental Modelling & Software: With Environment Data News*, vol. 148, Article ID 105279, 2022.

[14] H. Rahardjo, A. Satyanaga, and E. C. Leong, "Effects of rainfall characteristics on the stability of tropical residual soil slope," *E3S Web Conferences*, vol. 9, Article ID 15004, 2016.

[15] A. Satyanaga and H. Rahardjo, "Role of unsaturated soil properties in the development of slope susceptibility map," *Proceedings of the Institution of Civil Engineers—Geotechnical Engineering*, vol. 175, no. 3, pp. 276–288, 2022.

[16] H. Rahardjo and A. Satyanaga, "Sensing and monitoring for assessment of rainfall-induced slope failures in residual soil," *Proceedings of the Institution of Civil Engineers—Geotechnical Engineering*, vol. 172, no. 6, pp. 496–506, 2019.

[17] W. Cui, Z. Zhang, and S. Yan, "Dry and wet cyclic properties of expansive soil and their application in slope stability analysis," *Journal of Water Resources and Architectural Engineering*, vol. 8, no. 5, pp. 24–27, 2010.

[18] Z. Li, *Study on the Stability Evaluation Method of Red Clay Slopes under Dry and Wet Cycles*, Guizhou University, Guiyang, China, 2018.

[19] Y. Gao, *Numerical Simulation Analysis of Unsaturated Soil Slope Response under Dry and Wet Cycles*, Harbin Institute of Technology, Harbin, China, 2011.

[20] J. Zhou, H. Xu, and W. Hu, "Impact of wetting-drying cycle effects on stability of expansive soil slopes," *Proceedings of the 11th National Symposium on Numerical Analysis and Analytical Methods in Geotechnics*, vol. 35, no. S2, pp. 152–156, 2013.

[21] P. Y. Li and Z. J. Tang, "Influence analysis of dry-wet circulation effect on the stability of expansive soil slope," *Shanxi Science & Technology of Communications*, vol. 6, no. 6, pp. 18–22, 2018.

[22] R. Ahmadi, M. El May, and M. Dlala, "Ultimate slope design in open pit phosphate mine using geological and

geomechanical analysis: case study of Jebel Jebbeus," *Arabian Journal of Geosciences*, vol. 12, 2019.

[23] X. Shen, B. Liu, M. Henderson, L. Wang, M. Jiang, and X. Lu, "Vegetation greening, extended growing seasons, and temperature feedbacks in warming temperate grasslands of China," *Journal of Climate*, vol. 3, pp. 1–51, 2022.

[24] X. Shen, Y. Liu, B. Liu et al., "Effect of shrub encroachment on land surface temperature in semi-arid areas of temperate regions of the Northern Hemisphere," *Agricultural and Forest Meteorology*, vol. 320, Article ID 108943, 2022.

[25] K. S. Chen, "Study of shear strength characteristics for red clay under wet and dry cycles," *Highway*, vol. 61, no. 2, pp. 45–49, 2016.

[26] Y. Liu, S. Zhu, and C. Wang, "The investigation of the effect of dry density on the development and strength of fractures in expansive soil under dry and wet cycles," *Soil Engineering and Foundation*, vol. 33, no. 5, pp. 571–574, 2019.

[27] W. Liu and Z. Zhang, "Experimental characterization and quantitative evaluation of slaking for strongly weathered mudstone under cyclic wetting-drying condition," *Arabian Journal of Geosciences*, vol. 13, no. 2, 2020.

[28] S. Wang, K. Zhang, L. Chao et al., "Exploring the utility of radar and satellite-sensed precipitation and their dynamic bias correction for integrated prediction of flood and landslide hazards," *Journal of Hydrology (Amsterdam)*, vol. 603, Article ID 126964, 2021.

[29] K. Zhang, A. Ali, A. Antonarakis et al., "The sensitivity of North American terrestrial carbon fluxes to spatial and temporal variation in soil moisture: an analysis using radar-derived estimates of root-zone soil moisture," *Journal of Geophysical Research Biogeosciences*, vol. 124, no. 11, pp. 3208–3231, 2019.

[30] K. Zhang, S. Wang, H. Bao, and X. Zhao, "Characteristics and influencing factors of rainfall-induced landslide and debris flow hazards in Shaanxi Province, China," *Natural Hazards and Earth System Sciences*, vol. 19, no. 1, pp. 93–105, 2019.

[31] C. Kristo, H. Rahardjo, and A. Satyanaga, "Effect of hysteresis on the stability of residual soil slope," *International Soil and Water Conservation Research*, vol. 7, no. 3, pp. 226–238, 2019.

Hindawi
Advances in Meteorology
Volume 2022, Article ID 3336257, 14 pages
https://doi.org/10.1155/2022/3336257

Research Article

Evaluation of Hydropower Generation and Reservoir Operation under Climate Change from Kesem Reservoir, Ethiopia

Kinfe Bereda Mirani,[1] **Mesfin Amaru Ayele**,[2] **Tarun Kumar Lohani**,[2] **and Tigistu Yisihak Ukumo**[2]

[1]*Faculty of Water Supply and Environmental Engineering, Arba Minch University, Arba Minch, Ethiopia*
[2]*Faculty of Hydraulic and Water Resources Engineering, Arba Minch University, Arba Minch, Ethiopia*

Correspondence should be addressed to Tarun Kumar Lohani; tarun.kumar@amu.edu.et

Received 21 March 2022; Revised 19 May 2022; Accepted 23 May 2022; Published 22 June 2022

Academic Editor: Upaka Rathnayake

Climate changes significantly cause the precipitation deficiency and in turn reduce the inflow amount in reservoir affecting hydroelectric power generation. The primary objective of this study was to evaluate hydropower generation and reservoir operation under climate change from Kesem reservoir. Recent Representative Pathway (RCP) scenarios were used to evaluate the impact of climate change on power generation. Power transformation equation and variance scaling approach were amalgamated to adjust the bias correction of precipitation and temperature, respectively. Bias, root mean square error, and coefficient of variation were used to check the accuracy of projected rainfall. The base and future precipitation, temperature, and evaporation trend was analysed using the Mann–Kendall test. The flow calibration and validation were carried out by the Hydrologic Engineering Center-Hydrologic Modelling System (HEC-HMS), and hydropower generation was evaluated with reservoir simulation model (MODSIM 8.1) under climate scenarios. The performance of the model was found good with Nash–Sutcliffe coefficient (NSE) of 0.72 and coefficient of determination (R^2) of 0.73 for calibration and NSE of 0.74 and R^2 of 0.75 for validation. Projected future climate scenarios predicted increasing and decreasing trend of temperature and precipitation, respectively. For RCP4.5 climate scenario, the average energy generation is likely to decrease by 0.64% and 0.82% in both short-term (2021–2050) and long-term (2051–2080), respectively. In case of RCP8.5 climate scenario, the average energy generation will be decreased by 1.06% and 1.35% for short-term and long-term, respectively. Remarkable reduction of energy generation was revealed in RCP8.5 with relation to RCP4.5 scenario. This indicates that there will be high energy fluctuation and decreasing trend in the future energy generation. The research finding is crucial for decision-makers, power authorities, governmental and nongovernmental organizations, and watershed management agencies to take care for sustainability in the future hydropower generation in the Kesem reservoir.

1. Introduction

The well-known impacts of global warming on the water recourses system include changes in the hydrologic cycle and the water availability in the reservoir [1–3]. The variations in the hydrologic system are also articulated by increasing/decreasing frequencies of streamflow and reservoir storage capacity. Hydrologic risks and the unusual frequency of water catastrophe in turn affect the country's economic activities [4].

In electricity generation, hydropower contributes about 16% of energy which is more than other renewable electricity sources in the world [5]. About 200 Giga Watts (GW) of hydroelectricity can be explored from runoff river mini hydropower plants worldwide [6]. However, the world's precipitation is affected by the variability of climate due to global warming that makes the hydropower plant vulnerable [7–9]. The impacts of climate change on water resources development are becoming a critical issue for many developing and developed countries in the world [10].

The optimal development of hydroelectric power generation and water resources systems are the major factors for economic development of developing countries [11]. Total hydropower production capacity in Africa is around 70 Giga Watts (GW), while almost 25% of the hydroelectric power plants are presently not operating due to poor maintenance [12]. In sub-Saharan Africa (SSA) countries, the capacity of hydropower installed is 27 GW with the additional capacity of 15 GW hydropower plants under construction [13]. According to the International Energy Agency (IEA) [14], the forecasted hydropower capacity in SSA will reach 95 GW by 2040 and this shows the fast growing of hydropower industry. For hydropower-dependent countries, hydropower represents more than 50% of electricity production providing 45% to the SSA population through grids [15].

Ethiopia is one of the developing countries, and it has rich water resources to generate hydropower with a capacity of 45,000 Mega Watt (MW) [16]. The country's power producing capacity had expanded from 850 MW to 4300 MW in 2017 [17]. The Ethiopian government is seriously working on a number of hydropower projects to make the country become a power center in East Africa with the Gibe-III hydropower (1870 MW) constructed on Omo Gibe River and Grand Ethiopian Renaissance Dam with the installed capacity of 5000 MW which is constructing on Abay (Blue Nile) River [18]. Currently, the demand is increasing consistently resulting to frequent power cut [19]. In addition, the government is working hard to control the country's 2030 GHG emissions to today's 1500 Mt CO_2 [20]. The emission is largely due to the traditional and unsustainable ways of using natural resources [21].

In Ethiopia, exploiting hydropower potential is considered as a key issue to bring economic growth to the country [22]. The importance of the function of operating maximum efficiency of reservoir is one of the most important requirements [23]. Temperature, precipitation, and stream flow are altering due to climate change [24, 25], and its impact is significant on hydropower scheme, dam/reservoir design horizons, and life periods [26, 27]. Reservoirs are important in balancing water scarcity in the system and sustaining the hydropower generation [28–30]. Reservoir operating rules are widely used to give directions for basin releases in order to maintain the best benefits of a reservoir with specific inflow and storage levels [31]. The impacts of climate change variables such as precipitation and temperature have hydrological impacts on reservoir operation and hydropower production [32, 33]. Countries like Ethiopia which poured a substantial investment in hydroelectric power generation were concerned about the climate change in terms of variation in precipitation and temperature on hydropower production besides forecasting the future power generation [34]. Most of the developing countries are vulnerable to climate change impact due to lack of financial development and organized capacity [35]. In Ethiopia, the effects of climate change may seriously affect the reservoir operation and hydropower production [26, 36]. Currently, there is considerable uncertainty about the performance of reservoirs in the context of climate change in Ethiopia [37]. About 50% of capacity reduction is perceived in hydropower projects due to climate change in the Ethiopia [17] by which the total installed capacity has to increase by 5.9% compared with the New Policy Scenario (169 GW to 179 GW), though the energy generation from the hydroelectric power had decreased by 2.7% compared with the New Policy Scenario (517 TWh to 503 TWh).

According to the Working Group II Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC AR6), the risks are projected for the near-term, mid-term, and long-term, at different global warming levels and for pathways that overshoot 1.5°C global warming level for multiple decades [38]. Increasing weather and climate extreme events have exposed millions of people to face severe food insecurity and reduced water security and energy, with the largest impacts cited in many locations and/or communities in Africa, Asia, Central and South America, Small Islands, and the Arctic [38].

The impact of climate variation on hydropower production is studied by many researchers using General circulation models (GCMs)/Regional Climate Models (RCMs) [39–41]. Dynamical downscaling is one of the methods that transfer data from GCMs to smaller scales by using a higher resolution regional climate models (RCMs) [42, 43].

Kesem watershed is one of the largest tributaries of Awash basin characterized by high climatic variation and water demands [44]. The climatic variation and water demands like irrigation, water supply for industries, and livestock put additional stress on Kesem reservoir and hydropower generation. There is no significant research on hydropower generation under climate change in the Kesem watershed for future periods. This study was initiated and focused on evaluating the future hydropower generation and reservoir operation under climate change.

2. Materials and Methods

2.1. Study Area. Kesem River basin is one of the sub-basins of Awash River basin located in the eastern part of Ethiopia. The basin is roughly found between 8°–10°N latitudes and 39°–40°E longitudes and covers a drainage area of 2885 km^2 with the mean elevation of 1825 m above mean sea level (Figure 1) [45]. The land use types of the Kesem watershed are acacia, agricultural land, bare land, eucalyptus, forest, settlement, shrub land, water body, and grass land [46]. The dominant soil types are vertic cambisols, lithic leptosols, eutric vertisols eutric leptosols, eutric cambisol, and chromic luvisols [47]. The climate of the Kesem dam site is hot and semiarid to arid climatic zone (mean annual maximum and minimum temperature of 26.8°c and 10.56°c, respectively) with the mean annual rainfall of 882 mm [46]. In this watershed, rainfall is categorized by two different seasons like spring (February to May) and summer (July to September) [46].

2.2. Data Collection. Meteorological data (precipitation, temperature, relative humidity, wind speed, and sunshine) for five meteorological stations within and around the Kesem watershed were obtained from the Ethiopia National

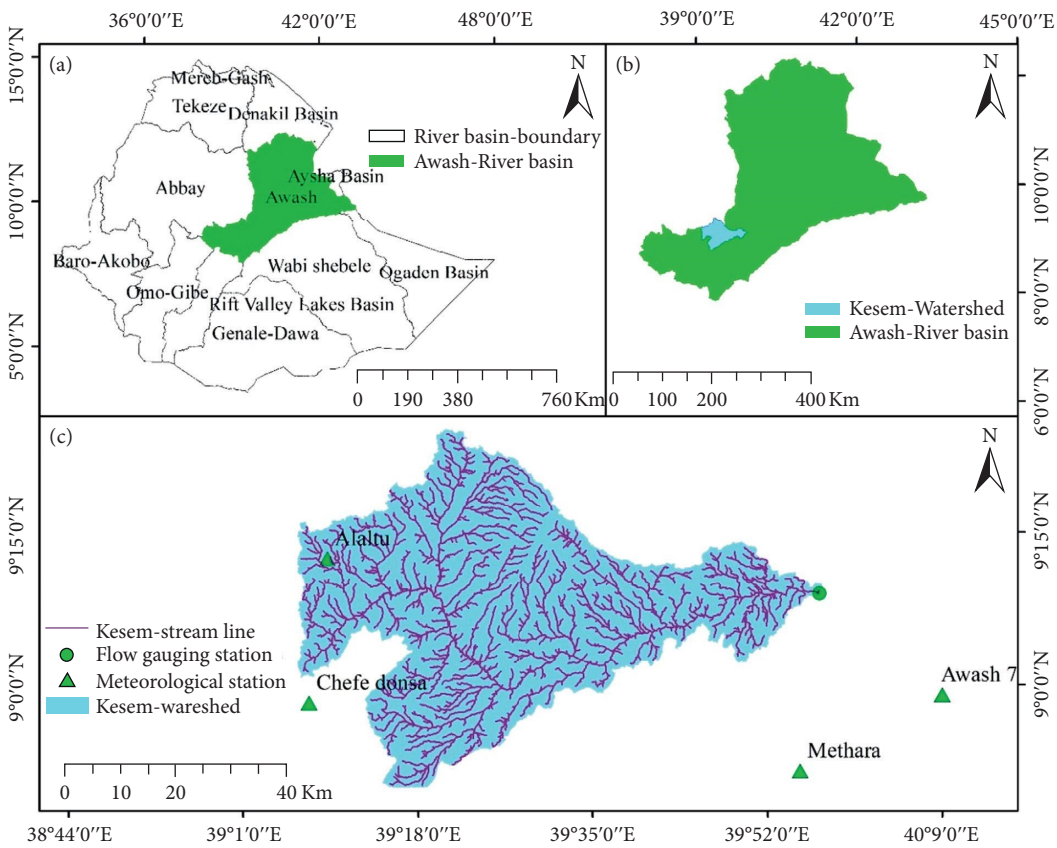


Figure 1: (a) Major river basins in Ethiopia, (b) Awash river basin, and (c) Kesem watershed.

Meteorology Agency. And also, hydrological data (streamflow) were collected from Ethiopia ministry of Water, Irrigation, and Energy. Dynamically downscaled outputs of General Circulation Model climate data were obtained by using the CORDEX-Africa program (http://wcrp-cordex.ipsl.jussieufr) for Representative concentration pathway (RCP 4.5 and RCP 8.5) scenarios for the period (1951 to 2100). IPCC used the new RCP scenarios in the fifth assessment report to represent emissions trajectories of RCP 2.6 (low emissions), RCP 4.5 (intermediate emissions), RCP 6 (intermediate emissions), and RCP 8.5 (high emissions scenario) [48]. In this research, the precipitation, minimum and maximum temperature, evapotranspiration, and evaporation for short-term (2021–2050) and long-term (2051–2080) under the RCP 4.5 and RCP 8.5 scenarios with respect to base period (1971–2000) were projected. 30 m × 30 m digital elevation model (DEM) was downloaded from United States Geological Survey (USGS), SRTM (Shuttle Radar Topography Mission) website (http:/earthexplorer.usgs.gov/on 10 October 2021) and used to delineate watershed by ArcGIS 10.5.

2.3. Data Processing

2.3.1. Bias Correction of RCP Data. The power transformation equation (Equation (1)) and the variance scaling (VARI) method were used for precipitation and temperature bias correction, respectively, since these methods found to be better at capturing the coefficient of variation and standard deviation of observed rainfall and temperature [49].

$$P^* = aP^b, \tag{1}$$

where P^* is the corrected precipitation, a and b are the calibration parameters in the baseline period applied to the projected period, and P is the power transform constant for the reference point of the precipitation bias correction.

The VARI method was used to adjust both the average and variance of normally distributed temperature data [49]:

$$T^* = \overline{Tob} + \frac{\delta(T_o)}{\delta(\overline{T}_R)}\left(T_{\mathrm{Rcp}} - \overline{T}_{\mathrm{ob}}\right) + \left(\overline{T}_{\mathrm{ob}} - \overline{T}_{\mathrm{Rcp}}\right), \tag{2}$$

where T_{Rcp} is the daily temperature, uncorrected (RCP data value), $\overline{T}_{\mathrm{ob}}$ is the observed daily average temperature, and $\overline{T}_{\mathrm{Rcp}}$ denotes the matching RCP basin mean temperature. In this equation, an over bar represents the average over the time period under consideration, as well as the standard deviation.

2.4. Accuracy of Rainfall Simulations from RCP Data. In this study, the accuracy of rainfall simulation of RCP data is checked by systematic error in rainfall amount (Bias), root mean square error (RMSE) and coefficient of variation. Each performance measures is expressed in equations (3)–(5) [50].

$$\text{Bias} = 100^{*}\frac{\overline{\text{Rrcp}} - \overline{\text{Robs}}}{\overline{\text{Robs}}}, \tag{3}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{k=1}^{N}(\text{Rrcp} - \text{Robs})^2}{N}}, \tag{4}$$

$$\text{CV} = 100^{*}\frac{\delta\text{Rrcp}}{\delta\text{Robs}}, \tag{5}$$

where N is the analysis period; $\overline{\text{Rrcp}}$ and $\overline{\text{Robs}}$ are the average rainfall amount obtained from RCP and observation data, respectively; and δ refers the standard deviation.

2.5. Trend Analysis. Trend analysis has been used to detect the impacts of climate change on precipitation, temperature, and evapotranspiration in the Kasem watershed. The Mann–Kendall test is a nonparametric test that does not require the data to follow normal distribution [51]:

$$S = \sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\text{sign}(Yj - Yi), \tag{6}$$

where Y_j and Y_i show the sequential precipitation or temperature values in average month/year where $i = 1, 2, 3, \ldots, n-1$ and $j = i+1, i+2, i+3\ldots n$. j and i $(j > i)$.

S is used to calculate the current and future annual precipitation and temperature trends.

The variance of S is calculated using the following equation:

$$\text{Var}(S) = \frac{n(n-1)(2n+5)}{18}. \tag{7}$$

If ties occur, n is the number of tied group and m is the number of tied values. For ties, the variance of S is calculated as

$$\text{Var}(S) = \frac{\left[n\left(n - 10(2n+5) - \sum_{i=1}^{n} ti(i)(i-1)(2i+5)\right)\right.}{18}. \tag{8}$$

For this study, an investigation was carried out to identify trends in climate data for 5 stations distributed over the whole of the Kesem watershed. Then, all the above equations are used to compute the Mann–Kendall Trend analysis [52].

2.6. Potential Evapotranspiration. The Penman and Hargreaves methods were used to compute evapotranspiration (ETo) on the Kesem watershed for the base and future periods [53]. The adjustment factor between the Penman and Hargreaves methods for future scenarios was used. The correction factor to be well-suited with the approach was adopted during model calibration, validation, and reservoir simulation [53]. In this research, evapotranspiration was calculated using both the Penman and Hargreaves methods for the base and future periods. A regression equation was developed to estimate the future potential evaporation:

$$\text{RVE} = \left[\frac{\sum_{i=1}^{n}(Q_s - Q_o)}{\sum_{i=1}^{n}Q_o}\right] * 100. \tag{9}$$

2.7. Reservoir Evaporation. Evaporation from reservoirs cannot be measured directly; it should be determined indirectly by number of methods [54]. In this study, the Penman Monteith method was used to estimate the monthly evapotranspiration rate at Kesem reservoir. The meteorological data (precipitation, temperature, relative humidity, wind speed, and sunshine) are collected at Metehara synoptic station, which is adjacent to the basin and is used to estimate monthly evapotranspiration (estimated evapotranspiration) from Kesem Reservoir.

2.8. HEC-HMS 4.2.1 Model Development. Input data used in the HEC-HMS model were stream flow, precipitation, evapotranspiration, and different watershed characteristics (digital elevation model, soil, and slope) obtained from Arc Hydro tools and the Geospatial Hydrologic Modelling Extension (HEC-GeoHMS). HEC-GeoHMS can be used under the GIS 10.5 environment, which is a geospatial hydrology tool allowing users to determine sub-basin streams as an input for the HEC-HMS hydrological model. Arc Hydro tools, an extension in ArcGIS, are used to process terrain data, define streams, and delineate the watershed of interest. The Kesem watershed is classified into three sub-basins such as sub-basin 1, sub-basin 2, and sub-basin 3 (Figure 2).

In addition, an outlet point (dam site) was used to estimate the total simulated flow. The Kesem watershed sub-basins, area, and the rainfall contribution of each sub-basin are illustrated in Table 1.

2.9. HEC-HMS Model Performance. The Nash–Sutcliffe coefficient (NSE) was used to evaluate model performance and values varying from zero to one. NSE of one and zero indicates the excellent and poor model performance, respectively [55]. The values between 0.6 and 1.0 are considered to be very good performance:

$$\text{NSE} = 1 - \frac{\sum_{i=1}^{n}(Q_O - Q_s)^2}{\sum_{i=1}^{n}(Qo - \overline{Q})^2}. \tag{10}$$

The coefficient of determination (R^2) was used to determine the strength of observed and simulated flow [56]:

$$R^2 = \frac{\left|\sum_{i=1}^{n}(Q_s - \overline{Q_s})(Q_o - \overline{Q_o})\right|^2}{\left|\sum_{i=1}^{n}(Q_s - \overline{Q_s})(Q_o - \overline{Q_o})\right|^2}. \tag{11}$$

Relative volume error (RVE) was used for quantifying the volume errors between observed and simulated value. When the RVE value is zero, its performance is the best which means there is no difference between simulated and observed runoff [57]:

$$\text{RVE} = \left[\frac{\sum_{i=1}^{n}(Q_s - Q_o)}{\sum_{i=1}^{n}(Q_s - Q_o)}\right]^{*} 100, \tag{12}$$

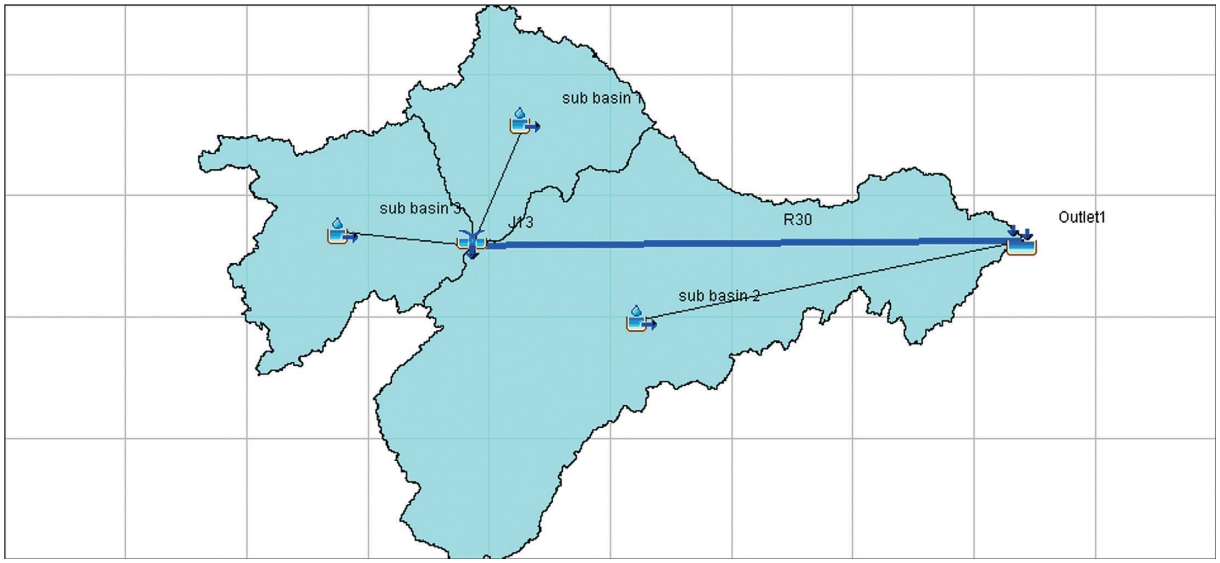


Figure 2: HEC-HMS model development.

Table 1: Kesem watershed sub-basins, area, and rainfall contribution.

Sub-basin	Area (km^2)	Contributing rainfall station
Sub-basin 1	488	Debre berhan Aleletu
Sub-basin 2	1880	Awash7 Metehrara Chefedonsa
Sub-basin 3	606	Chefedonsa Aleletu

where Q_O is the observed flow, Q_S is the simulated flow, $\overline{Q_o}$ is the average of observed flow, $\overline{Q_s}$ is the average of simulated flow, and n is the number of data points. RVE ranges between $-\infty$ and $+\infty$. The model performance is very good for RVE between −5% and 5%, while RVE between −10% and −5% and 5% and 10% suggests satisfactory performance and indicates a model with reasonable performance [57].

2.10. MODSIM 8.1 Model. MODSIM model is river basin management decision support system and used for river basin management and reservoir operation and to determine hydroelectric power generation [45, 58, 59]. The MODSIM model includes a powerful, interactive graphical user interface for creating, locating, and connecting eiver basin network components [60]. In this study, the input data used in the MODSIM model were stream flow, water demand, environmental release, net evaporation, maximum reservoir capacity, minimum and initial reservoir capacity, reservoir area and elevation, power plant capacity, load factor, tail water discharge, efficiency, and reservoir node properties.

3. Results and Discussion

3.1. RCP Rainfall Data Performance Evaluation. The mean annual observed rainfall of the basin is 882 mm/year, and the RCP rainfall value is 1115 mm (Table 2). The accuracy of RCP rainfall data performance showed that the rainfall of RCP data overestimated by 26% and root mean square (RMSE) performed 31 mm/year compared with the observed value.

3.2. Trends in Historical Climate. Figure 3 shows that the rainfall has decreased during the period from 1989 to 2003 except 1993 and 1998 (Figure 3(a)). The rainfall value was remarkably decreased by 6.706 mm/year. The Kesem watershed received its lowest rainfall amount in the periods of 2002. In the case of temperature, there was a considerable increasing trend (Figures 3(c) and 3(d)). The mean annual maximum and minimum temperature in the watershed has increased at rate of 0.42°C and 0.38°C per decade, respectively. The potential evapotranspiration (PET) shows an increasing trend between 1999 and 2005 (Figure 3(b)).

The bias correction result showed that there was a significant systematic error between the observed and climatic data in the months of April, May, June, July, August, and September (Figure 4). However, there was a small error revealed in RCP rainfall data in the months of January, February, March, October, November, and December.

The RCP average maximum temperature and observed temperature data show small underestimates for all months (Figure 5).

Figure 6 shows the RCP average minimum temperature and the observed minimum temperature.

The RCP average minimum temperature was slightly underestimated in most of the months except January, February, March, April, and May (Figure 6).

3.3. Future Scenarios of Precipitation and Temperature. Figure 7 shows the average monthly precipitation in the Kesem watershed under RCP4.5 and RCP8.5 scenarios with the baseline period (1971–2000) and future periods (2021–2050 and 2051–2080). Average monthly precipitation

Table 2: Accuracy of rainfall from RCP data.

	Mean annual rainfall (mm)	Bias (%)	CV (%)	RMSE (mm year^{-1})
Observed	882	—	9.6	—
RCP	1115	26	7.8	31

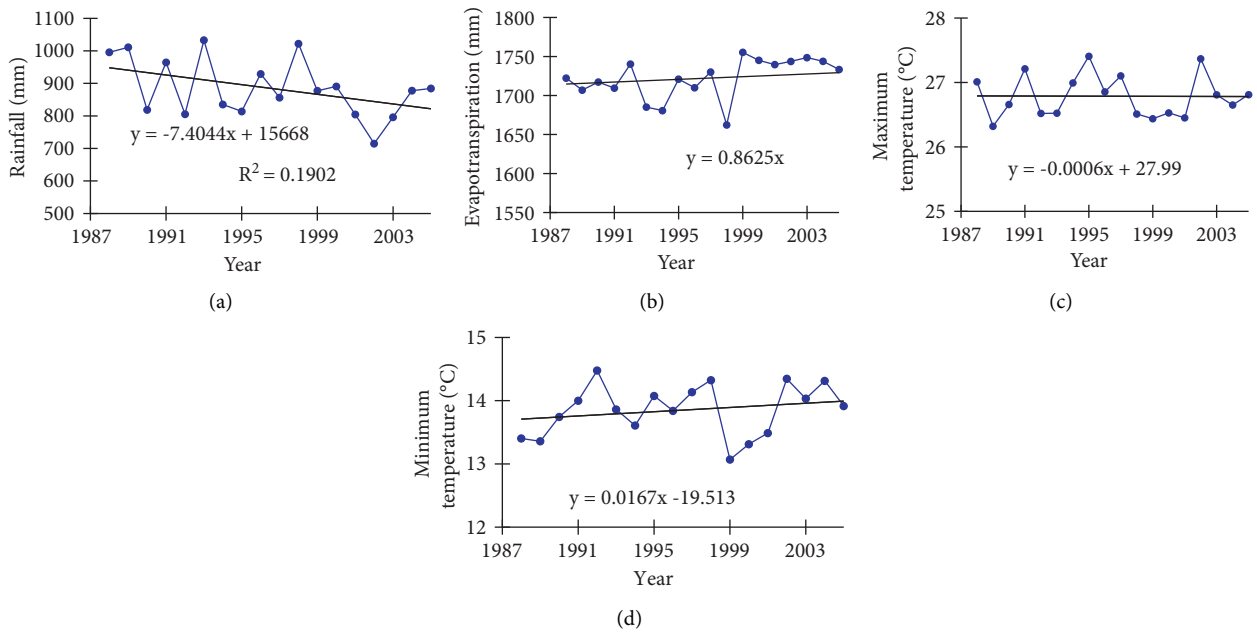


Figure 3: Trends in historical climatic: (a) rainfall, (b) evapotranspiration, (c) maximum temperature, and (d) minimum temperature data. Bias correction of precipitation and temperature data.

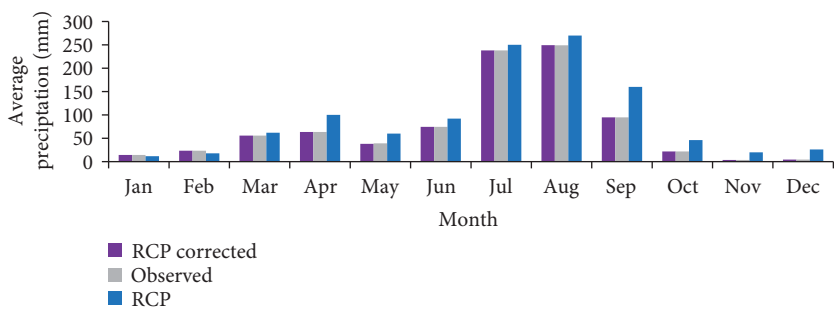


Figure 4: Bias correction of average monthly precipitations.

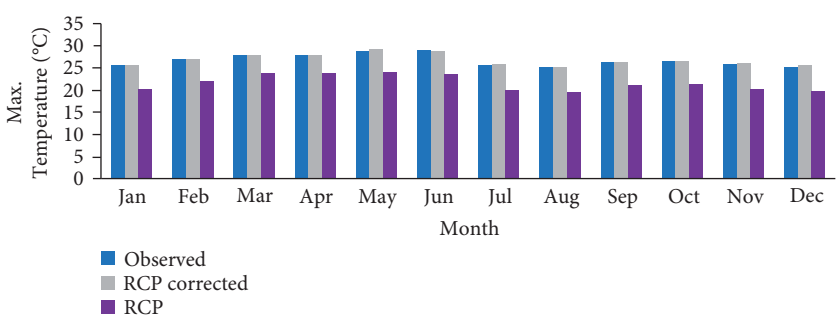


Figure 5: Bias correction of average monthly maximum temperature.

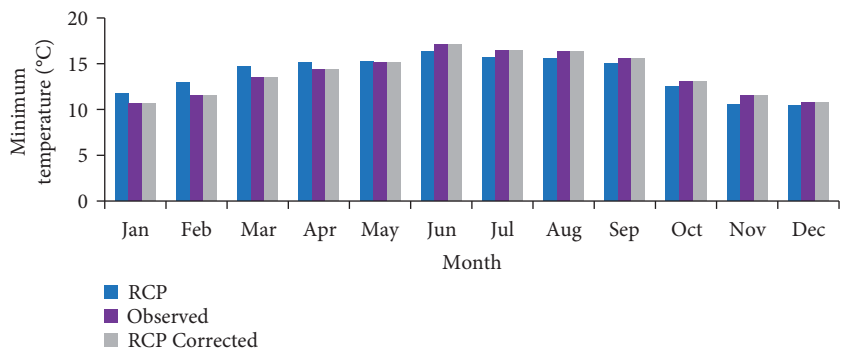


Figure 6: Bias correction of average monthly minimum temperature.

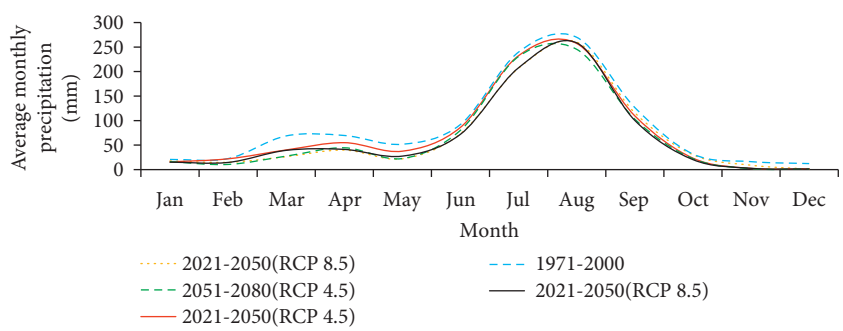


Figure 7: Average monthly precipitation of future scenarios for Kesem watershed.

for RCP4.5 scenario was changed by −9.7% and 16% in (2021–2050) and (2051–2080), respectively, and also for RCP8.5 average monthly precipitation was decreased by 16.4% and 17.6% in (2021–2050) and (2051–2080), respectively.

The maximum temperatures in the climate scenario RCP4.5 and RCP8.5 for the baseline period (1971–2000) and future periods (2021–2050 and 20051–2080) are illustrated in Figure 8. The monthly average maximum temperature rises by 0.2°C and 0.5°C for RCP4.5 (2021–2050) and (2051–2080), respectively, and by 0.7°C and 1°C for RCP8.5 (2021–2050) and (2051–2080), respectively.

Figure 9 shows the minimum temperature in the climate scenarios of RCP4.5 and RCP8.5 for the baseline period (1971–2000) and future periods (2021–2050 and 20051–2080). Monthly average minimum temperature was increased by 0.4°C and 0.7°C for RCP4.5 scenario for (2021–2050) and (2051–2080), respectively. Moreover, RCP8.5 monthly average minimum temperature was increased by 0.5°C and 0.6°C in future periods (2021–2050) and (2051–2080), respectively.

3.4. Climate Change Impact in Kesem Watershed. Precipitation was changed by −9.7% (2021–2050) and 16% (2051–2080) in the RCP4.5 scenario. In the RCP8.5 scenario, precipitation indicates a decreasing trend and the result shows 16.4% (2021–2050) and 17.5% (2051–2080). Climate change impacts on the Kesem watershed were compared with the observed period (1971–2000) and future periods (2021–2050 and 20051–2080) in the RCP4.5 and RCP8.5 scenarios (Figure 10).

Maximum temperature was increased by 0.2% (2021–2050) and 0.5% (2051–2080) for RCP4.5 scenarios, respectively (Figure 11). The average monthly temperature for the RCP8.5 scenario shows rising temperatures and fluctuated by 0.7% (2021–2050) and 1% (2051–2080).

Figure 12 demonstrates that during the future time period, the mean monthly minimum temperature over the Kesem watershed increased by 0.2% (2021–2050) and 0.3% (2051–2080) under RCP4.5 scenario. The average monthly minimum temperature change for the RCP8.5 scenario shows an increase in temperature.

3.5. Evapotranspiration. Figure 13 shows RCP4.5 and RCP8.5 scenarios of average monthly evapotranspiration increasing trend in the future scenarios. From Figure 13, it is observed that the average monthly evapotranspiration was increased by 2.5% (2021–2050) and 4.4% (2051–2080) for RCP4.5 and also for RCP8.5 increased by 3.4% (2021–2050) and 6.8% (2051–2080).

3.6. Evaporation. In the RCP4.5 and RCP8.5 scenarios, the rate of monthly evaporation in the Kesem watershed was significantly increasing for both short-term (2021–2050) and long-term (2051–2080) (Figure 14).

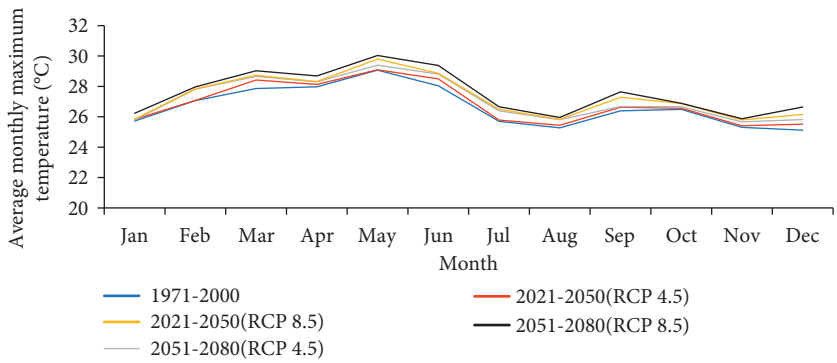


Figure 8: Average monthly maximum temperatures for future scenarios.

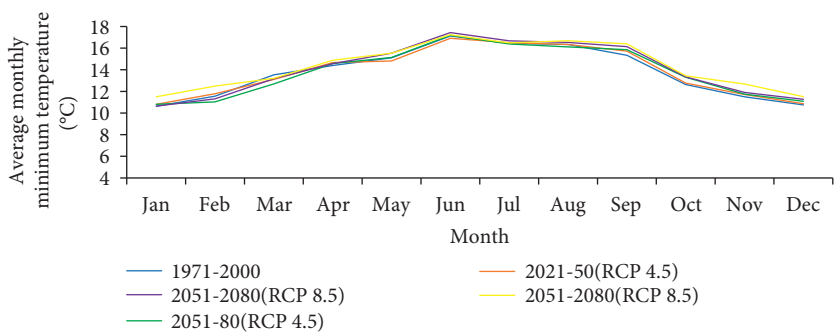


Figure 9: Average monthly minimum temperatures for future scenarios.

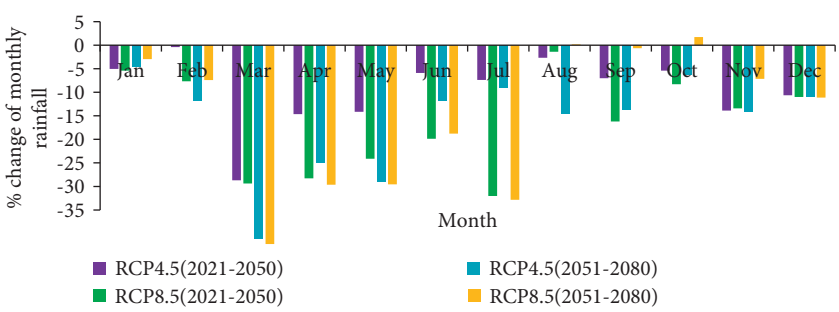


Figure 10: Change in average monthly rainfall over the period (2021–2050 and 2051–2080).

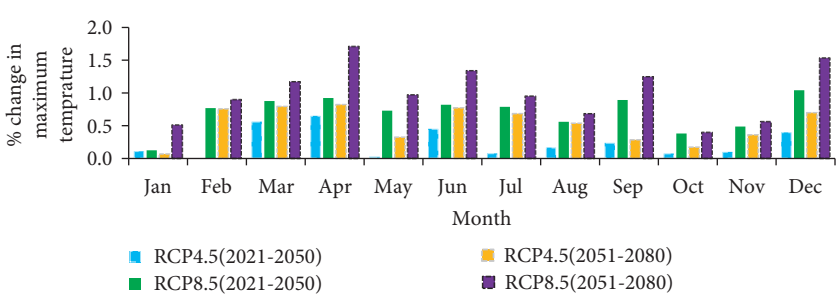


Figure 11: Change in average monthly maximum temperature.

3.7. HEC-HMS Model Sensitivity Analysis, Calibration, and Validation. Sensitivity analysis can be used to identify more sensitive model parameters for further model calibration. The effects of each ±30% modification in the model parameter values were investigated [61]. To test the sensitive parameters, each model parameters can have its value

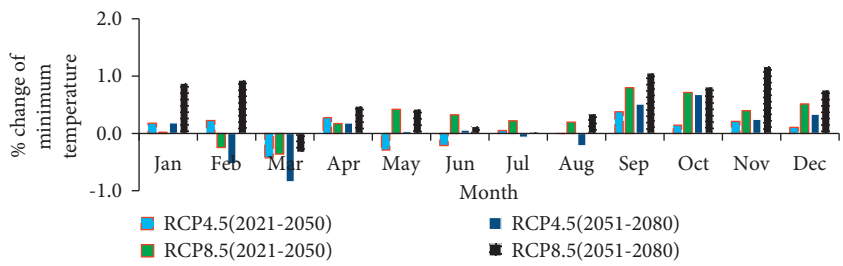


Figure 12: Change in average monthly minimum temperature.

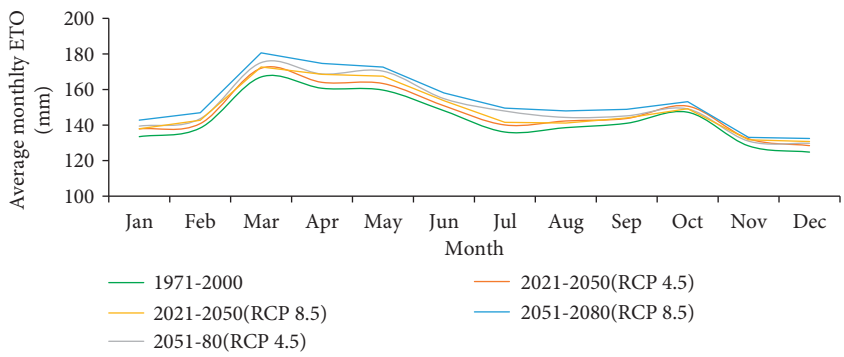


Figure 13: Changes in average monthly evapotranspiration.

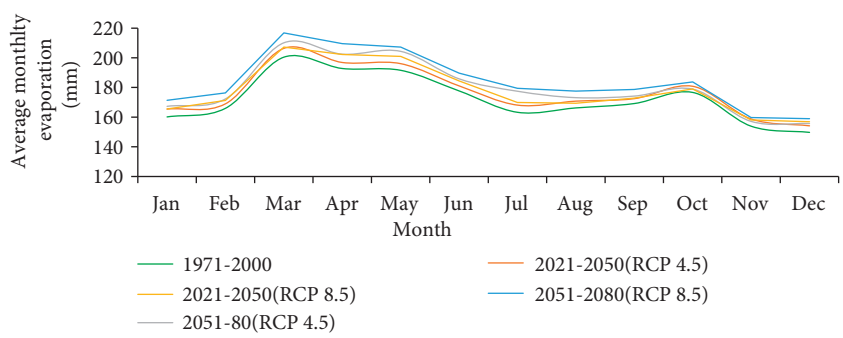


Figure 14: Changes in average monthly evaporation from Kesem reservoir.

increased or decreased by up to ±30%. After the model parameters were adjusted by ±30% of peak volume, the sensitive parameters were identified. The constant rate (CR), storage coefficient (SC), and dimensionless weight (X) were more sensitive parameters (Figure 15). The recession constant (RC), the initial loss (IL), time of concentration (TC), Initial discharge (ID), recession constant (RC), Ratio-to-peak (RP), and Travel time (K) all were less sensitive parameters.

Streamflow data were used to calibrate (1992 to 2003) and validate (2004 to 2009) the HEC-HMS with warm up period (1990 to 1991) (Figure 16).

Stream flow calibration process is used to establish the most suitable parameter in modelling studies and an iterative process that compares simulated and observed data through parameter evaluation. The observed areal precipitation, areal evapotranspiration, and observed stream flow are calibrated using automatic and manual calibration process for Kesem reservoir. The calibration result (Figure 16) showed that the model has good agreement between the observed and simulated stream flow, and the peak is well captured most of year. However, there is some underestimation in the years of 1993, 1997, and 2000 for simulated flow.

After calibration of the model, Nash–Sutcliffe coefficient (NSE) value of 0.72, coefficient of determination (R^2) of 0.73, and relative volume error (RVE) of 0.063% were observed. Moreover, for validation, the model performed with NSE of 0.74, R^2 of 0.75, and RVE of −4.7%. Based on Moriasi et al. [55] model evaluation criteria, the aforementioned model performance result was viewed as acceptable level and good correlation of data. Figure 16 shows the peak stream flow

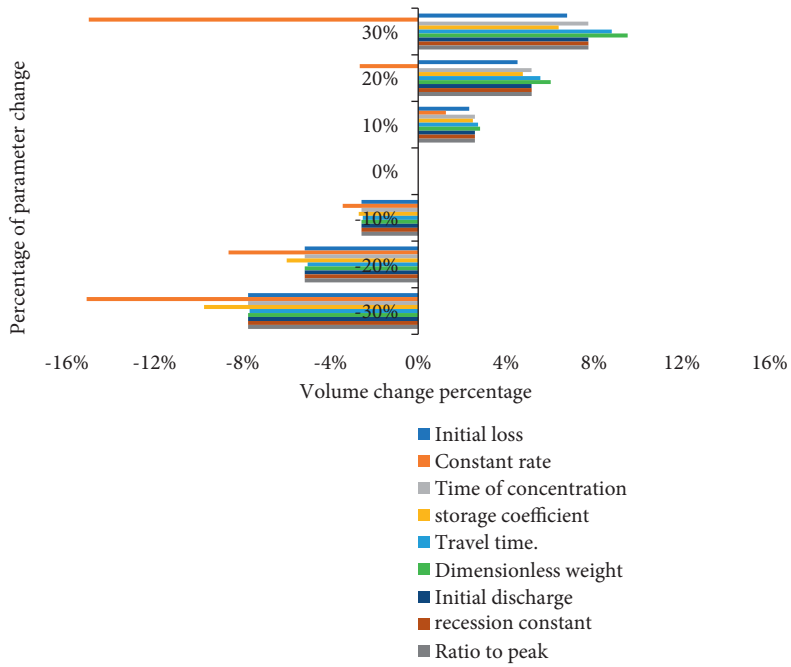


Figure 15: Percentage change in volume verses percentage of parameter variation.

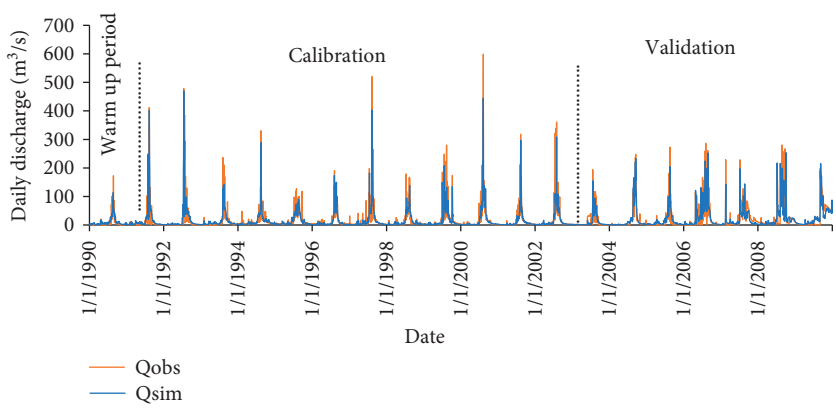


Figure 16: Observed and simulated stream flow at Kesem dam.

well captured in the years of 2004 and 2006. However, there is underestimation in 2005, 2003, and 2008 and overestimation in 2009.

3.8. Reservoir Inflow. Reservoir inflow (stream flow data) was used to forecast the future scenarios in different time horizons including the baseline, short-term, and long-term. Average stream flow will change by −12.93% (2021–2050) and −17.01% (2051–2080) RCP4.5 as compared with the observed flow. Average monthly stream flow will likely decrease by −21.83% (2021–2050) and −24.87% (2051–2080) for RCP8.5 scenario, and this shows a downward trend (Figure 17).

3.9. Hydropower Generation with Observed Data. The hydropower production anticipated from the observed flow data results depicts that there is a high fluctuation of average power generation and energy. In the observed period (1987–2005), the reservoir simulation showed that the maximum energy generated was 378.613 MWH with the average energy of 300.093 MWH (Figure 18).

3.10. Hydropower Generation with RCPs Scenarios. The reservoir simulation focused on energy variation between observed period (1987–2005) and future (2021–2050) and (2051–2080) periods for RCP4.5 and RCP8.5 scenarios. In RCP4.5 scenario, energy generated for short-term period (2021–2050) will be 376.213 MWH with the average energy of 321.392 MWH. In the case of long-term (2051–2080), energy generated in the long-term scenarios (2051–2080) will be 370.513 MWH with the average energy of 307.682 MWH. This indicates that the average energy generation will be decreased by 0.64% and 0.82% in both short-term and long-term, respectively.

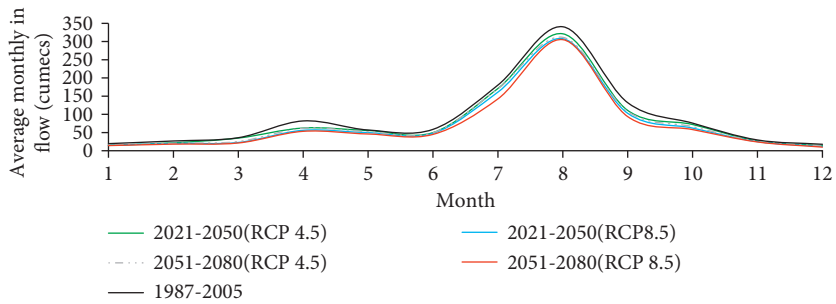


Figure 17: Kesem reservoir inflow discharge at different time horizons.

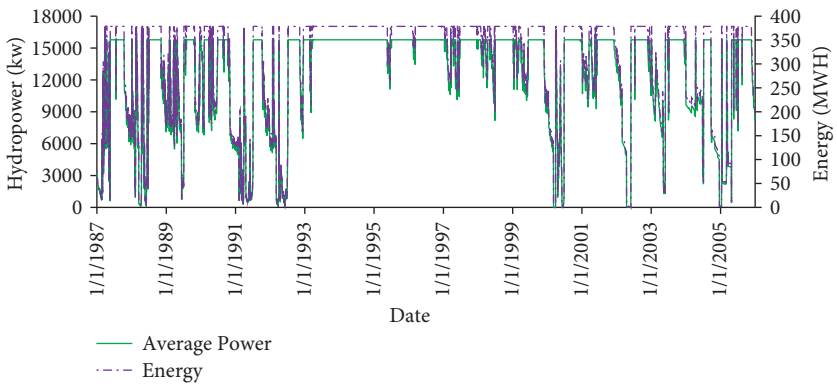


Figure 18: Average hydropower verses energy for base period scenario.

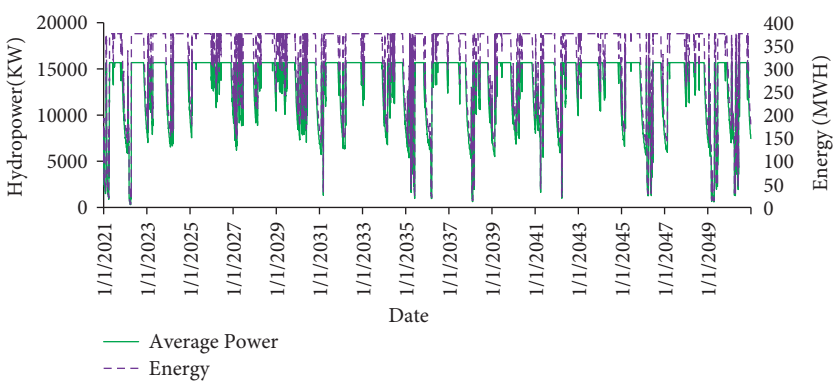


Figure 19: Daily hydropower generation and energy.

The RCP8.5 scenario result showed that the maximum energy generated in short-term scenarios will be 368.605 MWH, with an average energy of 320.411 MWH. In the case of long-term, for RCP8.5, the maximum energy generated will be 363.492 MWH and average energy of 306.584 MWH. In both short-term and long-term, the average energy generated will likely decrease by 1.06% and 1.35%, respectively. Figure 19 shows the sample of the MODSIM output of daily hydropower generation production and energy.

3.11. Average Hydropower Generation. The result of this study reveals that the Kesem reservoirs daily hydropower generation and water volume reservoir had a good hydropower generation potential for baseline condition. However, as a result of the influence of climate change on the Kesem watershed, the inflow, storage volume, and hydropower output have all declined. Figure 20 shows that on average the hydropower generated will likely change or slightly decrease between the observed period −0.64% (2021–2050) and −0.82% (2051–2080) RCP4.5 scenario, respectively. The RCP8.5 scenario of average monthly hydropower change shows decreased power on average; the power will likely change by −1.06% (2021–2050) and −1.35% (2051–2080), respectively.

3.12. Reservoir Operation Guide Curve. Depending on Chaleeraktrakoon and Chinsomboon [62] reservoir guide

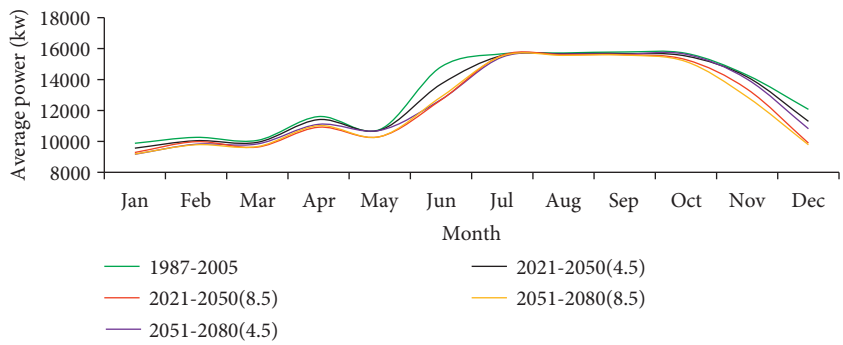


Figure 20: Average monthly power generation.

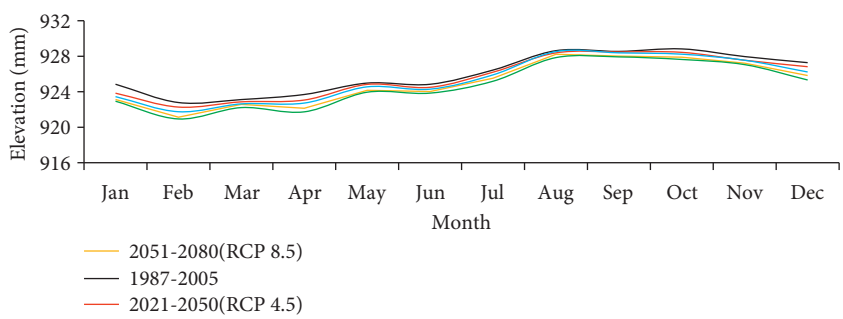


Figure 21: Reservoir operation guide curve.

curve, the reservoir rule curves were developed for Kesem reservoir and classified into three periods. These were the observed period (1987–2005), short-term (2021–2050), and long-term (2051–2080) for RCP4.5 and RCP8.5 scenarios to achieve the target demand of the projects. Figure 21 shows the reservoir rule guide curves for Kesem reservoir, and the result shows the decreasing trend for future scenarios compared with the observed period. The findings of the research work indicate that in Kesem, watershed climate change affects the reservoir inflow and reservoir operation and in turn reduces the future hydropower and energy production. This research finding is important for undertaking climate change mitigation strategies, watershed management practices, and to control the future change of hydropower production.

4. Conclusion

The evaluation of hydropower generation and reservoir operation under climate change can play a significant role by providing the baseline and future hydropower production evidences in the particular hydropower plant. Climate change is likely to have severe effects on water availability of Kesem reservoir. The result of climate projection reveals that the RCP data can be seen considering climate scenarios of RCP4.5 and RCP8.5 for short-term and long-term.

The HEC-HMS model performance during calibration period for stream flow was good with the NSE of 0.72, R^2 of 0.73, and RVE of −0.063%. During validation period, model performance was also good result with NSE of 0.74, R^2 of 0.75, and RVE of −4.7%. Reservoir operation for base and future period was analysed using the reservoir simulation model (MODSIM). The model had simulated the amount of release and hydropower generation to determine reservoir operation rule curve for both RCP4.5 and RCP8.5 scenarios.

Accordingly, the output generated will be used for the hydropower production and reservoir operation rule curve for different climate periods that includes baseline (1987–2005), short-term (2021–2050), and long-term (2051–2080). The quantified maximum energy for RCP4.5 will be 376.213 MWH and 370.513 MWH in short and long period, respectively. Moreover, in case of RCP8.5, the maximum energy of 368.605 MWH and 363.492 MWH will likely generate for the short and long period, respectively. This indicates that there is fluctuation in energy generation and decreasing trend in the future energy production from the Kesem reservoir. The outcome of this study is important to hydroelectric power authorities and watershed management agencies to ensure sustainability in the future hydroelectric power production in the Kesem reservoir and it is highly suggested that other scientific researchers can apply the same approach for similar type of reservoirs.

Data Availability

All the data sets used to support the findings of the study are available from the corresponding author(s) upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest among the authors regarding the publication of this paper.

References

[1] J. P. Bruce, *Meteorology and Water Resources*, Water Resources of Canada, 2017, https://www.google.com/search?rlz=1C1GCEB_enIN1005IN1005&q=Berlin&stick=H4sIAAAAAAAAOPgE-LSz9U3MDIvyEpLU-IAsQ1NzQq1tLKTrfTzi9IT8zKrEksy8_NQOFYZqYkphaWJRSWpRcWLWNmcUotyMvN2sDLuYmfiYAAAPJb5vVUAAAA&sa=X&ved=2ahUKEwiT2oCPs4T4AhVJT2wGHYtaBmkQmxMoAXoECF4QAw.

[2] J. D. Hunt, G. Falchetta, B. Zakeri et al., "Hydropower impact on the river flow of a humid regional climate," *Climatic Change*, vol. 163, no. 1, pp. 379–393, 2020.

[3] T. G. Huntington, "Evidence for intensification of the global water cycle: review and synthesis," *Journal of Hydrology*, vol. 319, no. 1-4, pp. 83–95, 2006.

[4] D. C. Sohoulande, S. Djebou, and V. P. Singh, "Impact of climate change on the hydrologic cycle and implications for society," *Environment and Social Psychology*, vol. 1, no. 1, pp. 36–49, 2016.

[5] WEC (World Energy Council), *Energy Resources: Hydropower*, WEC (World Energy Council), London, UK, 2019.

[6] Organization for Economic Cooperation and Development (OECD), *Medium-term Renewable Energy Market Report*, OECD, Paris, France, 2017.

[7] J. L. Chiang, H. C. Yang, Y. R. Chen, and M. H. Lee, "Potential impact of climate change on hydropower generation in Southern Taiwan," *Energy Procedia*, vol. 40, pp. 34–37, 2013.

[8] P Pilesjo and S. S. Al-Juboori, "Modelling the effects of climate change on hydroelectric power in Dokan, Iraq," *International Journal of Energy and Power Engineering*, vol. 5, no. 2, p. 7, 2016.

[9] R. Schaeffer, A. S. Szklo, A. F. Pereira de Lucena et al., "Energy sector vulnerability to climate change: a review," *Energy*, vol. 38, no. 1, pp. 1–12, 2012.

[10] N. M. Lin and M. Rutten, "Optimal operation of a network of multi-purpose reservoir: a review," *Procedia Engineering*, vol. 154, pp. 1376–1384, 2016.

[11] S. Saied, H. Magsi, F. Marri, and H. Rejeb, "Under what management practices and mechanisms water resource can be considered as a public good for sustainable development in the developing countries," *Journal of Water Sustainability*, vol. 3, no. 1, pp. 1–16, 2013.

[12] K. Findt, D. B. Scott, and C. Lindfeld, *Sub-Saharan Africa Power Outlook*, KPMG, Amstelveen, Netherlands, 2014.

[13] International Hydropower Association (IHA), *Hydropower Status Report*, IHA Central Office, London, UK, 2019.

[14] International Energy Agency, *Africa Energy Outlook - A Focus on Energy Prospects in Sub-saharan Africa*, IEA, Paris, France, 2014.

[15] G. Falchetta, D. E. H. J. Gernaat, J. Hunt, and S. Sterl, "Hydropower dependency and climate change in sub-Saharan Africa: a nexus framework and evidence-based review," *Journal of Cleaner Production*, vol. 231, pp. 1399–1417, 2019.

[16] International Energy Agency, *Africa Energy outlook report*, IEA, Paris, France, 2019.

[17] T. W. Mekonnen, S. T. Teferi, F. S. Kebede, and G. Anandarajah, "Assessment of impacts of climate change on hydropower-dominated power system: the case of Ethiopia," *Applied Sciences*, vol. 12, no. 4, p. 1954, 2022.

[18] MoWIE, "National electrification program 2.0 national electrification," 2019, https://www.africa-energy-forum.com/article/ethiopia-national-electrification-program-20report.

[19] M. A. Tiruneh and G. B. Worku, "Identification of system vulnerabilities in the Ethiopian electric power system," *Global Energy Interconnect*, vol. 1, pp. 358–365, 2018.

[20] INDC, "Intended nationally determined contribution of the federal democratic republic of Ethiopia," 2015, https://www.climatelearningplatform.org/intended-nationally-determined-contribution%20indc-federal-democratic%20republic-ethiopia-0.

[21] G. N. Taka, T. T. Huong, I. H. Shah, and H.-S. Park, "Determinants of energy-based CO2 emissions in Ethiopia: a decomposition analysis from 1990 to 2017," *Sustainability*, vol. 12, no. 10, p. 4175, 2020.

[22] S. B. Awulachew, A. D. Yilma, M. Loulseged, W. Loiskandl, M. Ayana, and T. Alamirew, "Water resources and irrigation development in Ethiopia," *International Water Management Institute*, p. 78, Colombo, Sri Lanka, 2007.

[23] J. Anand, A. K. Gosain, and R. Khosa, "Optimisation of multipurpose reservoir operation by coupling soil and water assessment tool (SWAT) and genetic algorithm for optimal operating policy (case study: ganga River Basin)," *Sustainability*, vol. 10, no. 5, p. 1660, 2018.

[24] B. Khaniya, I. Jayanayaka, P. Jayasanka, and U. Rathnayake, "Rainfall trend analysis in Uma oya basin, Sri Lanka, and future water scarcity problems in perspective of climate variability," *Advances in Meteorology*, vol. 2019, Article ID 3636158, 10 pages, 2019.

[25] B. Khaniya, C. Karunanayake, M. B. Gunathilake, and U. Rathnayake, "Projection of future hydropower generation in samanalawewa power plant, Sri Lanka," *Mathematical Problems in Engineering*, vol. 2020, Article ID 8862067, 11 pages, 2020.

[26] F. Abera, D. Asfaw, A. Engida, and A. Melesse, "Optimal operation of hydropower reservoirs under climate change: the case of Tekeze reservoir, Eastern Nile," *Water*, vol. 10, no. 3, p. 273, 2018.

[27] A. Ranzani, M. Bonato, E. Patro, L. Gaudard, and C. De Michele, "Hydropower future: between climate change, renewable deployment, carbon and fuel prices," *Water*, vol. 10, no. 9, p. 1197, 2018.

[28] M. Mamman, "Recalibration of kainji reservoir operating rule for optimal operation," *Annals of Reviews & Research*, vol. 1, no. 1, pp. 23–29, 2018.

[29] Y. Meng, J. Liu, Z. Wang, G. Mao, K. Wang, and H. Yang, "Undermined co-benefits of hydropower and irrigation under climate change," *Resources, Conservation and Recycling*, vol. 167, Article ID 105375, 2021.

[30] L. A. Scola, R. H. C. Takahashi, and S. A. A. G. Cerqueira, "Multipurpose Water Reservoir Management: An Evolutionary Multiobjective Optimization Approach," *Mathematical Problems in Engineering*, vol. 2014, Article ID 638259, , 2014.

[31] M. Visweswararao and G. K. Viswanadh, "Operation of a upstream reservoir using rule curve to improve the performance of the downstream reservoirs," *International Journal of Civil Engineering & Technology*, vol. 10, no. 7, pp. 14–21, 2019.

[32] N. Norouzi, "Climate change impacts on the water flow to the reservoir of the Dez Dam basin," *Water Cycle*, vol. 1, pp. 113–120, 2020.

[33] V. Sýs, P. Fošumpaur, and T. Kaspar, "The impact of climate change on the reliability of water resources," *Climate*, vol. 9, no. 11, p. 153, 2021.

[34] S. Shrestha, M. Khatiwada, M. S. Babel, and K. Parajuli, "Impact of climate change on river flow and hydropower

production in Kulekhani hydropower project of Nepal," *Environmental Processes*, vol. 1, no. 3, pp. 231–250, 2014.

[35] I. Niang, O. C. Ruppel, M. A. Abdrabo, C. Lennard, J. Padgham, and P. Urquhart, "Africa. Climate Change 2014: Impacts, Adaptation and Vulnerability: Part B: Regional Aspects," *Working Group II Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, vol. 2, pp. 1199–1266, 2015.

[36] Z. Belay and G. Getaneh, "Climate change in Ethiopia variability, impact, mitigation, and adaptation," *Journal of Social Science and Humanities Research*, vol. 2, no. 4, pp. 66–84, 2016.

[37] A. G. Mulat, S. A. Moges, and M. A. Moges, "Evaluation of multi-storage hydropower development in the upper Blue Nile River (Ethiopia): regional perspective," *Journal of Hydrology: Regional Studies*, vol. 16, pp. 1–14, 2018.

[38] H. O. Pörtner, D. C. Roberts, and E. S. Poloczanska, "Climate change impacts, adaptation, and vulnerability," in *Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC)* Cambridge University Press, Cambridge, UK, 2022.

[39] M. S. Khan, P. Coulibaly, and Y. Dibike, "Uncertainty analysis of statistical downscaling methods," *Journal of Hydrology*, vol. 319, no. 1-4, pp. 357–382, 2006.

[40] B. Khaniya, H. G. Priyantha, N. Baduge, H. M. Azamathulla, and U. Rathnayake, "Impact of climate variability on hydropower generation: a case study from Sri Lanka," *ISH Journal of Hydraulic Engineering*, vol. 26, no. 3, pp. 301–309, 2018.

[41] P. Qin, H. Xu, M. Liu et al., "Climate change impacts on three Gorges reservoir impoundment and hydropower generation," *Journal of Hydrology*, vol. 580, Article ID 123922, 2020.

[42] T. H. Alemseged and R. Tom, "Evaluation of Regional Climate Model Simulations of Rainfall over the Upper Blue Nile Basin," *Atmospheric Research*, vol. 161–162, 2015.

[43] W. Yang, J. Andréasson, L. Phil Graham, J. Olsson, J. Rosberg, and F. Wetterhall, "Distribution-based scaling to improve usability of regional climate model projections for hydrological climate change impacts studies," *Hydrology Research*, vol. 41, no. 3-4, pp. 211–229, 2010.

[44] D. Bekele, T. Alamirew, A. Kebede, G. Zeleke, and A. M. Melesse, "Modeling climate change impact on the hydrology of keleta watershed in the Awash River basin, Ethiopia," *Environmental Modeling & Assessment*, vol. 24, no. 1, pp. 95–107, 2019.

[45] F. T. Berhe, A. M. Melesse, D. Hailu, and Y. Sileshi, "MODSIM-based water allocation modeling of Awash River Basin, Ethiopia," *Catena*, vol. 109, pp. 118–128, 2013.

[46] N. Tessema, A. Kebede, and D. Yadeta, "Modeling land use dynamics in the Kesem sub-basin, Awash River basin, Ethiopia," *Cogent Environmental Science*, vol. 6, no. 1, Article ID 1782006, 2020.

[47] C. S. Otto and D. Josef, "World reference base for soil resources," *An Introduction with Special Reference to Soils of Tropical Forest Ecosystems (Working Paper and Preprint, 95/14)*, vol. 6, 2000.

[48] G. Fan, A. Sarabandi, and M. Yaghoobzadeh, "Evaluating the climate change effects on temperature, precipitation and evapotranspiration in eastern Iran using CMPI5," *Water Supply*, vol. 21, no. 8, pp. 4316–4327, 2021.

[49] M. Luo, T. Liu, F. Meng et al., "Comparing Bias Correction Methods Used in Downscaling Precipitation and Temperature from Regional Climate Models: A Case Study from the Kaidu River Basin in Western China," *Water*, vol. 10, 2018.

[50] T. H. Alemseged and R. Tom, "Evaluation of regional climate model simulations of rainfall over the upper blue nile basin," *Atmospheric Research*, vol. 161-162, pp. 57–64, 2015.

[51] J. Mallick, A. R. M. T. Islam, B. Ghose et al., "Spatiotemporal trends of temperature extremes in bangladesh under changing climate using multi-statistical techniques," *Theoretical and Applied Climatology*, vol. 147, 2021.

[52] Z. A. Alemu and M. O. Dioha, "Climate change and trend analysis of temperature: the case of Addis Ababa, Ethiopia," *Environmental Systems Research*, vol. 9, no. 1, p. 27, 2020.

[53] M. Gentilucci, M. Bufalini, M. Materazzi et al., "Calculation of potential evapotranspiration and calibration of the Hargreaves equation using geostatistical methods over the last 10 Years in Central Italy," *Geosciences*, vol. 11, no. 8, p. 348, 2021.

[54] M. Alazard, C. Leduc, Y. Travi, G. Boulet, and A. Ben Salem, "Estimating evaporation in semi-arid areas facing data scarcity: example of the El Haouareb dam (Merguellil catchment, Central Tunisia)," *Journal of Hydrology: Regional Studies*, vol. 3, pp. 265–284, 2015.

[55] D. N. Moriasi, J. G. Arnold, M. W. Van Liew, R. L. Bingner, R. D. Harmel, and T. L. Veith, "Model evaluation guidelines for systematic quantification of accuracy in watershed simulations," *Transactions of the ASABE*, vol. 50, no. 3, pp. 885–900, 2007.

[56] D. R. Legates and G. J. McCabe, "Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation," *Water Resources Research*, vol. 35, no. 1, pp. 233–241, 1999.

[57] B. B. Goswami, R. P. M. Krishna, P. Mukhopadhyay, M. Khairoutdinov, and B. N. Goswami, "Simulation of the Indian summer monsoon in the super parameterized climate forecast system version 2: preliminary results," *Journal of Climate*, vol. 28, no. 22, 9012 pages, 2015.

[58] W. Kitlasten, E. D. Morway, R. G. Niswonger et al., "Integrated hydrology and operations modeling to evaluate climate change impacts in an agricultural valley irrigated with snowmelt runoff," *Water Resources Research*, vol. 57, no. 6, 2021.

[59] J. Labadie, *MODSIM 8.1: River Basin Management Decision Support System User Manual and Documentation*, Department of Civil and Environmental Engineering Colorado State University Ft, Fort Collins, Colorado, 2010.

[60] E. Goharian and S. J. Burian, "Developing an integrated framework to build a decision support tool for urban water management," *Journal of Hydroinformatics*, vol. 20, no. 3, pp. 708–727, 2018.

[61] M. B. Gunathilake, P. Panditharathne, A. S. Gunathilake, and N. D. Warakagoda, "Application of a HEC-HMS model on event-based simulations in a tropical watershed," *Journal of Engineering and Applied Science Research*, vol. 47, no. 4, pp. 349–360, 2020.

[62] C. Chaleeraktrakoon and Y. Chinsomboon, "Dynamic rule curves for flood control of a multipurpose dam," *Journal of Hydro-Environment Research*, vol. 9, no. 1, pp. 133–144, 2015.

Hindawi
Advances in Meteorology
Volume 2022, Article ID 1523198, 17 pages
https://doi.org/10.1155/2022/1523198

Research Article

Solar GHI Ensemble Prediction Based on a Meteorological Model and Method Kalman Filter

Yuanyuan Liu

Department of Mathematics, Luliang University, Luliang 033001, China

Correspondence should be addressed to Yuanyuan Liu; hnzykys@126.com

Received 15 April 2022; Revised 20 May 2022; Accepted 21 May 2022; Published 10 June 2022

Academic Editor: Upaka Rathnayake

The intensity of light emanating from sun is determined by using a meteorological version and is altered with the numerical version, and the forecast accuracy is improved in advance by using Kalman Filter. As the accuracy of the version output related to its specific position is often questionable, group prediction constituting three members is suggested and agreed upon measurement. Also, this ensemble prediction provides an estimation of the solar global horizontal irradiance uncertainty (i.e., coverage rate of the prediction interval), which can be useful to provide flexible energy production forecasts. This article displays how the method Kalman filter could be used as an error correction way to alter the predicted irradiance value. The Kalman filter ameliorates the prediction of solar global horizontal irradiance as well as its interval. As the empirical coverage rate increases and closes to the nominal coverage rate, the interval size reduces.

1. Introduction

Having been installed and connected to power grids, a large amount of photovoltaic (PV) systems with their prognostication in power generation or the prediction of the solar irradiance have been used in handling the electric grids more efficiently to make the energy provision steadier. With the help of dependable and conforming prediction data [1–9], system operators ameliorate the electrical management to balance the users and power generation [10–12]. Also, with the development of the forecasting models, support is offered in an efficient way for electricity suppliers in managing, dispatching, and planning power plants [13]. By applying a meteorological model, the authors have advanced a system in ensemble prediction of solar global horizontal irradiance [14], and not only the irradiance but also its prediction interval has been forecasted [15–17]. The interval corresponds to the reliability of the forecasting and will be also useful for power grid management. With the use of model WRF (the Weather Research and Forecasting), Liu et al. proposed solar global horizontal irradiance and verified its forecasting accuracy [14]; the simulated WRF irradiance included RMSEs, which imply root mean square errors [18–26] as well as positive biases. In order to improve solar global horizontal irradiance forecasting accuracy, the meteorological calculation used in a post-\processing technique is introduced [27–30]; and, the result in the amelioration on the forecasting of the prediction interval is also discussed in this work [31–34].

Different approaches for processing in post-period, with examples of MOS (model output statistics) [35–39] and Kalman filter, are put into use in weather forecast in a frequent way for the sake of modifying the anticipatory output in gaining possession by NWP models which stand for numerical weather prediction. In 1995, Homleid depicted a model Kaman Filter to estimate temperature forecasting corrections in day time [34]. In 2011, De Carvalho et al. [35] applied the model Kalman filter for correcting the mean temperature errors which are assessed by the model PRECIS, a system developed by the English Hadley Centre used for climate predictions in regional area. In 2008, Monache et al. [36] estimated systematic errors in surface ozone forecasts. Galanis et al. in 2006 [37] as well as Cassola and Burlando in 2012 [38] put the method Kalman filtering procedure into removing bias in forecasting for wind speed. Pelland et al. in 2013 [39] as well as Diagne et al.

in 2014 [40] used the model Kalman filter to remove bias in solar irradiance prediction. In 2018, Rincón with his teammates [41] made a combination with model Kalman filter and MOS for bias correction for the sake of improving the entire predicted values in solar irradiance. The creation of correction function method on linear regression traditionally depends mainly on MOS. According to Kalman [42] and Homleid [34], the theory of the Kalman filter offers equations in modifying the assessments in a recursive way through observing.

The probabilistic forecasting of solar global horizontal irradiance can have an important impact on the management of the power grid with many PV systems and its risk analysis. For this case, the method Kalman filter is used as a posttreatment technique to rectify the system flaws of prediction of solar gleam and, on the other hand, influences the development on anticipating the gauge interval employing a group anticipating approach also inspected.

2. Meteorological Model and Solar Global Horizontal Irradiance Forecasting

2.1. Model WRF Configuration. As is shown previously in the research work [14], the performance with the WRF model, which means Weather Research and Forecasting by Skamarock et al. 2008 [43], was supervised from July 2013 to June 2014 for the prediction of inexhaustible energy resources, for example photovoltaics or wind power generations [44–47]. The model WRF is also applied in an extensive way, for the reason that it can be used to simulate the weather in reality with the technique of high resolution. The calculation back drop of meteorology with WRF are summed up in Table 1.

The simulation was driven based on the analysis of data in a global scope for the GSM-JP standing for area of Global Spectral Model for Japan. The spatial resolution of the analyses was $0.2^\circ \times 0.25^\circ$, while the temporal resolution was 3 h. The WRF exemplary composition constitute three consecutive installed realms with 18 kilometers of grid spacing, which is the grid in coarser style showed in domain 1 [48]. Also, the 6 kilometers of grid spacing will be showed in domain 2 and 2 kilometers of grid spacing is in domain 3, which is the best grid, accompanied with 50 vertical levels. Estimates which correspond to the most inside domain grid points (domain 3) enclosing the stations of experimental radiometric were put into use in the assessment.

Physical parameterizations in a wide range are enclosed in the model WRF, allowing setting the WRF model to depict the physical processes in a better way which are based on resolution, model domain, application, and location. Under such circumstances, various problems were chosen to pursue. Shimada et al. [49] supervised an assessment study of the achievement of distinct problems of the exemplary in the identical researching field.

Thus, the model WRF was functioned with the MYJ planetary boundary layer scheme [50], the Betts-Miller-Janjic cumulus scheme [51], WSM 6 microphysical scheme [52], the Noah land surface model, the scheme of Dudhia for shortwave radiation scheme [53], and the RRTM scheme for long-wave radiation scheme [54]. With regard to every day, a 3-day (72 h) ahead prediction was fulfilled, which started at 12 h Universal Coordinated Time (UTC).

2.2. Study Area and Observed Data. The central region of Japan is the study field for the solar global horizontal irradiance prediction; the research area includes Aichi, Nagano, Mie, Shizuoka, and Gifu Prefectures. Results of the present research have been assessed with the observation data which are collected at 61 radiometric observation stations displayed in Figure 1. The main reasons why these stations are chosen is because they are located near to regions where the solar power plants are running or under construction. The solar global horizontal irradiance (GHI) data were collected by EKO pyranometer MS–402 in these stations. In this research, to get data in 30 minutes, measurements in 10 seconds were averaged. Moreover, since researches were operated in one field where the commercial electric power grid locates, the GHI averaged data for GHI from all 61 stations of Figure 1 are used for the assessment. The data at night, which refer to time period from sunset in the previous day to sunrise in the intraday time, are excluded from the assessment [55–57].

2.3. Forecasted Results with WRF. With the method Kalman filter as a posttreatment technique of solar global horizontal irradiance, calculation by model WRF is correlated with the checked information. As is explained before, the solar GHI averaged values over 61 stations in the Figure 1 and with 30-minute frequency is used for the checking. The data at night are not used for the evaluation.

Figure 1 shows correlation charts between the forecasted solar global horizontal irradiance and the observation from the beginning in July 2013 to the end of June in 2014. Such statistical error indexes as biases, root mean square errors, and correlation coefficients, which are abbreviated as Bias, RMSE, and CORR respectively, are listed in Table 2. The WRF computes 72-h ahead forecasting in one operation, which includes the intraday $M_{i,0}$, the next-day $M_{i,1}$ and the 2-day ahead $M_{i,2}$ prediction. Subscript i indicates operation day of the forecasting. The labels "$M_{i,j}$" only refers to the simulated GHI with model WRF, and "$M_{i,j}$ – KF" refers to the result which is adjusted with method Kalman filter. The outcomes are studied separately for $M_{i,0}$, $M_{i,1}$, and $M_{i,2}$ forecasting in this figure and this table. For the same target day, the ensemble average value is computed as a mean of the members, which consist of $M_{i,0}$, $M_{i-1,1}$, and $M_{i-2,2}$, standing for the very day forecasting in the operation day i.

3. Kalman Filter for Improvement of WRF Forecasting

3.1. Method Kalman Filter. Designed to collect a signal in an efficient way from data in noisy environment that consists of statistical noise and other inexact ones, Kalman filter method is therefore expected to show a more exact outcome by using training data [40, 58, 59]. This approach is ordinarily employed to rectify systems flaws of weather forecast

Table 1: Configuration of the WRF model.

Input data	JMA GSM-JP (3-hourly, 0.2° × 0.25°) NCEP GFS (3-hourly, 0.5° × 0.5°)
Vertical layer	50 levels (surface to 100 hPa)
Nesting	2-way nesting
Domain	Domain 1, D01 (18 km, 120 × 110 grids), domain 2, D02 (6 km, 130 × 130 grids), and domain 3, D03 (2 km, 160 × 229 grids)
Physical options	Mellor-Yamada-Janjic PBL parameterization Betts-Miller-Janjic cumulus parameterization WSM 6-class graupel scheme Dudhia short wave radiation scheme RRTM long wave radiation scheme (domain 1 and domain 2 only) Noah land surface model

further and has been ratified as one of establishment approach of meteorology forecast supervision; such meteorological parameters as amount of precipitation, temperature, and probability of precipitation are studied based on this method, according to Kunitsugu [60]. The application of Kalman filter in the meteorological field is different from other fields. The correction coefficient of the observation and prediction error of solar irradiance are obtained, not physical quantities. Compared with other methods, Kalman filter is not so complex and is suitable for the solar GHI forecasting corrections. Meanwhile, the Kalman filter has not been widely applied in this field before. Also, the results of Kalman filter could be realized by MATLAB efficiently and conveniently.

By and large, previous values before the time k have an effect on the arbitrary instantaneous k time value to some extent. The method Kalman filter follows a trail of the previous data as the state of the system and renews it with the observation value. The Kalman filter method usually contains observation equations and state equations.

$$z_k = Hx_k + v_k, \tag{1}$$

$$x_k = Ax_{k-1} + w_k. \tag{2}$$

On the basis of state equation, equation (1), it is assumed in Kalman filter that the state at time k is generated from the previous one at time $k-1$. In equation (2), x_k refers to a state variable, A refers to a state transition matrix, and w_k refers to a state transition noise. On the basis of observation equation, equation (1), the observation data z_k is computed by utilizing the state variable x_k. H refers to a state-to-observation matrix and v_k means the noisy data in the observation. It is assumed that noise plays a key role in method Kalman filter. The two white noises w_k and v_k are compliant with Gaussian zero mean.

Figure 2 displays the calculation flow in Kalman filter algorithm which is described as equations, referring to Kim and Huh [61]. The first step is to set the initial values, which include the error covariance matrix P_0, an internal variable initial value, and evaluated state variable $\hat{x}_0$. Then, the estimated state variable $\hat{x}_k$ at time k is evaluated and updated with the observation data z_k by following the algorithm.

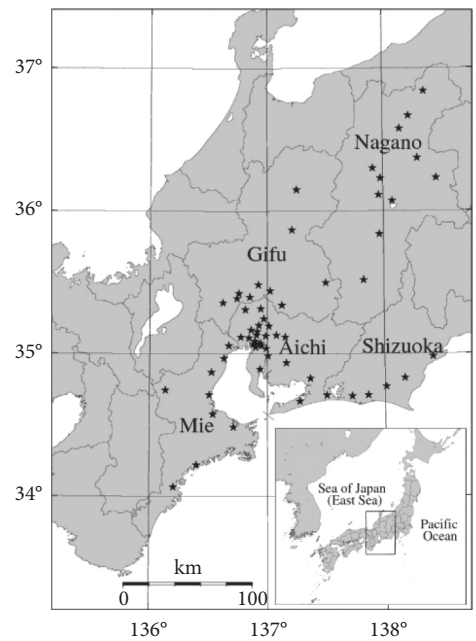


Figure 1: Study region and distribution of the radiometric stations with global irradiance measurements (dark stars).

In this case, we make use of the method Kalman filter to correct the unfairness of the predicted solar global horizontal irradiance in accordance with the system flaws of prediction. Predicting the future bias by using a linear relationship between the estimated previous bias present and the forecasting error, Kalman filter is such a mathematical approach [49]. After the estimation of current bias, to ameliorate the forecasting, it is detached from the solar global horizontal irradiance which is forecasted with model WRF. It is assumed that, the simulated solar global horizontal irradiance with model Kalman filter is assessed from forecasted GHI based upon model WRF according to the following equation:

$$\hat{I}_{\mathrm{WRF}} = aI_{\mathrm{WRF}} + b, \tag{3}$$

where a and b are correction coefficients. On this assumption, the state variable is defined as $x_k = \begin{pmatrix} a & b \end{pmatrix}^T$. Also, the observation equations and the state corresponding to equations (1) and (2) can be inferred in the following equations.

Table 2: Statistical indexes for solar GHI different forecasts.

Forecasting method and member	Bias [W/m^2] (Relative bias)	RMSE [W/m^2] (Relative RMSE)	CORR [−]
$M_{i,0}$	49.7 (28.8%)	102.9 (59.6%)	0.950
$M_{i,0}$ − KF	−2.4 (−1.4%)	79.1 (45.8%)	0.951
$M_{i,1}$	51.4 (29.7%)	118.9 (68.9%)	0.929
$M_{i,1}$ − KF	−6.0 (-3.5%)	94.6 (54.7%)	0.930
$M_{i,2}$	50.8 (29.4%)	129.8 (75.2%)	0.911
$M_{i,2}$ − KF	−8.5 (−4.9%)	106.6 (61.7)	0.911
Ensemble mean	50.6 (29.3%)	104.9 (60.7%)	0.946
Ensemble mean−KF	−5.6 (−3.2%)	81.8 (47.4%)	0.947
Persistent	0.3 (0.15%)	191.7 (110.9%)	0.718

$$z_k = Hx_k + v_k, \quad = \begin{pmatrix} I_{\mathrm{WRF}} & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}_k + v_k, \tag{4}$$

$$x_k = Ax_{k-1} + w_k, \quad = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}_{k-1} + w_k, \tag{5}$$

where $z_k = \widehat{I}_{\mathrm{WRF}}$. With regard to the function of Kalman filter, v_k in the equation (4) and its relevant covariance matrix R, and w_k in the equation (5) and its relevant covariance matrix Q in the equations Figure 2 are essential impact factors. To perform different WRF simulations, various matrices R and Q are input to gain their appropriate values. Figure 3 displays the solar GHI's RMSE during the intraday forecasting under different matrices R and Q, which reduce the predicted RMSE. According to Figure 3, the adequate rate between the impact factors R and Q could be found out. From this result, we gain that R is equal to $1 \times 10^{11.5}$ and Q is equal to $1 \times 10^0 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

The time series of the correction coefficients a and b in the equation (3) is displayed in Figure 4. Compared with the solar GHI's order about 100 or 1,000 W/m^2, b can be ignored. From this figure, b increases over time, and a varies by season. In the filter, the two coefficients vary with time, working to regulate the solar global horizontal irradiance which is predicted with Model WRF.

3.2. Forecasting Results with WRF and Kalman Filter. Figure 5 shows correlation diagrams of the observed GHI and the forecasted data. Ensemble means shown in Figures 5(g) and 5(h) are irradiances averaged in the same target day. Also, the persistent model as a reference one is displayed in Figure 5(i). Table 2 enlists scientific mistakes indices of weather prediction. Both of the results in CORRs and RMSEs based on model WRF forecasts are superior to those which are acquired by the persistent model. The validity of the forecasting with model WRF are uncovered in this result. But the bias value of model WRF forecasts is larger than that of the persistent model, which is zero in theory.

Biases of $M_{i,0}$, $M_{i,1}$ and $M_{i,2}$ forecasting by using model WRF are 49.7 W/m^2, 51.4 W/m^2, and 50.8 W/m^2, respectively. According to these outcomes, a notable positive bias is found in the solar GHI forecasting with model WRF, which signifies that the solar GHI with WRF model is overestimated [62–65].

As found in this figure and this table, the bias values are decreased clearly by utilizing the method Kalman filter; for instance, with the help of method Kalman filter, the bias of $M_{i,0}$ forecasting turns from 49.7 W/m^2 to −2.4 W/m^2. By comparing with the solar irradiance mean intensity, the biases could be practically negligible. With the use of method Kalman filter, the two indices RMSEs and biases decrease while CORRs raise as is shown in Table 2. Those relative RMSEs and biases in this table are normalized with their mean values. These relative indices are also improved by applying Kalman filter. The function of the method Kalman filter to the forecast precision is proved with the amelioration of these indices.

The correction coefficient a of the Kalman filter in Figure 4 is nearly 0.88, and smaller than unity. The method Kalman filter decreases the solar global horizontal irradiance intensity which is predicted with model WRF and rectifies the overestimation of model WRF output with the value a.

The intraday forecasting of daily operation is associated and stated in the form of time series in Figure 6. The observed solar global horizontal irradiance is also plotted in this figure. The irradiance which is forecasted with model WRF seems a little greater than the measured one, but after

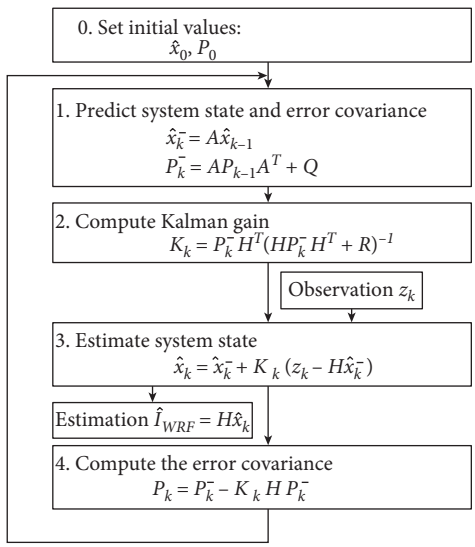


Figure 2: Algorithm of Kalman filter.

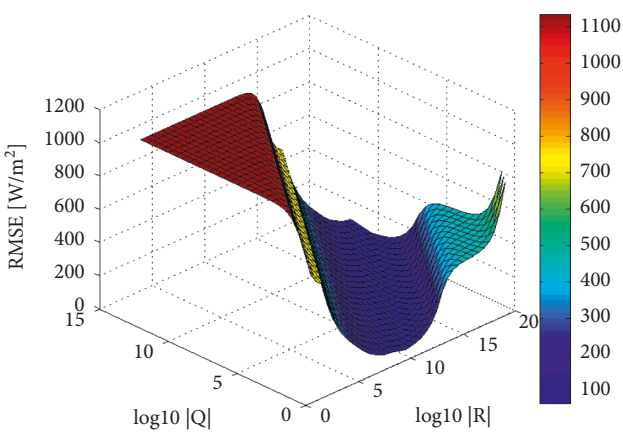


Figure 3: RMSE of the solar GHI intraday forecasting under the different values of covariance matrices R and Q.

the application of Kalman filter to the WRF's results, the evaluated irradiance reduces and approaches the observation data. The method Kalman filter has an effect on the amelioration of the solar global horizontal irradiance forecast shown in this figure.

4. Solar Irradiance Ensemble Prediction

4.1. Ensemble Prediction Method and Forecasting Conditions. Everyday group prediction of solar global horizontal irradiance is controlled with WRF, similar to previous work proposed by Liu et al. [14], with the exception of the application of method Kalman filter, which is a postprocessor for the WRF model. We use similar calculations and circumstances as the latter case in order to contrast [66–71]. The detailed process of the ensemble prediction is stated in the previous research [14]; therefore, we give only the outlines here [72].

The lagged averaged forecast (LAF) which is made up of members with distinct original situation and time is used for group prediction. Figure 7 shows the chart of the cycle of operation for the ensemble prediction of solar GHI. $M_{i,j}$ – KF indicate the members of the ensemble prediction based on WRF output and Kalman filter, in which i refers to the running day and amounts to the day of the termination of calculation, and j represents the scope of the forecasting day. Hence, $j = 0$ refers to the result of the intraday; $j = 1$ refers to the result of the next-day; and $j = 2$ refers to the results of the 2-day ahead forecasting from the running day. As the figure displayed, GHI of the intraday forecasting in the running day i is predicted based on three members: $M_{i,0}$ – KF, $M_{i-1,1}$ – KF, and $M_{i-2,2} - KF$. GHI of the next-day forecasting in the running day i is predicted based on two members: $M_{i,1}$ – KF and $M_{i-1,2}$ – KF. GHI of the 2-day ahead forecasting in the running day i is predicted based on one member: $M_{i,2}$ – KF.

The spread of the forecasting members is also defined with the following equation:

$$\text{Spread}_{i,j} = \sqrt{\frac{1}{N-j} \sum_{k=0}^{N-j-1} \left(\text{GHI}_{i-k,j+k} - \overline{\text{GHI}}\right)^2}, \qquad (6)$$

where j equals to 0, 1, and 2; N represents the number of forecasting day in one running, and it is equal to 3 in this work; $N - j$ means the member number of ensemble prediction; $\overline{\text{GHI}}$ is the averaged GHI value of the ensemble members; and $\text{GHI}_{i-k,j+k}$ refers to the predicted GHI value of the member $M_{i-k,j+k}$. This spread serves as the evaluation factors of uncertainty of the forecasting [14].

4.2. Solar Global Horizontal Irradiance Ensemble Prediction. The certainty of the outcome pointed out for solar global horizontal irradiance prediction is examined in Figure 5 and Table 2. It is found that Bias, RMSE, and CORR, the statistical error indices of $M_{i,0}$ forecasting, are the optimal in those figure and table. In general, in the ensemble prediction method, the ensemble average is selected as the predicted value, brief noted as ensemble mean–KF in this research. Nevertheless, $M_{i,0}$ – KF, which represents the intraday forecasting value based on WRF and Kalman filter, is used as the predicted value in this work, for the reason that the precision of the intraday forecasting is optimum.

4.3. Evaluation Index of Ensemble Prediction. Figure 8 displays the link in between the distribution within a day prediction and its prediction flaws from the examination throughout the time from September 1 to November 30 2013. Also, the top and bottom peripheries in the prediction interval whose coverage rates are 50 percent, 80 percent, 90 percent, and 95 percent respectively are linked with lines in this chart. These lines are assessed during per 10 W/m^2 in the target spread on the one hand and, on the other hand, are assessed negative and positive errors, respectively. The plotted data gather around the original point of the chart. Meanwhile, the forecasting error values deviate in a wide way with the target spread getting large as figured with the periphery lines of the extents.

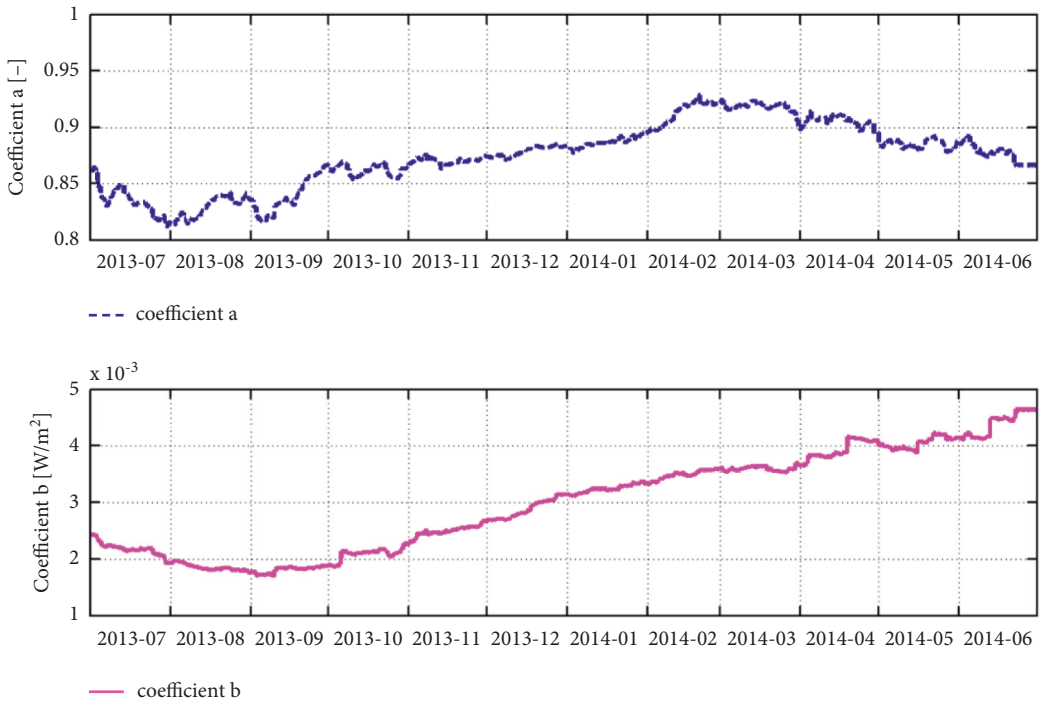


Figure 4: Time series of correction coefficients in the method Kalman filter for the intraday forecasting during the period from July 2013 to July 2014.

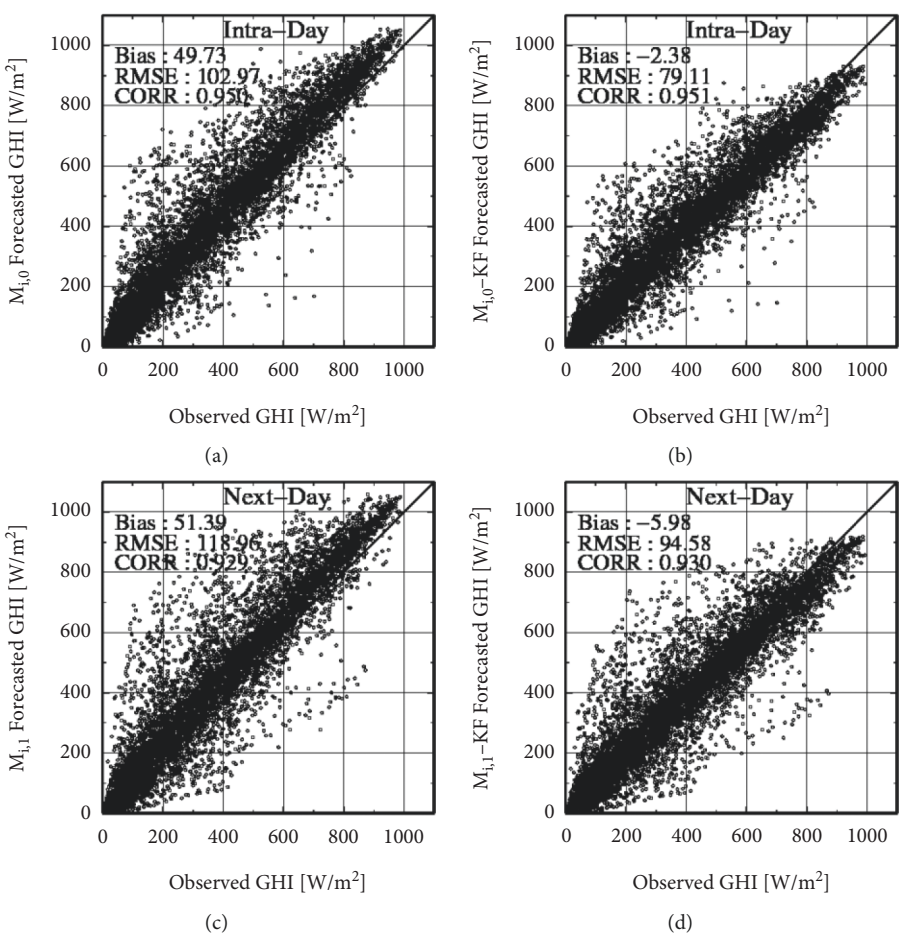


Figure 5: Continued.

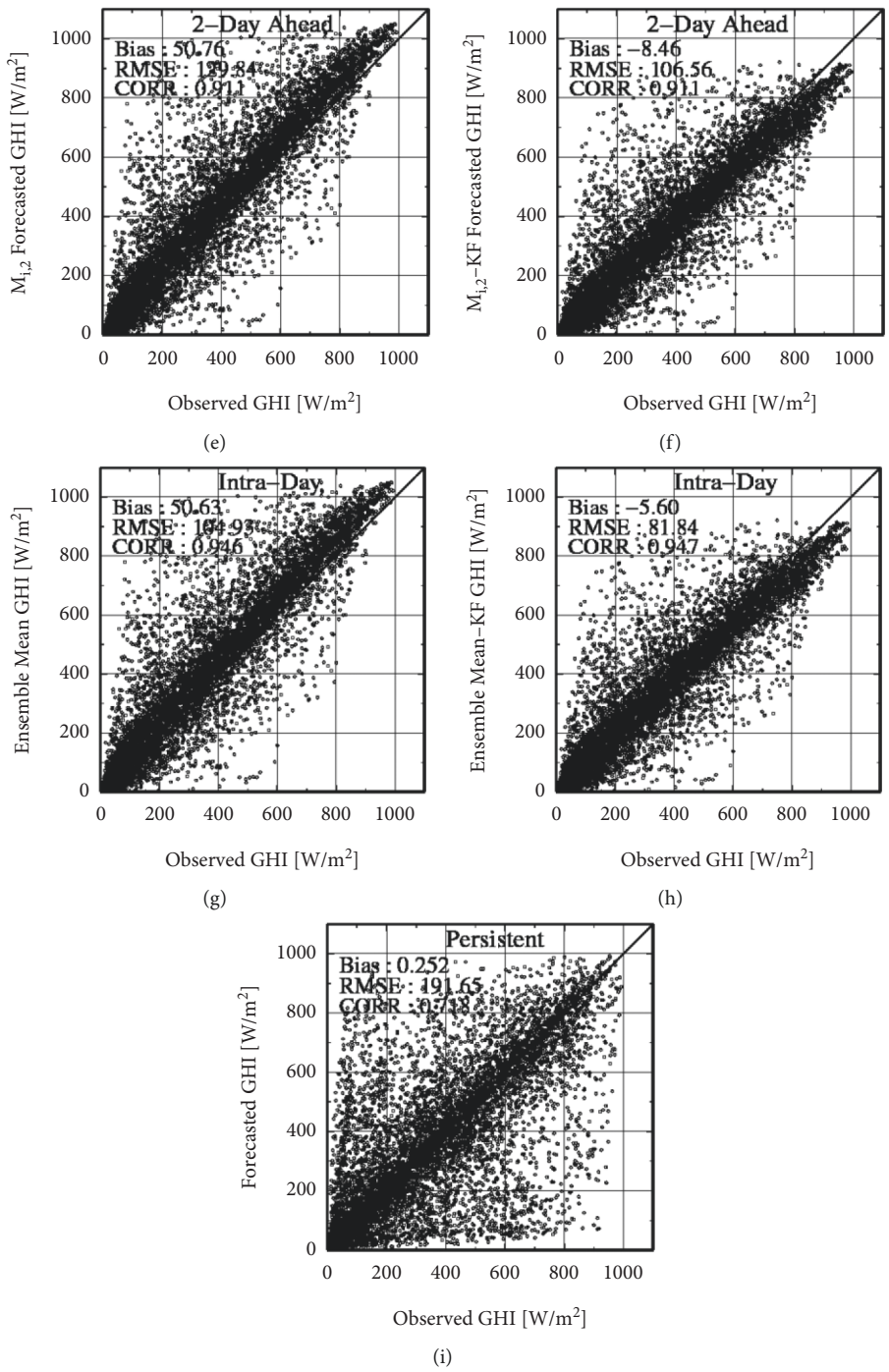


Figure 5: Correlation diagrams of the observed GHI and the forecasted data for the intraday prediction, next-day prediction, 2-day ahead prediction, and ensemble mean with the meteorological model WRF only ($M_{i,0}$, $M_{i,1}$, $M_{i,2}$, and ensemble mean), with WRF and post-treatment Kalman filter ($M_{i,0}$ – KF, $M_{i,1}$ – KF, $M_{i,2}$ – KF, and ensemble mean–KF), and the persistent prediction. (a) $M_{i,0}$ prediction, (b) $M_{i,0}$ – KF prediction, (c) $M_{i,1}$ prediction, (d) $M_{i,1}$ – KF prediction, (e) $M_{i,2}$ prediction, (f) $M_{i,2}$ – KF prediction, (g) ensemble mean, (h) ensemble mean –KF, and (i) persistent.

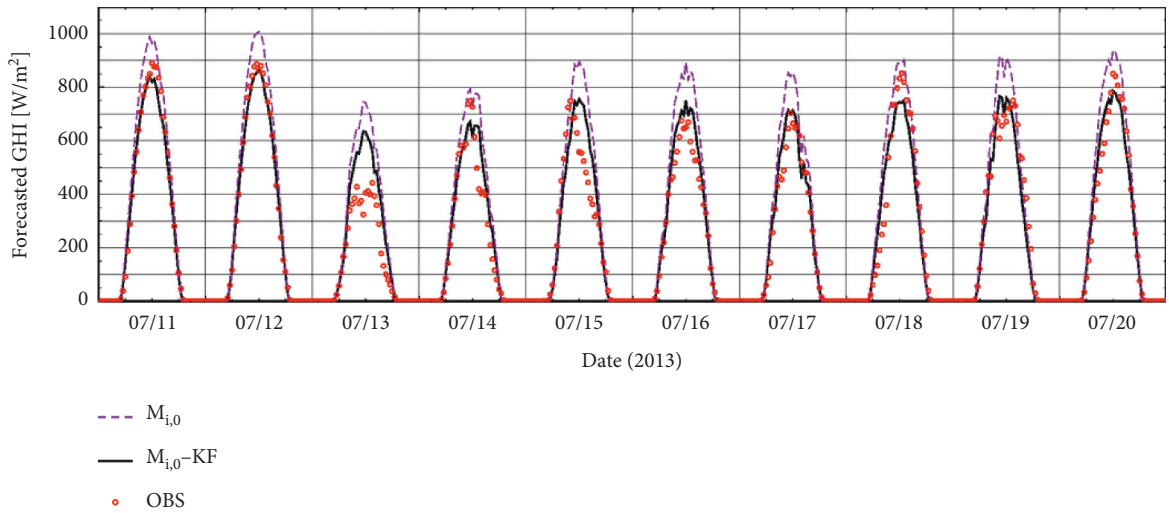


Figure 6: Results of intraday prediction of the solar GHI (global horizontal irradiance) by using WRF only $M_{i,0}$, based on WRF and method Kalman filter $M_{i,0}$ – KF and observed one during the period from July 11 to 20 in 2013.

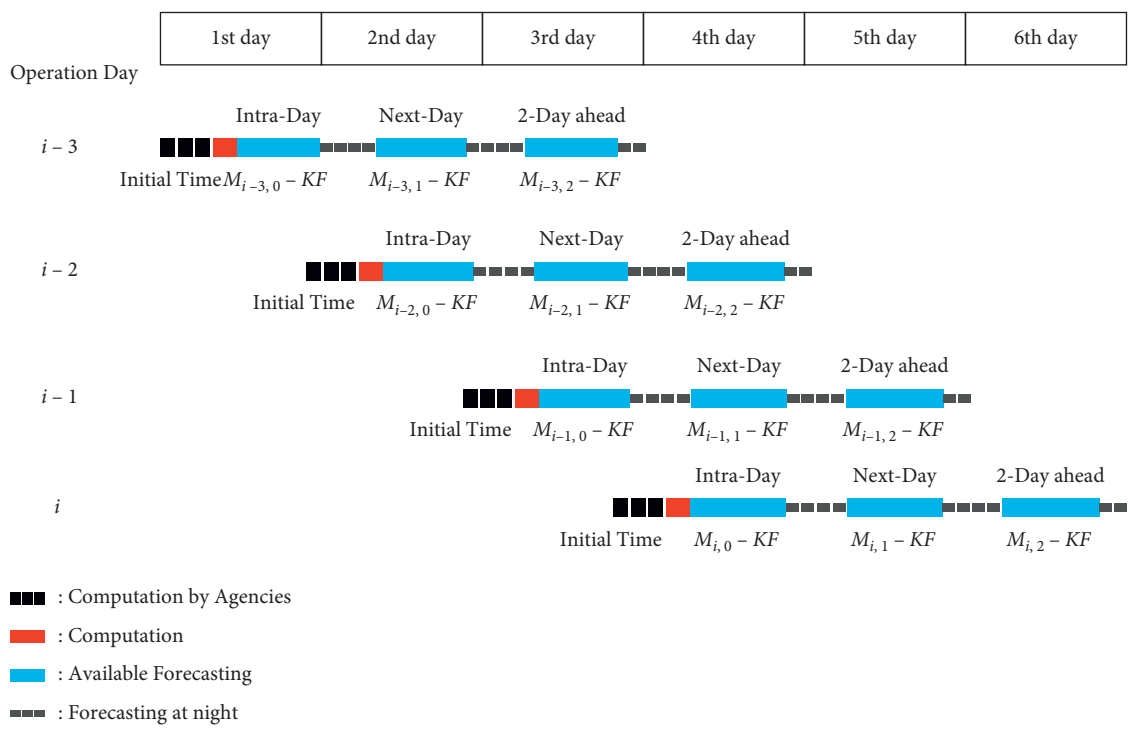


Figure 7: Cycle of operation for the solar GHI ensemble prediction based upon model WRF and method Kalman filter.

The periphery lines in the prediction extents are coarse, for the reason that three is the limitation for the ensemble member with the target spread containing random error, which results in the limitation range of the computing environment in this project. The increasement of ensemble members makes contribution to the amelioration of the interval evaluation [73–76].

The distribution of scatter diagrams and the periphery lines of $M_{i,0}$, which is the calculated outcomes with model WRF, deflects to the positive side, because WRF has the tendency to overestimate the solar global horizontal irradiance as explained before. The pure WRF forecasting is symmetric with respect to the horizontal line 30 W/m^2. On the other hand, the plots and the periphery lines, with an exception of the lines whose coverage rate is 95 percent, distribute in a symmetrical way relative to the axis of abscissas. This is the result of method Kalman filter, which corrects the overestimating calculation of model WRF.

4.4. Solar Global Horizontal Irradiance Forecasting with Prediction Interval. Every day solar global horizontal irradiance ensemble predictioning is performed, and the solar lustrous and its prognosis interval are forethought. Figure 9

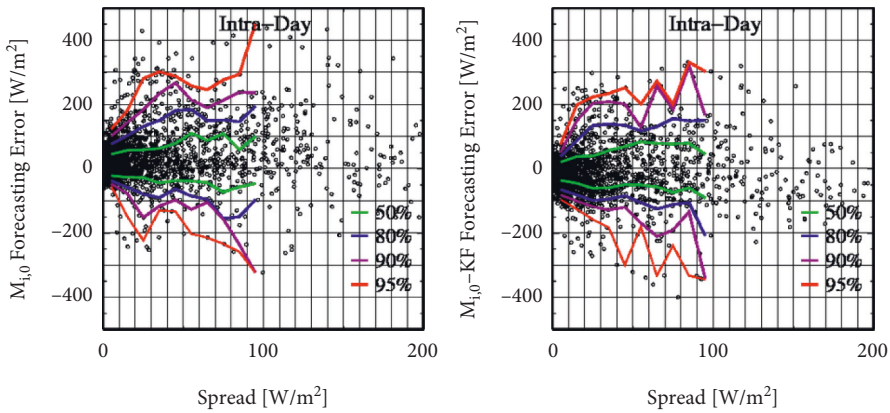


Figure 8: Relationship chart between the ensemble spread and solar GHI's forecasting error for the intraday prediction by using model WRF only $M_{i,0}$, based on WRF and method Kalman filter $M_{i,0}$ − KF during the period from September 1 to November 30 in 2013.

shows the irradiance forecasted with WRF only, and Figure 10 displays the irradiance with model WRF and method Kalman filter. In these charts, the calculated outcomes of the intraday forecasting data are linked up and shown during the time series between December 1 and December 10 2013. By respectively using black full lines and red dot lines in these charts, the data of the predicted irradiance and the data of observed ones are plotted. As a result of the overestimation of model WRF, the outcome of irradiance which is predicted with model WRF is greater than the observed data in Figure 9. On the other hand, the forecasting results with WRF and Kalman filter is suited to the observed data as indicated in Figure 10.

The ensemble spread of the members with model WRF and based on model WRF and method Kalman filter is also displayed by using dashed lines in green in Figures 9(a) and 10(a). The target spread value alters with time; for instance, the target spread on December 6 is not small, while the target spread on December 3 is close to zero in these charts. As the ensemble spread becomes smaller, the predicted GHI values come closer to the observed one.

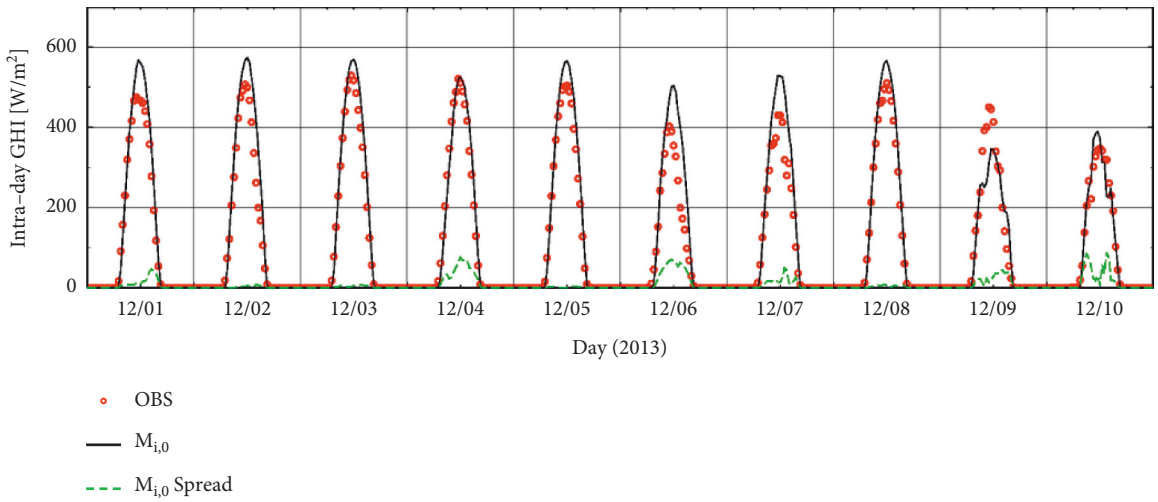
The top and bottom peripheries in the prediction interval are displayed by using two dashed lines in blue which situate from Figures 9(b)–9(e) and from Figures 10(b)–10(e), whose coverage rates are respectively 50 percent, 80 percent, 90 percent, and 95 percent. For the sake of evaluating the peripheries of the predicted intervals, the top and bottom peripheries in the prediction interval in Figure 8(a) or 8(b) are assessed based on the given rates of nominal coverage and the target spread in Figures 9(a) or 10(a) at first. And, defined as the prediction interval, they are drawn based on the forecasted irradiance. In these charts, accompanied with the decrease of target spread, the interval size, which can be found from the data on December third, is small. On the contrary, the interval size becomes large when the target spread increases, just like the data on December 6. As is displayed in these charts, the rate of nominal coverage gets large compared with the interval size in proportional way.

Table 3 displays the rate of daily empirical coverage, which refers to the ratio of the actual data number of observed data fallen in the estimated interval, to the total data number within the effective sunshine time, for the intraday solar GHI ensemble prediction during the period from December 1 to 10 in 2013. $M_{i,0}$ represents the predicted result based on model WRF, and $M_{i,0}$ − KF represents the predicted result based on model WRF and method Kalman filter in this table. It is revealed that the rate of empirical coverage is in proportion to the rate of nominal coverage, which is greater than the rate of daily empirical coverage of $M_{i,0}$ and approached by the rate of daily empirical coverage of $M_{i,0}$ − KF.

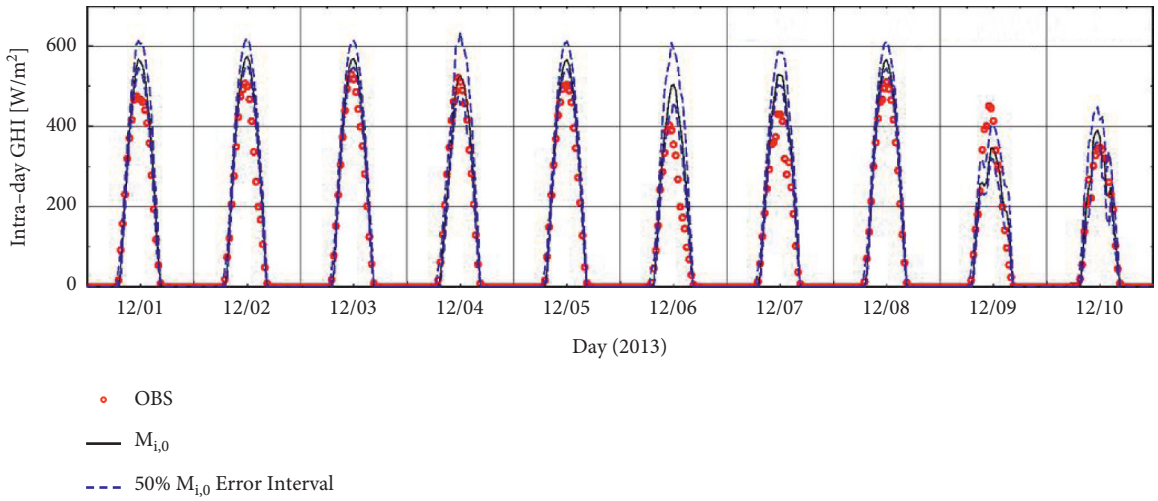
During the period from December 1 to 10 2013, Table 4 shows the daily prediction interval sizes for the intraday solar GHI ensemble prediction. For the size computation, the data at night are refused. It is revealed that the sizes get in proportion to the rates of nominal coverage in this table. The sizes of $M_{i,0}$ − KF forecasting is smaller than the ones of $M_{i,0}$ forecasting.

To improve the WRF calculation, solar global horizontal irradiance via Kalman filter is used. Applying it to the ensemble prediction brings about the size of the prognosis interval gradually little and the factual analysis quota rises and closes to the nominal one. It signifies that the amelioration of the forecasting provides a better assessment for prediction interval.

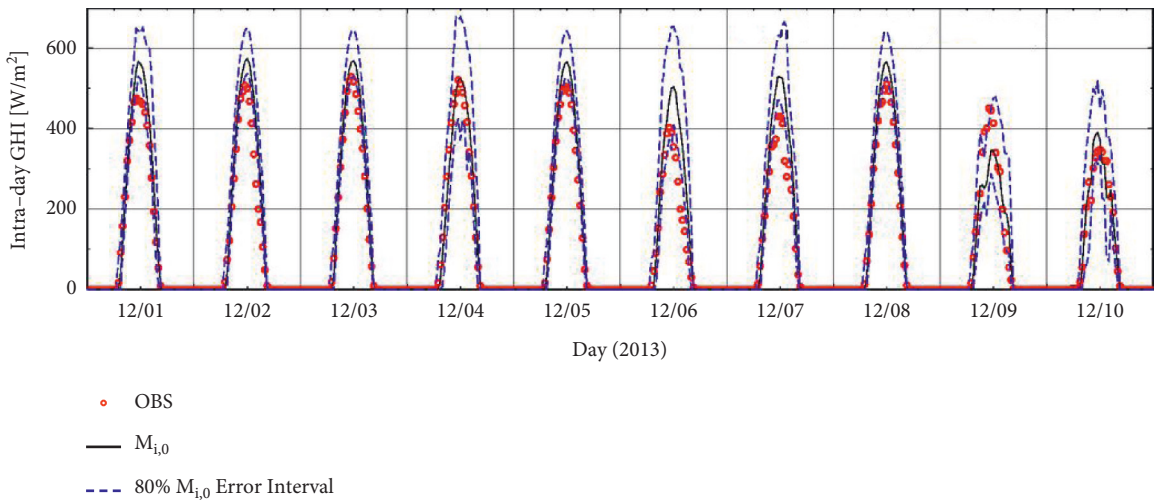
Furthermore, monthly empirical coverage rates and prediction interval sizes from October 2013 to June 2014 are illustrated in Figures 11 and 12. With the application of Kalman filter in improving the solar GHI forecasting, the prediction interval sizes get small, while the rates of empirical coverage add and become near the nominal rates in the long period. It means that the improvement in the forecasting works well for the improvement in the prediction interval estimation.



(a)

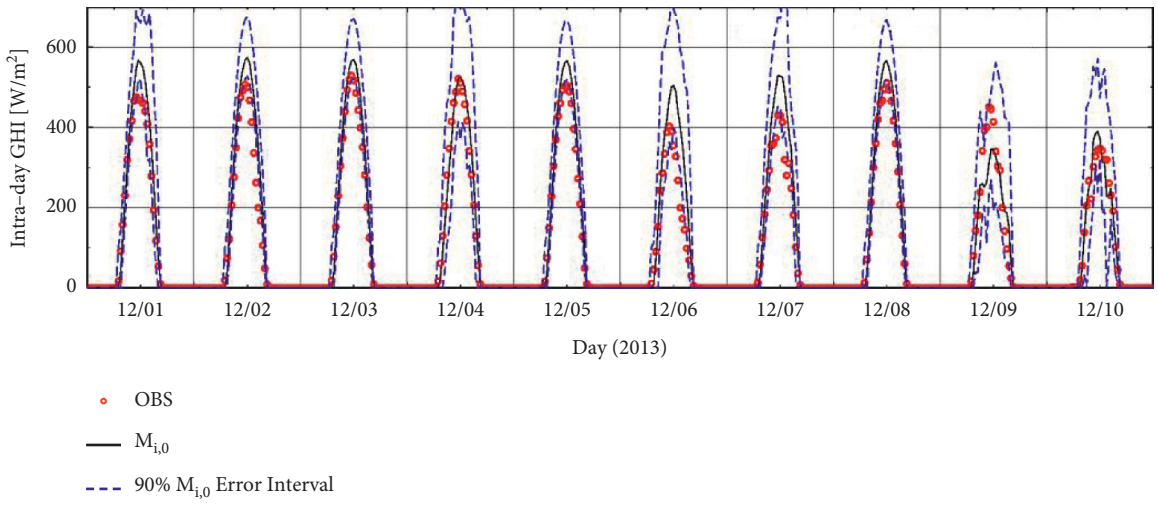


(b)

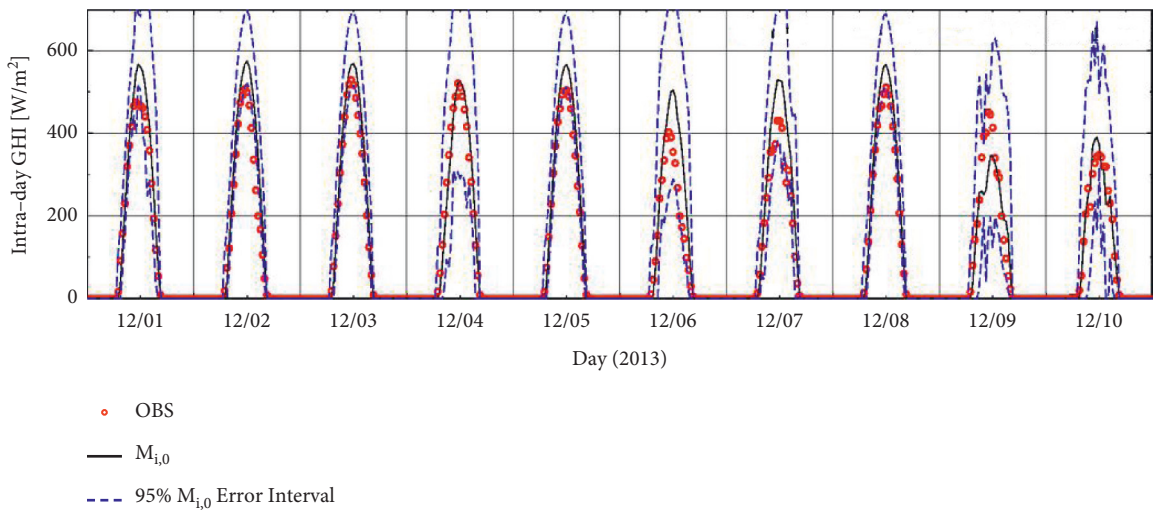


(c)

Figure 9: Continued.

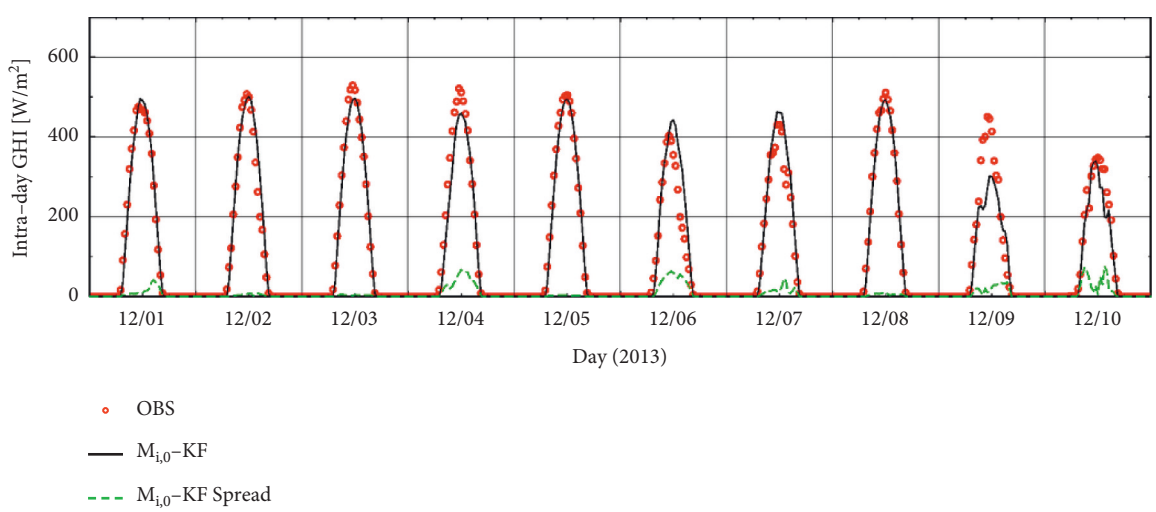


(d)



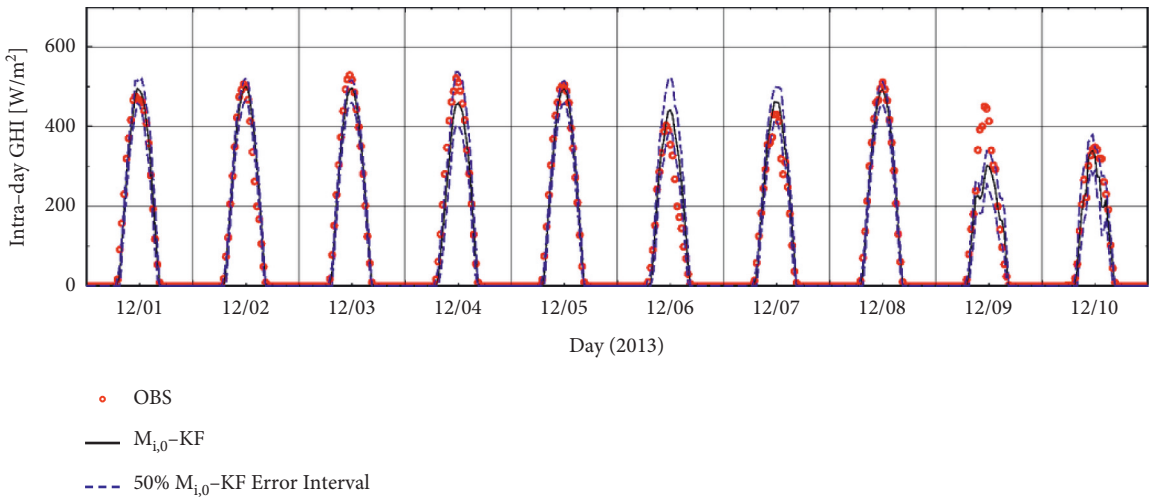
(e)

Figure 9: Time series of solar GHI intraday prediction by using model WRF only $M_{i,0}$ and its 50%, 80%, 90%, and 95% prediction interval during the period from December 1 to 10 in 2013. (a) $M_{i,0}$ prediction and ensemble spread, (b) $M_{i,0}$ prediction and 50% prediction interval, (c) $M_{i,0}$ prediction and 80% prediction interval, (d) $M_{i,0}$ prediction and 90% prediction interval, and (e) $M_{i,0}$ prediction and 95% prediction interval.

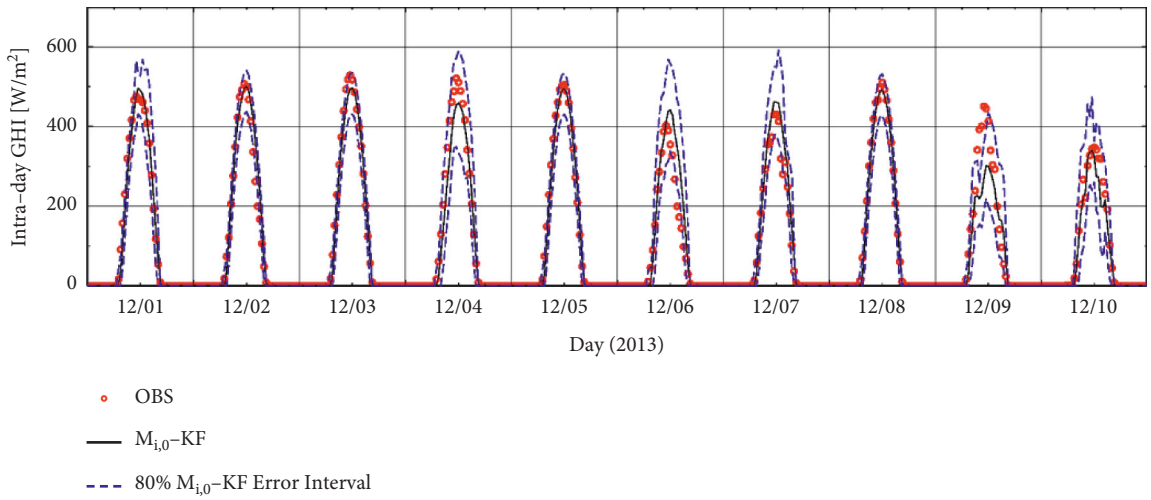


(a)

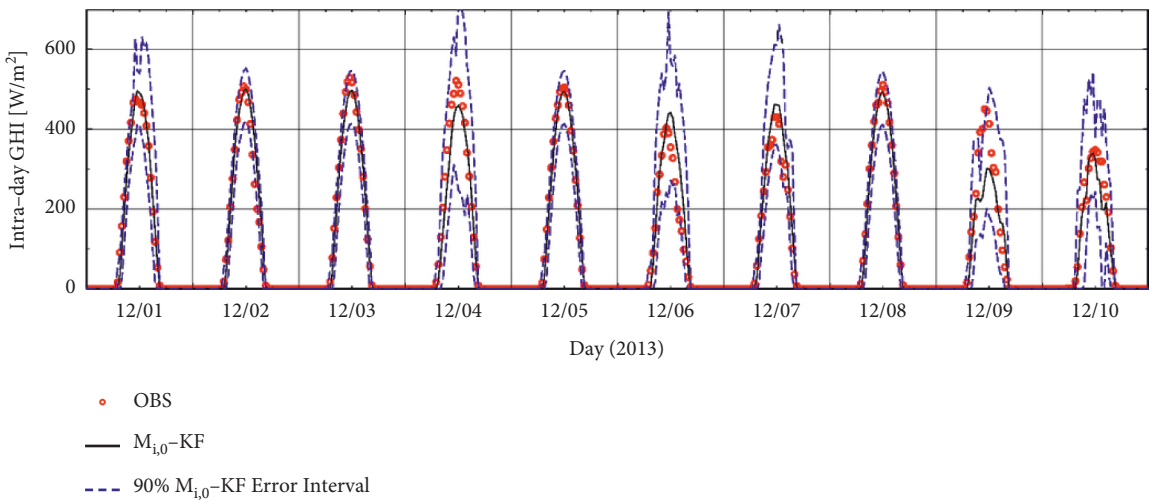
Figure 10: Continued.



(b)

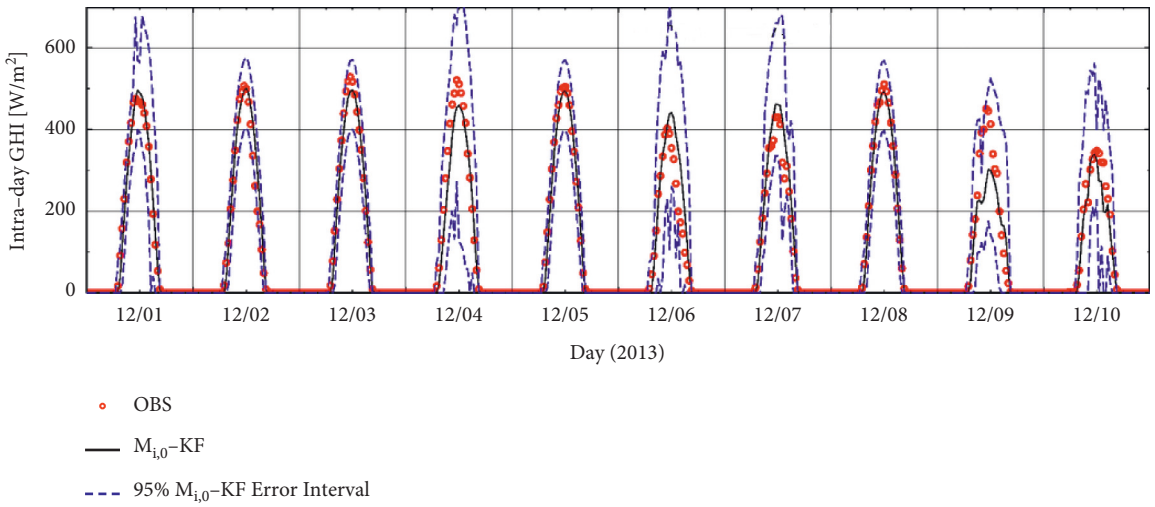


(c)



(d)

Figure 10: Continued.



(e)

Figure 10: Time series of solar GHI intraday prediction $M_{i,0}$ – KF based on model WRF and method Kalman filter, and its 50%, 80%, 90%, and 95% prediction interval during the period from December 1 to 10 in 2013. (a) $M_{i,0}$ – KF prediction and ensemble spread, (b) $M_{i,0}$ – KF prediction and 50% prediction interval, (c) $M_{i,0}$ – KF prediction and 80% prediction interval, (d) $M_{i,0}$ – KF prediction and 90% prediction interval, and (e) $M_{i,0}$ – KF prediction and 95% prediction interval.

Table 3: Rates of empirical coverage for the intraday solar GHI prediction with regard to the prediction interval of the ensemble prediction during the period from December 1 to 10th 2013.

Rate of nominal coverage for prediction interval (%)	Forecasting method and member	Rate of empirical coverage for prediction interval [%]										
		12/01	12/02	12/03	12/04	12/05	12/06	12/07	12/08	12/09	12/10	Ave
50	$M_{i,0}$	31.6	10.5	31.6	73.7	26.3	26.3	21.1	26.3	36.8	63.2	34.7
	$M_{i,0}$ – KF	57.9	52.6	42.1	57.9	57.9	47.4	63.1	57.9	36.8	68.4	54.2
80	$M_{i,0}$	63.2	26.3	52.6	89.5	36.8	57.9	36.8	36.8	73.7	94.7	56.8
	$M_{i,0}$ – KF	73.7	78.9	78.9	89.5	84.2	68.4	84.2	89.5	63.1	94.7	80.5
90	$M_{i,0}$	73.7	42.1	57.9	94.7	47.4	57.9	57.9	47.4	89.5	94.7	66.3
	$M_{i,0}$ – KF	84.2	78.9	94.7	94.7	94.7	89.5	94.7	94.7	78.9	94.7	90.0
95	$M_{i,0}$	84.2	47.4	63.2	94.7	52.6	94.7	78.9	47.4	89.5	94.7	74.7
	$M_{i,0}$ – KF	94.7	84.2	94.7	94.7	94.7	94.7	94.7	94.7	89.5	94.7	93.2

Table 4: Prediction interval sizes for the intraday solar GHI prediction with regard to the prediction interval of the ensemble prediction during the period from December 1 to 10 in 2013.

Rate of nominal coverage for prediction interval (%)	Prediction method and member	Prediction interval size [W/m^2]										
		12/01	12/02	12/03	12/04	12/05	12/06	12/07	12/08	12/09	12/10	Ave
50	$M_{i,0}$	79.9	65.4	65.4	112.9	65.4	120.9	82.9	65.4	90.8	104.7	85.4
	$M_{i,0}$ – KF	76.7	55.0	55.0	107.9	54.9	109.7	78.3	54.9	91.6	98.9	78.3
80	$M_{i,0}$	155.6	111.8	111.8	231.9	111.8	232.1	164.5	111.7	196.7	208.0	163.6
	$M_{i,0}$ – KF	147.2	99.6	99.6	212.2	99.5	196.9	155.8	99.4	188.4	190.6	148.9
90	$M_{i,0}$	217.0	144.5	144.5	316.4	144.4	304.6	233.3	144.3	287.1	284.4	222.1
	$M_{i,0}$ – KF	208.6	127.1	126.9	328.0	126.9	290.3	223.1	126.8	272.8	279.0	210.9
95	$M_{i,0}$	274.8	171.9	171.9	427.6	171.8	408.2	320.7	171.8	381.3	378.4	287.9
	$M_{i,0}$ – KF	266.8	163.5	163.3	431.1	163.3	398.9	288.6	163.2	328.8	338.1	270.6

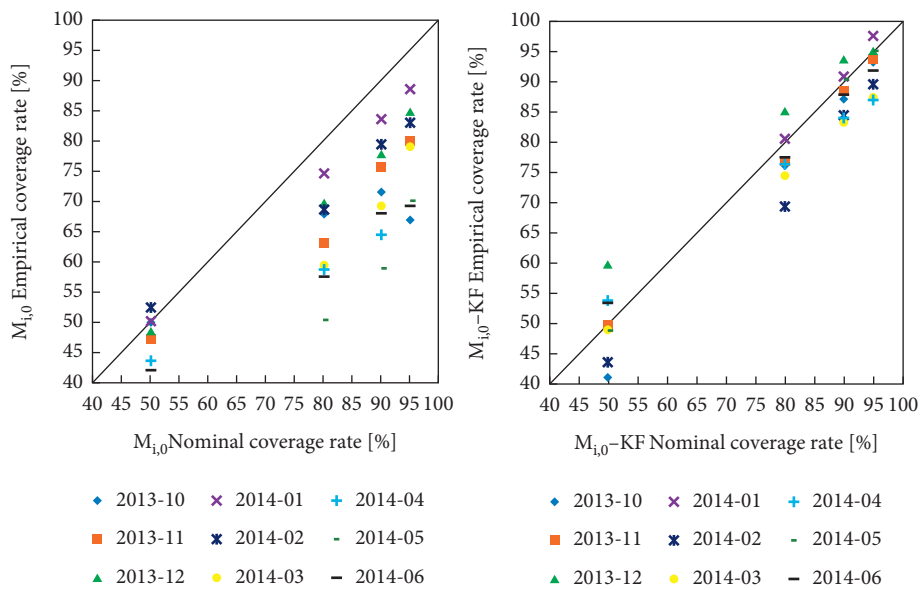


Figure 11: Association diagram between monthly rate of empirical coverage and monthly rate of nominal coverage for the intraday solar GHI prediction with regard to the prediction interval of the ensemble prediction during the period from October 2013 to June 2014.

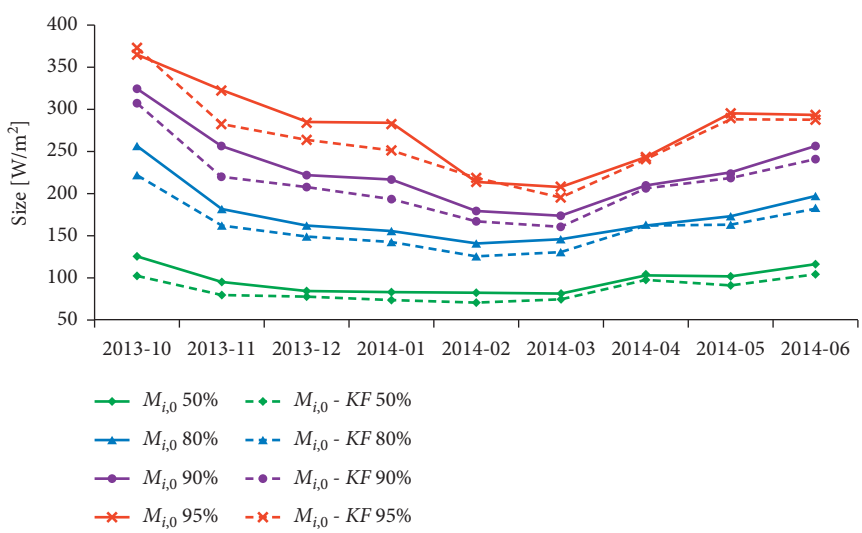


Figure 12: Monthly sizes of prediction interval for the intraday solar GHI prediction with regard to the prediction interval of the ensemble prediction during the period from October 2013 to June 2014.

5. Conclusion

This report portraits how solar global horizontal irradiance is accomplished with weather forecast model WRF and collecting data approach Kalman filter. The Kalman filter get rid of aberration of the irradiance calculated with WRF and enhance prognosticating. Also, the ensemble prediction is applied to the forecasting of the solar global horizontal irradiance as well as its prediction interval. The spread of the ensemble forecasting is calculated as a parameter corresponding to the unreliability of the forecasting. And the relation between the spread and the forecasting error is discussed to evaluate the prediction interval. With the prediction interval sizes reducing, its rate of empirical coverage raises and comes close to the nominal ones by applying the method Kalman filter. In the ensemble prediction, the amelioration through putting the method Kalman filter into use does not fit for the forecasting simply. Meanwhile, it is also suited for the prediction interval assessment. The probabilistic forecasting of solar irradiance presented in this article will be used to predict photovoltaic system generation for electric power grid management and its risk analysis. Compared with other methods, the first highlight of this article is the application of ensemble prediction method to achieve the probabilistic forecasting of solar irradiance. The second highlight is that the Kalman filter is applied into the prediction accuracy of solar irradiance, which is an innovation in solar irradiance prediction research, making the results more exact and reliable.

Further work is to improve the solar GHI prediction by using some interpretable machine learning tools to quantify the parameter importance. The work also seeks to achieve uncertainty quantification of the solar GHI prediction by transferred recurrent neural network for battery calendar health prognostics of energy transportation systems. Furthermore, it is also of great interest to compare and analyze the evaluation of the photovoltaic system installation impact to an electric power grid.

Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest

It is declared by the authors that this article is free of conflicts of interest.

Acknowledgments

The author was supported by the Scientific and Technological Innovation Programs of Higher Education Institutions in Shanxi Province (No. 2019L0983) and Introduction of High-Level Science and Technology Talent Program of Luliang Development Area in Shanxi Province (No. 2019107). The observed solar global horizontal irradiance (Global Horizontal Irradiance, GHI) data used in the work are measured by the Chubu Electric Power Co. Inc. under PV300 project, supported by the Ministry of Economy, Trade and Industry of Japan.

References

[1] Y. X. Xia, M. H. Feng, S. W. Wang, and C. Lu, "Risk assessment method of rock burst based on the combination of theory and field detection," *Journal of Mining and Strata Control Engineering*, vol. 3, no. 4, Article ID 043017, 2021.

[2] N. Z. Xu and C. Gao, "Study on the special rules of surface subsidence affected by normal faults," *Journal of Mining and Strata Control Engineering*, vol. 2, no. 1, Article ID 011007, 2020.

[3] X. Y. Yu, Z. S. Wang, Y. Yang, and X. W. Mao, "Numerical study on the movement rule of overburden in fully mechanized caving mining with thick depth and high mining height," *Journal of Mining and Strata Control Engineering*, vol. 3, no. 1, Article ID 013533, 2021.

[4] J. D. J. Rubio, "Stability analysis of the modified Levenberg-Marquardt algorithm for the artificial neural network training," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 8, pp. 3510–3524, 2021.

[5] J. D. J. Rubio, E. Lughofer, J. Pieper et al., "Adapting H-infinity controller for the desired reference tracking of the sphere position in the maglev process," *Information Sciences*, vol. 569, pp. 669–686, 2021.

[6] H. S. Chiang, M. Y. Chen, and Y. J. Huang, "Wavelet-based EEG processing for epilepsy detection using fuzzy entropy and associative petri net," *IEEE Access*, vol. 7, pp. 103255–103262, 2019.

[7] J. D. J. Rubio, M. A. Islas, G. Ochoa, D. R. Cruz, E. Garcia, and J. Pacheco, "Convergent newton method and neural network for the electric energy usage prediction," *Information Sciences*, vol. 585, pp. 89–112, 2022.

[8] D. M. Vargas, "Superpixels extraction by an Intuitionistic fuzzy clustering algorithm," *Journal of Applied Research and Technology*, vol. 19, no. 2, pp. 140–152, 2021.

[9] A. López González, J. A. Meda Campaña, E. G. Hernández Martínez, and P. P. Contro, "Multi robot distance based formation using parallel genetic algorithm," *Applied Soft Computing*, vol. 86, Article ID 105929, 2020.

[10] E. Lorenz, J. Hurka, D. Heinemann, and H. G. Beyer, "Irradiance forecasting for the power prediction of grid-connected photovoltaic systems," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 2, no. 1, pp. 2–10, 2009.

[11] J. Hashimoto and T. Kobayashi, "Solar spectral irradiance model for forecast of photovoltaic power generation," *Japanese Journal of Multiphase Flow*, vol. 25, no. 3, pp. 229–236, 2011.

[12] M. Diagne, M. David, P. Lauret, J. Boland, and N. Schmutz, "Review of solar irradiance forecasting methods and a proposition for small-scale insular grids," *Renewable and Sustainable Energy Reviews*, vol. 27, pp. 65–76, 2013.

[13] M. G. De Giorgi, M. Malvoni, and P. M. Congedo, "Comparison of strategies for multi-step ahead photovoltaic power forecasting models based on hybrid group method of data handling networks and least square support vector machine," *Energy*, vol. 107, no. 15, pp. 360–373, 2016.

[14] Y. Liu, S. Shimada, J. Yoshino, T. Kobayashi, K. Furuta, and Y. Miwa, "Ensemble forecasting of solar irradiance by applying a mesoscale meteorological model," *Solar Energy*, vol. 136, no. 15, pp. 597–605, 2016.

[15] W. Zhu and Y. H. Teng, "Study on the safety and mining influence of fully-mechanized caving mining with ultrathick seam under barrier lake," *Journal of Mining and Strata Control Engineering*, vol. 3, no. 1, Article ID 013525, 2021.

[16] W. C. Zhu, L. L. Niu, S. H. Li, and S. Li, "Creep-impact test of rock: status-of-the-art and prospect," *Journal of Mining and Strata Control Engineering*, vol. 1, no. 1, Article ID 013003, 2019.

[17] J. P. Zuo, M. L. Yu, S. Y. Hu et al., "Experimental investigation on fracture mode of different thick rock strata," *Journal of Mining and Strata Control Engineering*, vol. 1, no. 1, Article ID 013007, 2019.

[18] K. G. Jiang, L. Wang, S. S. Chi, T. Wei, and C. Jiang, "Boltzmann function prediction model coupled with the correction of inflection point offset and its parameter inversion method," *Journal of Mining and Strata Control Engineering*, vol. 3, no. 2, Article ID 023527, 2021.

[19] H. P. Kang, "Spatial scale analysis on coalmining and strata control technologies," *Journal of Mining and Strata Control Engineering*, vol. 22, 023538.

[20] Y. Li and G. Du, "Reasonable width of narrow coal pillars in roadway driving with gas drainage hole," *Journal of Mining and Strata Control Engineering*, vol. 2, no. 1, Article ID 013007, 2020.

[21] Z. J. Jiang, B. Sun, Y. Ping, and Z. D. Zhu, "Experimental on dynamic response characteristics of rock based on large diameter SHPB," *Journal of Mining and Strata Control Engineering*, vol. 3, no. 4, Article ID 043021, 2021.

[22] H. P. Kang, "Temporal scale analysis on coal mining and strata control technologies," *Journal of Mining and Strata Control Engineering*, vol. 3, no. 1, Article ID 013538, 2021.

[23] J. K. Li and H. Wang, "Ground support of interbedded rock roof in a deep roadway with fully-anchored cables," *Journal of*

Mining and Strata Control Engineering, vol. 2, no. 3, Article ID 033036, 2020.

[24] J. H. Li, H. J. Li, Z. C. Li, and Z. Q. Wu, "Research on river dike failure of short-distance coal seams mining under Hunchun river," *Journal of Mining and Strata Control Engineering*, vol. 2, no. 1, Article ID 013538, 2020.

[25] Y. X. Xia, C. Lu, G. Y. Yang et al., "Experimental study on axial fracture cutting and fracturing of abrasive jet in hard roof hole," *Journal of Mining and Strata Control Engineering*, vol. 2, no. 3, Article ID 033522, 2020.

[26] B. Xiao, Y. X. Wang, Z. X. Wang, and H. Yi, "Experimental on optimization of gangue particle size gradation in solid filling mining," *Journal of Mining and Strata Control Engineering*, vol. 4, no. 1, Article ID 013516, 2022.

[27] G. G. Xiao and W. B. Yao, "Law of pressure behavior of shallow buried coal seam in the gully terrain of madiliang coal mine," *Journal of Mining and Strata Control Engineering*, vol. 3, no. 4, Article ID 043023, 2021.

[28] J. L. Xu, D. Y. Xuan, W. B. Zhu, and X. Z. Wang, "Partial backfilling coal mining technology based on key strata control," *Journal of Mining and Strata Control Engineering*, vol. 1, no. 1, Article ID 013504, 2019.

[29] A. Rincón, O. Jorba, J. M. Baldasano, and L. Delle Monache, "Assessment of short-term irradiance forecasting based on post-processing tools applied on WRF meteorological simulations," in *Proceedings of the ES1002: Workshop March*, Barcelona, Spain, 2011.

[30] R. A. Verzijlbergh, P. W. Heijnen, S. R. De Roode, A. Los, and H. J. Jonker, "Improved model output statistics of numerical weather prediction based irradiance forecasts for solar power applications," *Solar Energy*, vol. 118, no. 16, pp. 634–645, 2015.

[31] E. Lorenz and D. Heinemann, "Prediction of solar irradiance and photovoltaic power," *Comprehensive Renewable Energy*, vol. 1, pp. 239–292, 2012.

[32] M. Pierro, F. Bucci, C. Cornaro et al., "Model output statistics cascade to improve day ahead solar irradiance forecast," *Solar Energy*, vol. 117, no. 14, pp. 99–113, 2015.

[33] S. Baran and A. Möller, "Bivariate ensemble model output statistics approach for joint forecasting of wind speed and temperature," *Meteorology and Atmospheric Physics*, vol. 129, no. 1, pp. 99–112, 2017.

[34] M. Homleid, "Diurnal corrections of short term surface temperature forecasts using the Kalman filter," *Weather and Forecasting*, vol. 10, no. 4, pp. 689–707, 1995.

[35] J. R. Porto de Carvalho, E. D. Assad, and H. S. Pinto, "Kalman filter and correction of the temperatures estimated by PRECIS model," *Atmospheric Research*, vol. 102, pp. 218–226, 2011.

[36] L. D. Monache, J. Wilczak, S. Mckeen et al., "A kalman-filter bias correction method applied to deterministic, ensemble averaged and probabilistic forecasts of surface ozone," *Tellus B: Chemical and Physical Meteorology*, vol. 60, no. 2, pp. 238–249, 2008.

[37] G. Galanis, P. Louka, P. Katsafados, I. Pytharoulis, and G. Kallos, "Applications of kalman filters based on non-linear functions to numerical weather predictions," *Annales Geophysicae*, vol. 24, no. 10, pp. 2451–2460, 2006.

[38] F. Cassola and M. Burlando, "Wind speed and wind energy forecast through kalman filtering of numerical weather prediction model output," *Applied Energy*, vol. 99, pp. 154–166, 2012.

[39] S. Pelland, G. Galanis, and G. Kallos, "Solar and photovoltaic forecasting through post-processing of the global environmental multiscale numerical weather prediction model," *Progress in Photovoltaics: Research and Applications*, vol. 21, no. 3, pp. 284–296, 2013.

[40] M. Diagne, M. David, J. Boland, N. Schmutz, and P. Lauret, "Post-processing of solar irradiance forecasts from WRF model at reunion Island," *Solar Energy*, vol. 105, no. 13, pp. 99–108, 2014.

[41] A. Rincón, O. Jorba, M. Frutos, L. Alvarez, F. P. Barrios, and J. A. González, "Bias correction of global irradiance modelled with weather and research forecasting model over Paraguay," *Solar Energy*, vol. 170, no. 8, pp. 201–211, 2018.

[42] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.

[43] W. C. Skamarock, J. B. Klemp, J. Dudhia et al., "A description of the advanced research WRF version 3," University Corporation for Atmospheric Research, Boulder, CO, USA, NCAR Technical Note NCAR/TN-475 + STR, 2008.

[44] D. Heinemann, E. Lorenz, and M. Girodo, *Solar Global Horizontal Irradiance Forecasting for the Management of Solar Energy systems. Energy and Semiconductor Research Laboratory, Energy Meteorology Group*, Oldenburg University, Oldenburg, Germany, 2006.

[45] Y. Che, X. Peng, L. Delle Monache, T. Kawaguchi, and F. Xiao, "A wind power forecasting system based on the weather research and forecasting model and kalman filtering over a wind-farm in Japan," *Journal of Renewable and Sustainable Energy*, vol. 8, no. 1, Article ID 013302, 2016.

[46] Y. P. Li, F. Cui, W. H. Yang, and C. X. Q. Wei, "Dynamic migration law and its control of roof in fully mechanized top coal caving mining in extremely steep and thick coal seams," *Journal of Mining and Strata Control Engineering*, vol. 2, no. 4, Article ID 043538, 2020.

[47] Y. X. Liu, M. S. Gao, H. S. Zhao, S. L. He, Z. G. Li, and Z. C. Zhang, "Detection of overlying rock structure and identification of key stratum by drilling and logging technology," *Journal of Mining and Strata Control Engineering*, vol. 2, no. 2, Article ID 023038, 2020.

[48] J. Y. Li, L. Wang, K. G. Jiang, and C. Q. Teng, "Parameter inversion method of probability integral model based on improved wolves algorithm," *Journal of Mining and Strata Control Engineering*, vol. 3, no. 1, Article ID 017038, 2021.

[49] S. Shimada, Y. Liu, H. Xia et al., "Accuracy of solar global horizontal irradiance simulation using the WRF-ARW model," *Solar Energy*, vol. 38, no. 5, pp. 41–48, 2012.

[50] G. L. Mellor and T. Yamada, "Development of a turbulence closure model for geophysical fluid problems," *Reviews of Geophysics*, vol. 20, no. 4, p. 851, 1982.

[51] Z. I. Janjic, "Comments on development and evaluation of a convection scheme for use in climate models," *Journal of the Atmospheric Sciences*, vol. 57, no. 21, p. 3686, 2000.

[52] J. Dudhia, S. Y. Hong, and K. S. Lim, "A new method for representing mixed-phase particle fall speeds in bulk microphysics parameterizations," *Journal of the Meteorological Society of Japan*, vol. 86, pp. 33–44, 2008.

[53] J. Dudhia, "Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model," *Journal of the Atmospheric Sciences*, vol. 46, no. 20, pp. 3077–3107, 1989.

[54] E. J. Mlawer, S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, "Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the long wave," *Journal of Geophysical Research: Atmospheres*, vol. 102, pp. 16663–16682, 1997.

[55] M. F. Cai, "Key theories and technologies for surrounding rock stability and ground control in deep mining," *Journal of Mining and Strata Control Engineering*, vol. 2, no. 3, Article ID 033037, 2020.

[56] Y. F. Cai, X. J. Li, W. N. Deng, W. Xiao, and W. K. Zhang, "Simulation of surface movement and deformation rules and detriment key parameters in high-strength mining," *Journal of Mining and Strata Control Engineering*, vol. 2, no. 4, Article ID 043511, 2020.

[57] D. F. Yang, Y. J. Zhang, S. Wang, C. H. Niu, and J. L. Chai, "Analysis of the influence of hidden fault dip angle on ground pressure behavior in shallow seam roof," *Journal of Mining and Strata Control Engineering*, vol. 2, no. 4, Article ID 043038, 2020.

[58] W. Z. Ma, X. M. Zhou, and S. Tan, "Study on failure characteristics of coal seam floor above confined water: a case study of shanxi yitang coal mine," *Journal of Mining and Strata Control Engineering*, vol. 2, no. 3, Article ID 033011, 2020.

[59] K. Fan, "Sudden deformation characteristic and cuttingroof support technology for double-used roadways in longtan Mine," *Journal of Mining and Strata Control Engineering*, vol. 2, no. 3, Article ID 033032, 2020.

[60] M. Kunitsugu, "Statistical guidance systems using Kalman filter technique," *Meteorological Society of Japan*, vol. 5, pp. 7–21, 1997.

[61] P. Kim and L. Huh, *Kalman Filter for Beginners with Matlab Examples*, CreateSpace Independent Publishing Platform, Scotts Valley, CA, USA, 2011.

[62] F. Q. Gao, "Use of numerical modeling for analyzing rock mechanic problems in underground coal mine practices," *Journal of Mining and Strata Control Engineering*, vol. 1, no. 1, Article ID 013004, 2019.

[63] Y. L. Bi, X. L. Li, and N. Guo, "Spatial heterogeneity of vegetation and soil nutrients induced by the open-pit mining in eastern grassland of China," *Journal of Mining and Strata Control Engineering*, vol. 2, no. 4, Article ID 047036, 2020.

[64] S. Fan, Y. Wang, S. Cao, B. Zhao, T. Sun, and P. Liu, "A deep residual neural network identification method for uneven dust accumulation on photovoltaic (PV) panels," *Energy*, vol. 239, Article ID 122302, 2022.

[65] S. Wang, K. Zhang, L. Chao et al., "Exploring the utility of radar and satellite-sensed precipitation and their dynamic bias correction for integrated prediction of flood and landslide hazards," *Journal of Hydrology*, vol. 603, Article ID 126964, 2021.

[66] F. Q. Gao, "Influence of hydraulic fracturing of strong roof on mining-induced stress-insight from numerical simulation," *Journal of Mining and Strata Control Engineering*, vol. 3, no. 2, Article ID 023032, 2021.

[67] B. L. Zhang, B. T. Shen, J. H. Zhang, and X. G. Zhang, "Experimental study of edge-opened cracks propagation in rocklike materials," *Journal of Mining and Strata Control Engineering*, vol. 2, no. 3, Article ID 033035, 2020.

[68] J. Chai, Y. Y. B. Ou, and D. D. Zhang, "Crackdetection method in similar material models based on DIC," *Journal of Mining and Strata Control Engineering*, vol. 2, no. 2, Article ID 023015, 2020.

[69] K. Zhang, S. Wang, H. Bao, and X. Zhao, "Characteristics and influencing factors of rainfall-induced landslide and debris flow hazards in Shaanxi Province, China," *Natural Hazards and Earth System Sciences*, vol. 19, no. 1, pp. 93–105, 2019.

[70] K. Zhang, A. Ali, A. Antonarakis et al., "The sensitivity of north american terrestrial carbon fluxes to spatial and temporal variation in soil moisture: an analysis using radar-derived estimates of root-zone soil moisture," *Journal of Geophysical Research. Biogeosciences*, vol. 124, no. 11, pp. 3208–3231, 2019.

[71] Y. Zhang, Y. He, H. Wang, L. Sun, and Y. Su, "Ultra-Broadband mode size converter using on-chip metamaterial-based luneburg lens," *ACS Photonics*, vol. 8, no. 1, pp. 202–208, 2021.

[72] Q. C. Zhao, B. J. Fu, and J. Yin, "Deformation mechanism and control technology of the surrounding rock of the floor roadway under the influence of mining," *Advances in Civil Engineering*, vol. 2020, Article ID 661303, 23115 pages, 2020.

[73] C. Jia and C. C. Hu, "Instability mechanism and control technology of longwall entries driving along the gob in a thick coal seam," *Journal of Mining and Strata Control Engineering*, vol. 2, no. 4, Article ID 043535, 2020.

[74] X. Fu and R. F. Wang, "Cooperative self-adaptive control model of fluid feeding system and hydraulic supports in working face," *Journal of Mining and Strata Control Engineering*, vol. 2, no. 3, Article ID 036031, 2020.

[75] J. Wang and X. L. Wang, "Seepage characteristic and fracture development of protected seam caused by mining protecting strata," *Journal of Mining and Strata Control Engineering*, vol. 3, no. 3, Article ID 033511, 2021.

[76] X. W. Zang, J. H. Xu, Z. B. Liu, and L. Sun, "Mechanical response and deformation law of weak interbedded mudstone under different loading paths," *Journal of Mining and Strata Control Engineering*, vol. 3, no. 4, Article ID 043022, 2021.

Hindawi
Advances in Meteorology
Volume 2022, Article ID 3140872, 12 pages
https://doi.org/10.1155/2022/3140872

Research Article

Statistical Learning-Based Spatial Downscaling Models for Precipitation Distribution

Yichen Wu,[1] Zhihua Zhang,[1,2] M. James C. Crabbe,[3,4,5] and Lipon Chandra Das[1,6]

[1]*Climate Modeling Laboratory, School of Mathematics, Shandong University, Jinan 250100, China*
[2]*MOE Key Laboratory of Environmental Change and Natural Disaster, Beijing Normal University, Bejing 100875, China*
[3]*Wolfson College, Oxford University, Oxford OX2 6UD, UK*
[4]*Institute of Biomedical and Environmental Science & Technology, University of Bedfordshire, Luton LU1 3JU, UK*
[5]*School of Life Sciences, Shanxi University, Taiyuan 030006, China*
[6]*University of Chittagong, Chittagong 4331, Bangladesh*

Correspondence should be addressed to Zhihua Zhang; zhangzhihua@sdu.edu.cn

Received 7 April 2022; Revised 12 May 2022; Accepted 20 May 2022; Published 7 June 2022

Academic Editor: Upaka Rathnayake

The downscaling technique produces high spatial resolution precipitation distribution in order to analyze impacts of climate change in data-scarce regions or local scales. In this study, based on three statistical learning algorithms, such as support vector machine (SVM), random forest regression (RF), and gradient boosting regressor (GBR), we proposed an efficient downscaling approach to produce high spatial resolution precipitation. In order to demonstrate efficiency and accuracy of our models over traditional multilinear regression (MLR) downscaling models, we did a downscaling analysis for daily observed precipitation data from 34 monitoring sites in Bangladesh. Validation revealed that R^2 of GBR could reach 0.98, compared with RF (0.94), SVM (0.88), and multilinear regression (MLR) (0.69) models, so the GBR-based downscaling model had the best performance among all four downscaling models. We suggest that the GBR-based downscaling models should be used to replace traditional MLR downscaling models to produce a more accurate map of high-resolution precipitation for flood disaster management, drought forecasting, and long-term planning of land and water resources.

1. Introduction

Global warming is significantly influencing the environment, hydrology, and ecosystem. Continued warming in the 21st century will significantly impact precipitation and monsoons and lead to the intensification of extreme rainstorm and drought events [1–3]. South Asia is a well-known summer monsoon region. The formation of the South Asian monsoon is mainly caused by the seasonal movement of the pressure belt and wind belt, as well as the influence of thermal differences between land and ocean as well as topographic factors. About 80% of precipitation in South Asia are closely linked with monsoons [4–6]. More than one billion people rely on monsoonal rainfall for agricultural production, hydroelectric generation, and other basic needs [7]. Especially, Bangladesh is located in one of the largest deltas in the world with a dense network of main rivers and their tributaries, resulting Bangladesh being a flood prone country. Due to the reliance on rain-fed agriculture, Bangladesh is extremely sensitive and high vulnerability to climate change. Since Bangladesh has only few and sparse precipitation monitoring stations, it is very important to generate high spatial resolution precipitation data to mitigate climate change impacts. However, only very limited downscaling research in Bangladesh was carried out by now: observed precipitation data in Bangladesh were downscaled by using multilinear regression as the core part of the downscaling algorithm [8, 9]. Since the nonlinear relation between the large and small-scale dynamics in these research studies was ignored, the obtained downscaling accuracy is unstable. Simulated precipitation data from ensemble climate models in Coupled Model

Intercomparison Project Phase 5 (CMIP5) were downscaled by using the method of model output statistics [10, 11], but this method can only be applied for simulated climate data.

Generally, for any country with few and sparse precipitation monitoring stations, downscaling is the key technique to generate high spatial resolution precipitation data. Downscaling can be divided into dynamical downscaling and statistical downscaling. Dynamic downscaling mainly depends on physical principles governing the climate system and high-resolution regional climate models, while statistical downscaling is based on statistical relation between local variables and large-scale variables [11]. Compared with dynamic downscaling relying on some local scale models or regional climate models, statistical downscaling uses a multilinear regression model to establish the correlation between local variables and large-scale variables. Since Earth's climate is a complex, the multidimensional multiscale system with different physical processes acting on different temporal and spatial scales, statistical downscaling cannot reveal complex nonlinear relationships between local variables and large-scale variables [12,13].

Compared with traditional statistical techniques, advanced statistical learning techniques have showed excellent performance on solving problems with complex nonlinear correlations between variables [14]. Statistical learning techniques can map the predictor(s) only rely on the existing relationship between the two rather than the explicit function [15]. Main statistical learning techniques include the following. (a) Support vector machine (SVM) uses a kernel function to map features to a high-dimensional space for classification and regression; the main advantage lies in that SVM can effectively solve small-sample, nonlinear and high-dimensional regression problems. (b) Random forest (RF) is an ensemble learning method based on bagging, which can handle classification and regression problems well. (c) Gradient boosting regressor (GBR) is an ensemble learning model based on boosting, which reduces the loss by fitting the residuals to obtain high prediction accuracy. Compared with other statistical learning techniques (e.g., neural networks), the SVM requires only small amount of samples and RF and GBR can avoid over fitting [13], so in this study, based on SVM, RF, and GBR, we propose a new downscaling approach to produce a finer spatial resolution precipitation map. In order to demonstrate efficiency and accuracy of our models over traditional multilinear regression (MLR) downscaling models, we use a downscaling analysis for daily observed precipitation data from 34 monitoring sites in Bangladesh. Moreover, based on obtained high spatial resolution precipitation distribution, we analyzed patterns and trend of Bangladesh's precipitation from 1989 to 2018.

2. Downscaling Methods

Based on three statistical learning algorithms, such as support vector machine (SVM), random forest regression (RF), and gradient boosting regressor (GBR), we proposed an efficient downscaling approach to produce high spatial resolution precipitation, especially for any country with few and sparse precipitation monitoring stations.

2.1. Three Known Statistical Learning Algorithms. Support vector machine (SVM) can map the complex data features into a high-dimensional space by using nonlinear mapping algorithms and separate data using optimal linear hyperplane [16–18]. For given n training data (x_1, y_1), $(x_2, y_2), \cdots, (x_n, y_n)$, the SVM is to find a regression function $f(x) = \omega, \Phi(x) + b$, such that $f(x_i)$ has at most ε deviation from the actual value y_i, where Φ is a kernel function mapping the input data to a high-dimensional space, and the parameters ω and b are the weight term and bias term, respectively. The basic algorithm to search $f(x)$ is to minimize the regression risk by the following formulas:

$$R_{reg}(f) = \frac{1}{2}\omega^2 + C\sum_{i=1}^{n}\Gamma\left(f(x_i) - y_i\right), \tag{1}$$

$$\text{subject to}\left|y_i - \omega, \Phi(x_i) - b\right| \le \varepsilon \quad i = 1, 2, \ldots, n, \tag{2}$$

where $\Gamma(\cdot)$ is a cost function, and the parameter C can balance the prediction error and model complexity to avoid the overfitting of training data.

Random forest (RF) uses the bagging (or bootstrap aggregation) technique and decorrelation technique to combine a series of small-scale decision trees into a single procedure for better regression prediction [19]. RF can overcome the disadvantage of single decision tree in overfitting to training data and can handle data with few missing values. By using one in a randomly chosen subset of m predictors from a total of n predictors, a new node in a decision tree of RF can be generated, where the bootstrap resampling technique is used to randomly select k samples from N original training samples as its training set, and the remaining N-k samples (i.e., out-of-bag samples) are used for cross validation. Each decision tree is only trained by m predictors and k training samples, and different decision trees are generated by different predictors and training samples which are randomly chosen. In order to reduce the variance of prediction results by decision trees, the optimal prediction by RF is the average of the predictions from all decision trees (i.e., so-called the aggregate procedure). The prediction accuracy and computing efficiency of RF models are mainly affected by the number of decision trees and the number of predictors/training samples in each decision tree [20].

Gradient boosting regressor (GBR) is an ensemble regression tree model which starts from a simple regression tree and adds a new regression tree again and again [21]. The GBR is a weighted sum of regression trees:

$$F(x) = \sum_{m=1}^{M}\gamma_m h_m(x), \tag{3}$$

where $h_m(x)$ is a mth regression tree for boosting predication accuracy. The core procedure in GBR is to continuously

reduce the loss by searching optimal parameters in the new regression tree to fit the negative gradient of the residual error of existing ensemble regression tree model. In detail.the F(x) in GBR can be estimated through an iterative procedure by using the following formula.

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x). \quad (4)$$

During each iteration, a new regression tree $h_m(x)$ is constructed to minimize the residual error by using a gradient descent method. The output of GBR can achieve better generalization performance than a single regression tree [22]. The idea behind GBF is very different from RF. The RF is to build all regression trees in parallel and the output of RF is the average of prediction results from all decision trees, while GBR is to build regression trees in a form of sequence and the output of GBR is the sum of prediction results from all regression trees.

2.2. Statistical Learning-Based Downscaling Technique. The widely used statistical downscaling techniques are usually based on traditional multiple linear regression (MLR), which cannot effectively deal with the instability of downscaling time series and the existence of collinearity between downscaling factors and makes the improvement of downscaling performance significantly limited. In this study, based on GBR, RF, and SVM, we propose an efficient downscaling method to produce high spatial resolution precipitation, where daily station-level precipitation data and longitude/latitude/altitude are used as the input of GBR/SVM/RF models. The output is the downscaled precipitation product. Our downscaling models can largely make up for the deficiencies of the MLR downscaling approach.

For the validation of our downscaling method, noticing that available observed precipitation data are small scale, and in order to avoid overfitting and use as much data as possible in model training, we utilized the 5-fold cross validation method [23]. The main model training process was to divide all data into five subsets; each time one subset was used for the test set and the remaining four subsets were used for training set, and finally, the average of five training errors is used as the result. The correlation of determination (R^2), mean absolute error (MAE), and root mean square error (RMSE) are used to assess the performance of different downscaling models. To demonstrate accuracy and efficiency of our models with traditional MLR downscaling models, we used a downscaling analysis for daily observed precipitation data from Bangladesh.

3. Study Area and Data

Bangladesh is located on deltas of large rivers flowing from the Himalayas, leading to that its topography is extremely flat (Figure 1). Traditionally, it is divided into seven regions (Figure 2). High humidity, warm temperature, and wide seasonal variability in precipitation are the main climate characteristics of Bangladesh. This climate is mainly caused by geographic location, north-south continental atmospheric pressure gradient, and fluctuation in terrestrial and sea surface temperature [24]. Due to significant high precipitation in monsoon seasons and flat and low delta plain

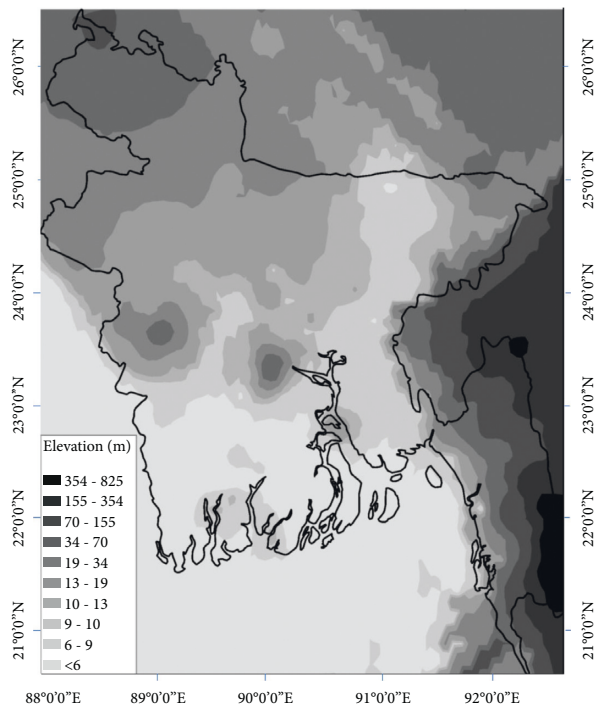


Figure 1: Elevation map of Bangladesh.

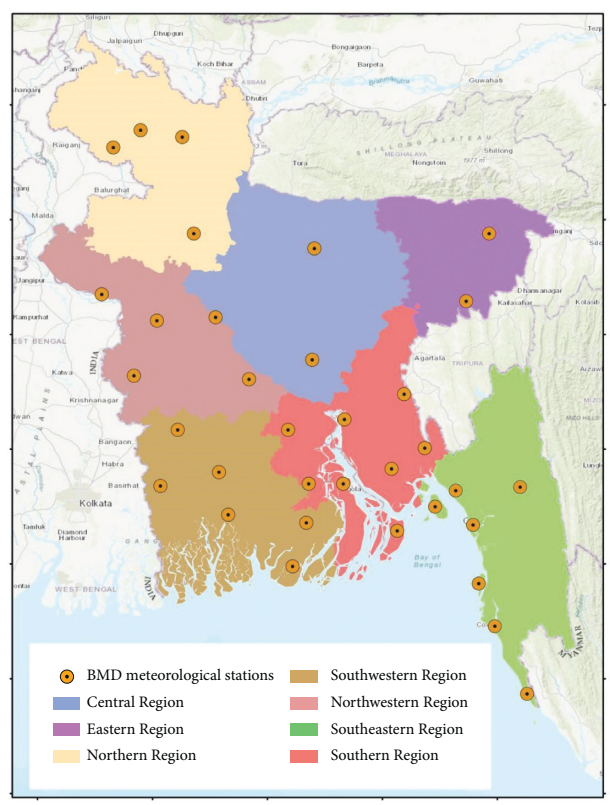


Figure 2: Seven regions of Bangladesh.

with a dense river network (Figure 1), floods and related disasters take place frequently [25]. Due to an agriculture-based economy, the high spatial resolution precipitation map can play a key role in Bangladesh's flood control, drought resistance, and water resource management. Since there are few and sparse precipitation monitoring stations in Bangladesh, it is necessary to conduct a downscaling analysis for observed precipitation data in Bangladesh. To achieve this aim, the daily precipitation data in Bangladesh were obtained from 34 monitoring stations (Figure 3) of the Bangladesh Meteorological Department, and the longitude, latitude, and elevation data of Bangladesh were extracted from Google Earth [26]. Based on statistical learning-based downscaling models in Section 2.2, we can produce high spatial resolution precipitation in Bangladesh.

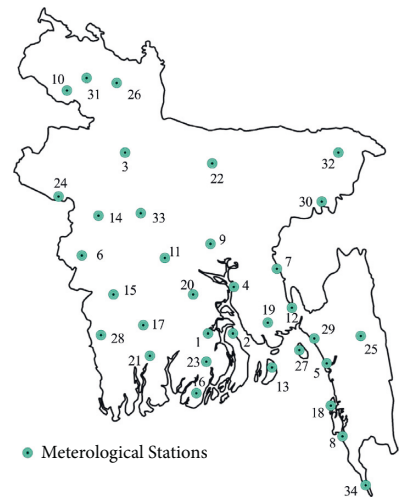
4. Results and Discussion

4.1. Optimal Statistical Learning-Based Spatial Downscaling Models. Based on daily precipitation data during 1989–2018 and longitude/latitude/elevation data in Bangladesh, we used our statistical learning-based downscaling models to produce higher spatial resolution precipitation data. Table 1 provides the validation results of our models and a traditional MLR downscaling model during 5-fold cross-validation processing. Our downscaling models demonstrate good performance over traditional MLR downscaling models. In terms of R^2 value, the downscaled data using GBR and RF showed good consistency with the original observation data. In validation analysis, the GBR downscaling model produced the highest R^2 (0.98) and the lowest RMSE (9.63) and MAE (7.24). Figure 4 shows the correlation between the downscaled products and the observed precipitation. The GBR downscaling model yielded the highest performance followed by RF, and the SVM downscaling model ranked the last.

In terms of spatial distribution, our downscaling models were better than the traditional MLR model (Figure 5). The spatial distribution maps of downscaled precipitation produced by GBR and RF are in high agreement with observations. The downscaling precipitation produced by SVM revealed only coarse spatial distribution characteristics: the precipitation gradually increased from western to central regions.

In summary, by using our downscaling model (GBR, RF, and SVM), to simulate the relationship between terrain variables and observed precipitation data in Bangladesh, it is clear that the GBR downscaling model performed best, compared with the RF model, the SVM model, and the traditional MLR model.

4.2. Spatial Variation Analysis of Downscaled Precipitation over Bangladesh. In order to analyze the seasonal variation of precipitation in Bangladesh, we used our GBR downscaling model to produce mean seasonal precipitation distribution during 1989–2018 (Figure 5). Bangladesh has significantly high precipitation during the monsoon season and low precipitation during the remaining three seasons



Meteorological Stations

1 Barisal	18 Kutubdia
2 Bhola	19 M.court
3 Bogra	20 Madaripur
4 Chandpur	21 Mongla
5 Chittagong	22 Maymensingh
6 Chuadanga	23 Patuakhali
7 Comilla	24 Rajshahi
8 Cox's Bazar	25 Rangamati
9 Dhaka	26 Rangpur
10 Dinajpur	27 Sandwip
11 Faridpur	28 Satkhira
12 Feni	29 Sitakunda
13 Hatiya	30 Srimangal
14 Ishurdi	31 Sydpur
15 Jessore	32 Sylhet
16 Khepupara	33 Tangail
17 Khulna	34 Teknaf

Figure 3: Location of 34 monitoring stations.

(Figure 6). In the winter season, the precipitation is significantly lower and is close to uniform spatial distribution; in the premonsoon season, the highest precipitation occurs in the middle region; in the monsoon season, higher precipitation occurs in the southwestern and southeastern regions; in the postmonsoon season, the precipitation distribution is particularly uneven and has high spatial variability. Relative dry conditions will occur in the northwestern and central regions.

Using downscaled precipitation by our GBR downscaling model, we demonstrated a difference between the seven regions of Bangladesh (Figure 7). The eastern region showed the highest fluctuation, followed by the southeastern region. The F-statistic value exceeds the critical point in analysis of variance (ANOVA) showing that these regional differences are statistically significant.

Based on the Mann–Kendall trend test and Sen's slope test (Table 2), eastern, southwestern, southern, and southeastern regions showed upward trends during 1989–2018, but these trends were not significant. The remaining three regions showed downward trends, where only one region showed a statistical significance. Among all seven regions, the northern region showed the highest downward trend with −13.38 mm/year, while southeastern region shows the highest upward trend with 4.24 mm/year.

Table 1: Validation results of downscaled precipitation data.

Regression models	R^2	RMSE	MAE
MLR	0.69	440.5	353.8
SVM	0.88	275.5	132.4
RF	0.94	197.7	134.24
GBR	0.98	9.63	7.24

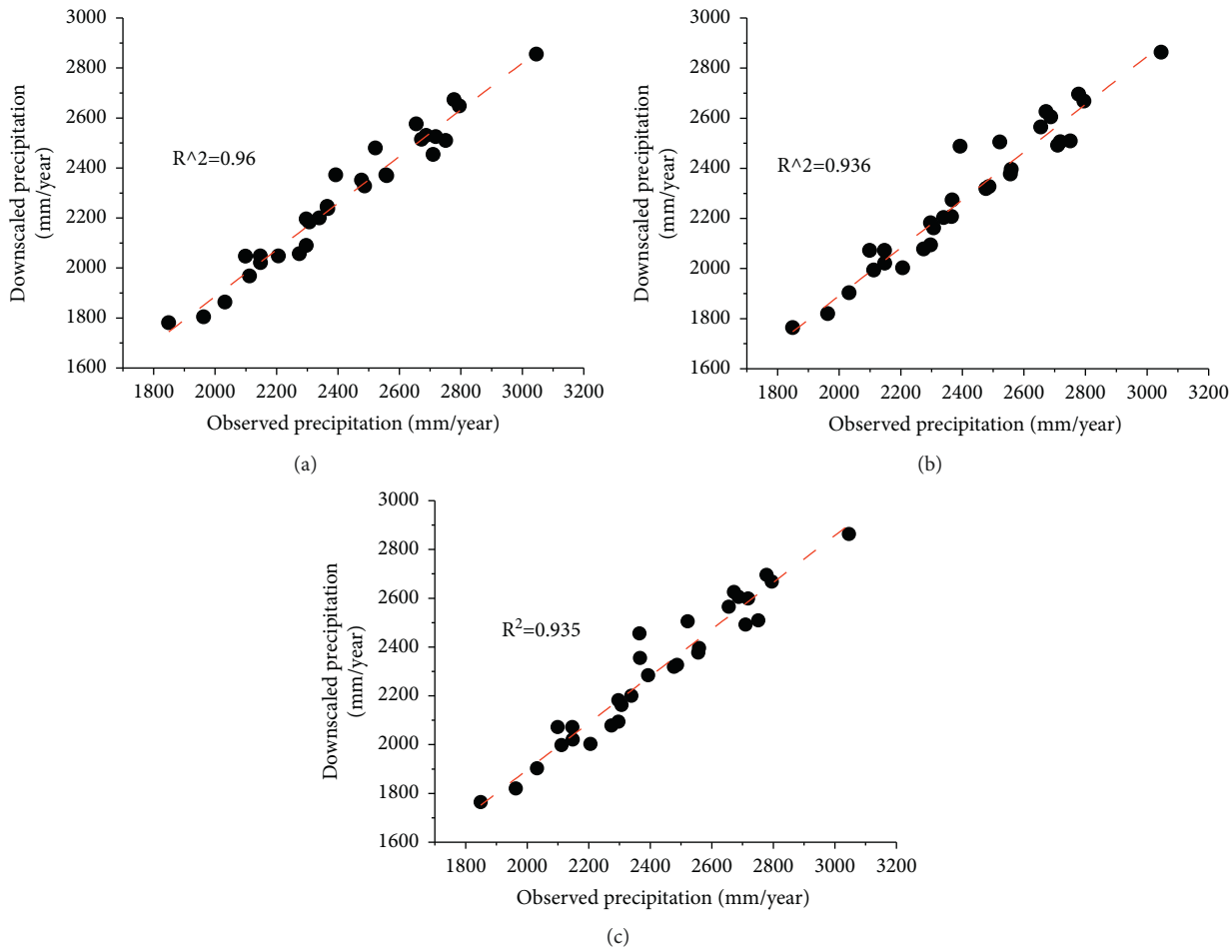


Figure 4: The correlation between downscaled precipitation by (a) GBR, (b) RF, (c) SVM and observed precipitation, respectively.

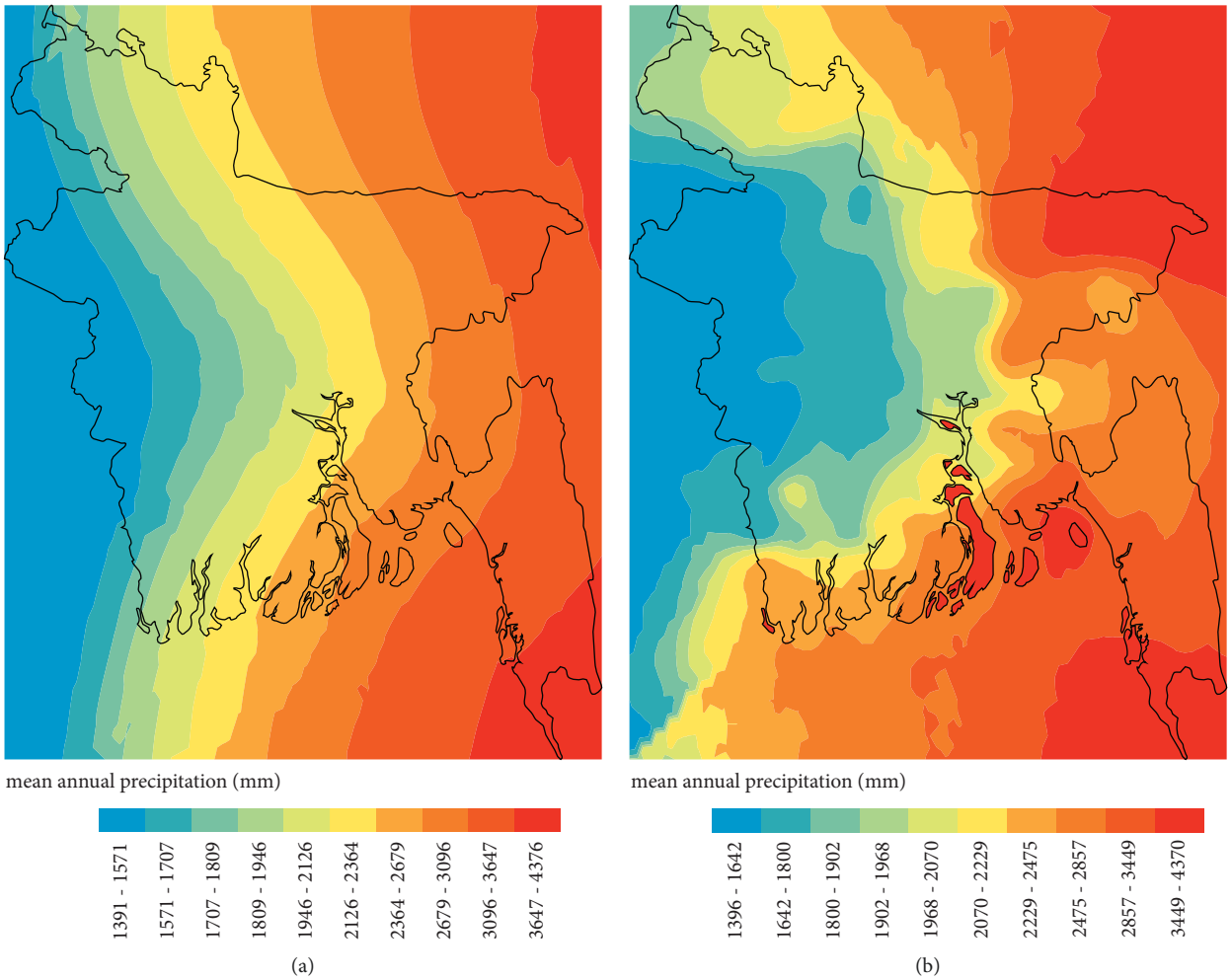


Figure 5: Continued.

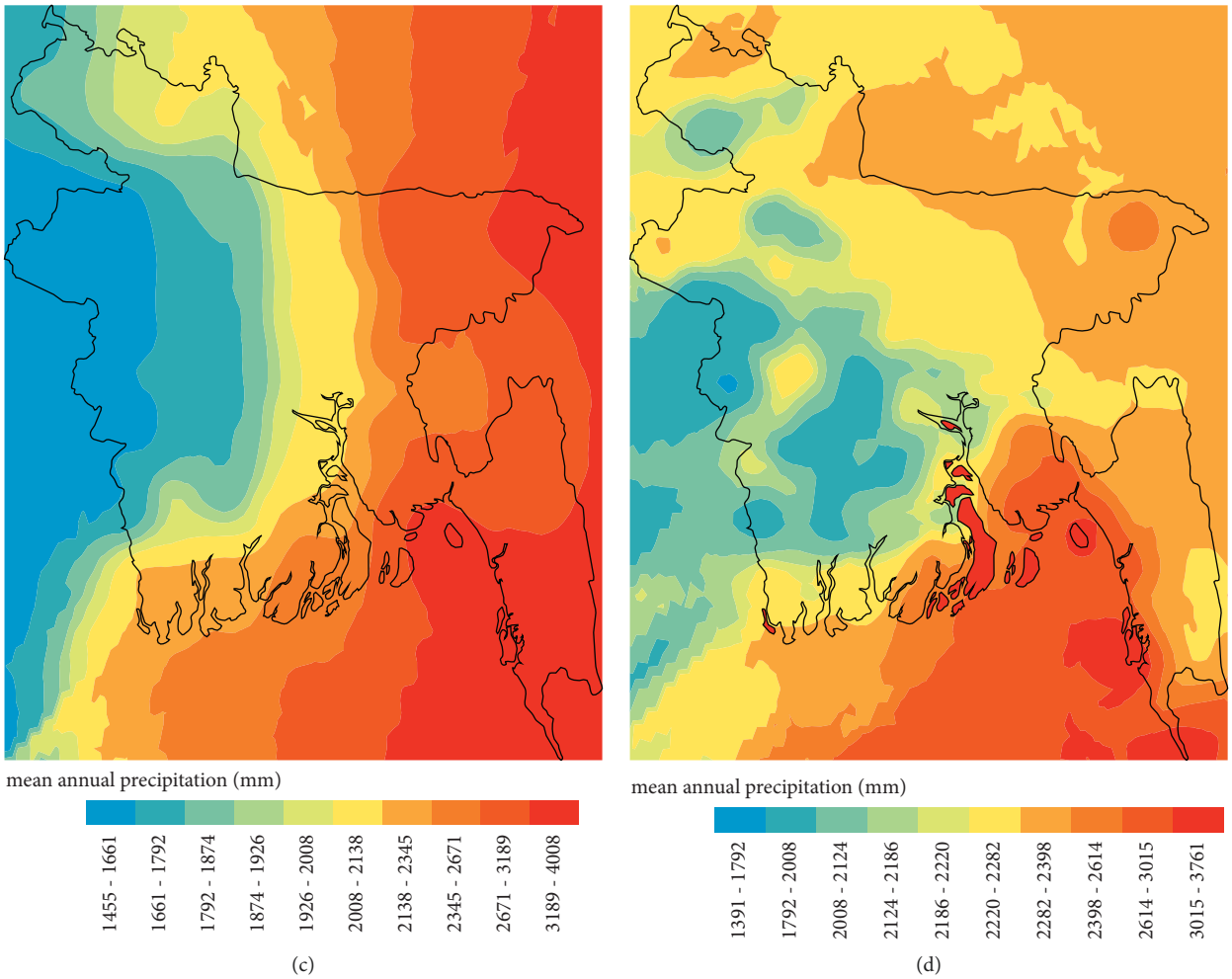


Figure 5: Continued.

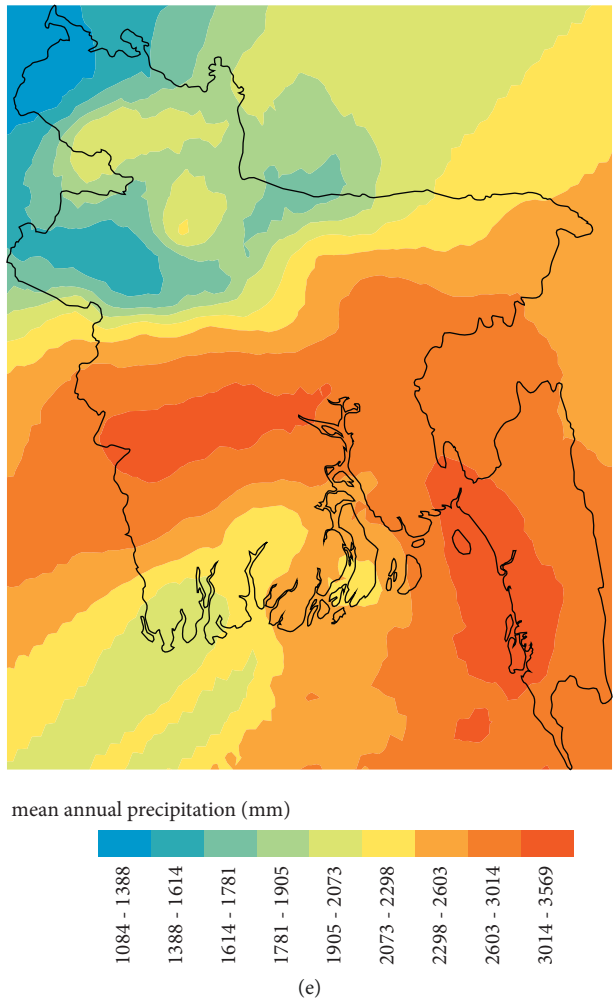
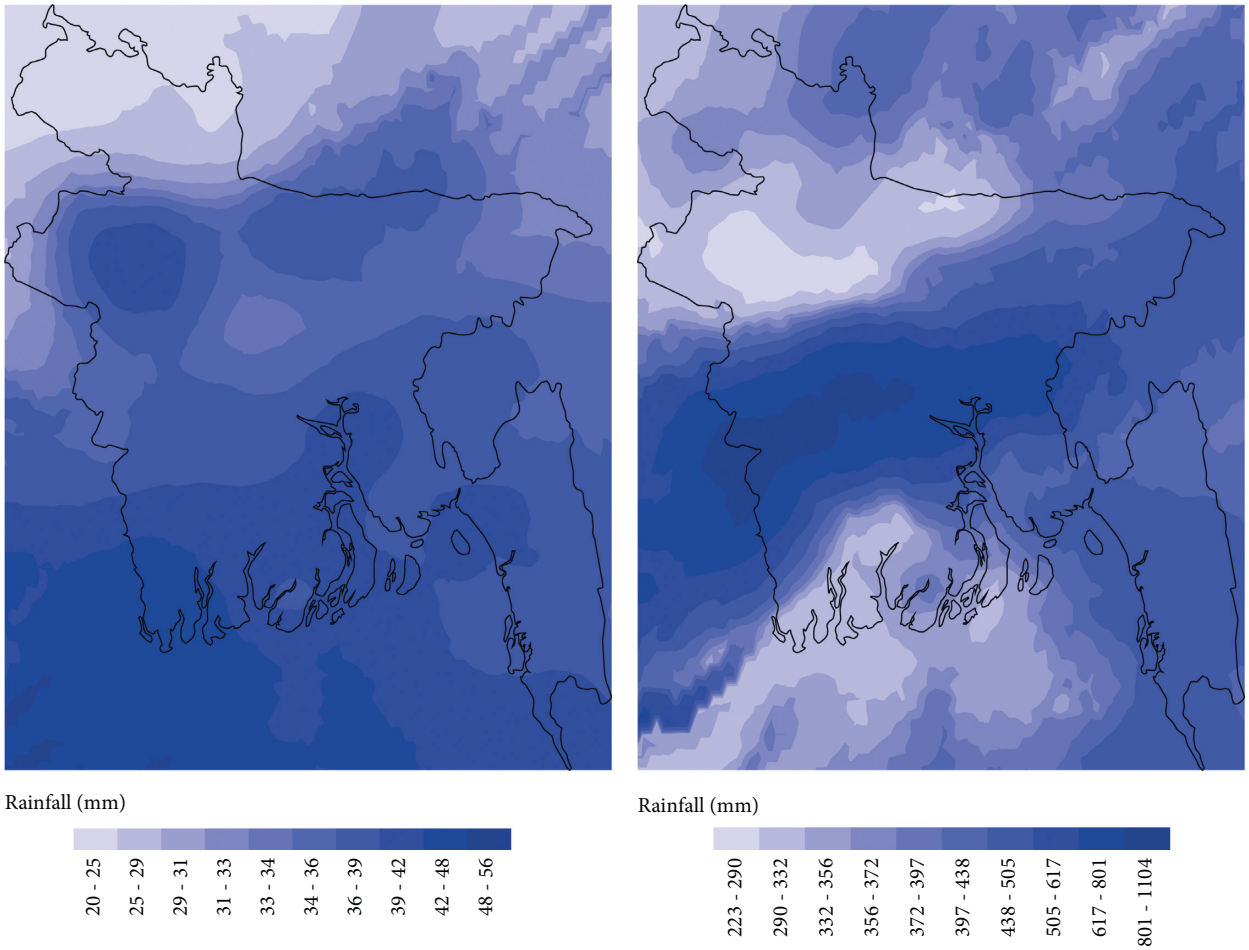


Figure 5: Spatial distribution of mean annual precipitation by (a) observation, (b) GBR, (c) RF, (d) SVM, and (e) MLR in Bangladesh from 1989 to 2018.



(a) (b)

Figure 6: Continued.

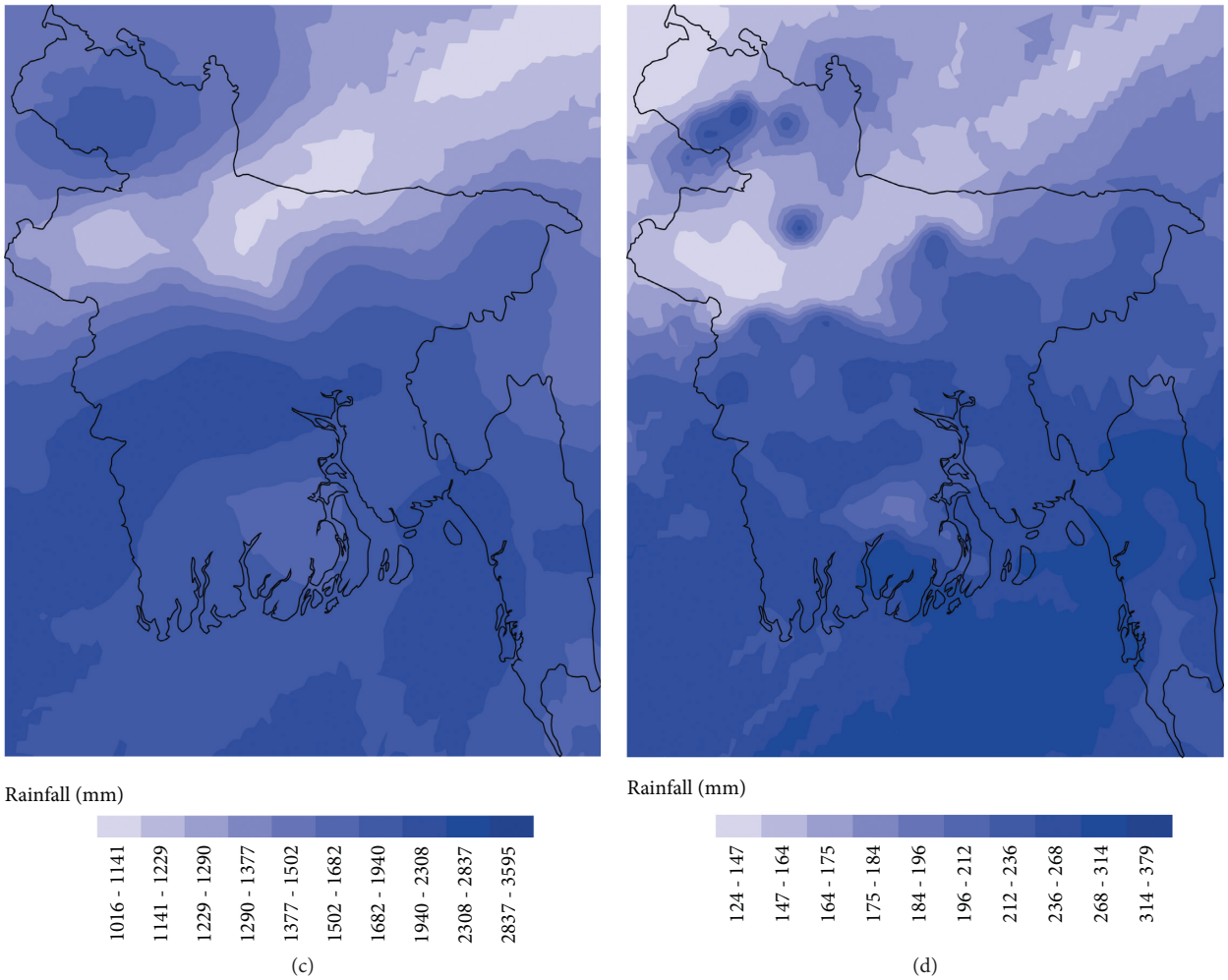


Figure 6: Mean seasonal precipitation distribution of (a) winter, (b) premonsoon, (c) monsoon, and (d) postmonsoon in Bangladesh during 1989–2018.

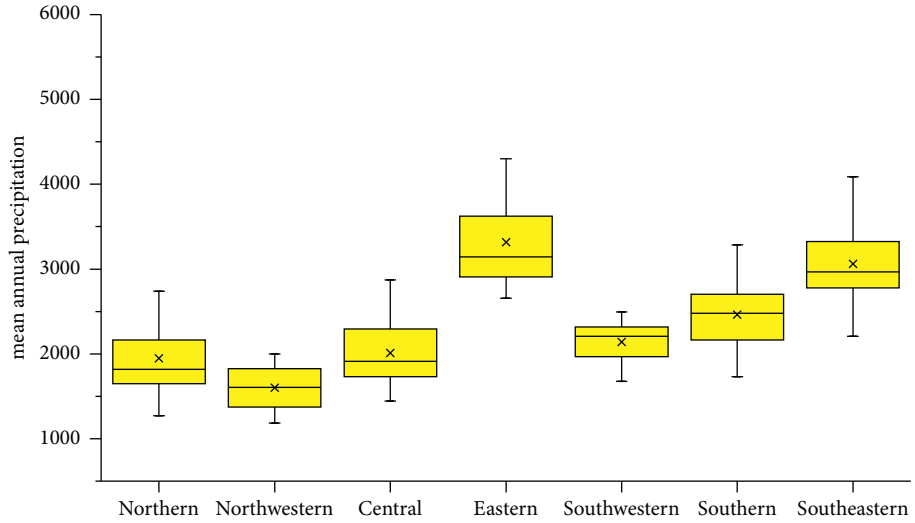


Figure 7: Regional difference in precipitation in Bangladesh.

Table 2: Trend analysis in difference regions of Bangladesh.

Region	Northern	Northwestern	Central	Eastern	Southwestern	Southern	Southeastern
Z_S	−1.78*	−1.05	−1.39	0.16	0.21	0.29	0.44
Q (mm/year)	−13.38	−9.99	−11.77	1.66	2.36	3.86	4.24

90% significance level*.

5. Conclusions

For an agriculture-based country like Bangladesh, water resource contributes the most to agricultural planning. Precipitation plays a more important role on agricultural development than other climatic and environmental variables. It can influence flood disaster management, drought resistance, long-term planning of land and water resources, and different kinds of infrastructure. Therefore, to produce high spatial resolution precipitation data is crucial in analyzing climate change impacts, especially for countries with few and sparse precipitation monitoring stations. Downscaling is an effective technique to solve this issue. The widely used statistical downscaling techniques are usually based on traditional MLR, which cannot effectively deal with the instability of downscaling time series and the existence of collinearity between downscaling factors, and make the improvement of downscaling performance significantly limited. In this study, based on GBR, RF, and SVM, we propose an efficient downscaling approach to produce high spatial resolution precipitation from daily station-level precipitation data and longitude/latitude/altitude data. In order to demonstrate the efficiency and accuracy of our models with traditional MLR downscaling models, we did a downscaling analysis for daily observed precipitation data from 34 monitoring sites in Bangladesh. Our downscaling models have clear advantages over traditional multilinear regression (MLR) downscaling models. The GBR-based downscaling model had the best performance in all four downscaling models. Therefore, we suggest that the GBR-based downscaling models should be used to replace traditional MLR downscaling models to produce a more accurate map of high-resolution precipitation for mitigating impacts of climate disasters, especially South Asian countries with few and sparse precipitation monitoring stations.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Disclosure

Yichen Wu and Zhihua Zhang are the co-first authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported by the European Commission's Horizon 2020 Framework Program (861584) and Taishan distinguished professorship fund.

References

[1] IPCC, *Climate Change 2021: The Physical Science Basis*, IPCC, Geneva, Switzerland, 2021.

[2] M. Latifur, N. Janet, S. A. Mansor et al., "Remote sensing an integrated method for identifying present status and risk of drought in Bangladesh," *Remote Sensing*, vol. 2020, no. 12, p. 2686, 2020.

[3] S. Shahid and H. Behrawan, "Drought risk assessment in the western part of Bangladesh," *Natural Hazards*, vol. 46, no. 3, pp. 391–413, 2008.

[4] V. H. Jamshadali, M. J. K. Reji, H. Varikoden, and R. Vishnu, "Spatial variability of south Asian summer monsoon extreme rainfall events and their association with global climate indices," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 221, Article ID 105708, 2021.

[5] S. S. C. Shenoi, "Differences in heat budgets of the near-surface Arabian sea and bay of bengal: implications for the summer monsoon," *Journal of Geophysical Research*, vol. 107, no. C6, pp. 3052–3061, 2002.

[6] A. G. Turner and H. Annamalai, "Climate change and the South Asian summer monsoon," *Nature Climate Change*, vol. 2, no. 8, pp. 587–595, 2012.

[7] M. R. Rahman and H. Lateh, "Climate change in Bangladesh: a spatio-temporal analysis and simulation of recent temperature and rainfall data using GIS and time series analysis model," *Theoretical and Applied Climatology*, vol. 128, no. 1-2, pp. 27–41, 2017.

[8] M. Alamgir, S. Hadi Pour, M. Mohsenipour, M. Mehedi Hasan, and T. Ismail, "Predictors and their domain for statistical downscaling of climate in Bangladesh," *Jurnal Teknologi*, vol. 78, no. 6-12, 2016.

[9] S. A. Shahriar, M. A. M. Siddique, and S. M. A. Rahman, "Climate change projection using statistical downscaling model over Chittagong Division, Bangladesh," *Meteorology and Atmospheric Physics*, vol. 133, no. 4, pp. 1409–1427, 2021.

[10] J. Fang, J. Du, W. Xu, P. Shi, M. Li, and X. Ming, "Spatial downscaling of TRMM precipitation data based on the orographical effect and meteorological conditions in a mountainous area," *Advances in Water Resources*, vol. 61, pp. 42–50, 2013.

[11] S. H. Pour, S. Shahid, E.-S. Chung, and X. J. Wang, "Model output statistics downscaling using support vector machine for the projection of spatial and temporal changes in rainfall of Bangladesh," *Atmospheric Research*, vol. 213, pp. 149–162, 2018.

[12] Z. Zhang and J. Li, *(Monograph) Big Data Mining for Climate Change*, Elsevier, Amsterdam, Netherlands, 2020.

[13] H. Zhang, P. Wu, A. Yin, X. Yang, M. Zhang, and C. Gao, "Prediction of soil organic carbon in an intensively managed reclamation zone of eastern China: a comparison of multiple linear regressions and the random forest model," *The Science of the Total Environment*, vol. 592, pp. 704–713, 2017.

[14] W. Jing, Y. Yang, X. Yue, and X. Zhao, "A comparison of different regression algorithms for downscaling monthly satellite-based precipitation over north China," *Remote Sensing*, vol. 8, no. 10, pp. 835–917, 2016.

[15] Y. Mei, V. Maggioni, P. R. Houser, Y. Xue, and T. Rouf, "A nonparametric statistical technique for spatial downscaling of precipitation over high mountain Asia," *Water Resources Research*, vol. 56, no. 11, Article ID e2020WR027472, 2020.

[16] V. Vapnik, S. E. Golowich, and A. Smola, "Support vector method for function approximation, regression estimation, and signal processing," *Advances in Neural Information Processing Systems*, vol. 9, pp. 281–287, 2008.

[17] A. G. Abo-Khalil and D. C. Lee, "MPPT control of wind generation systems based on estimated wind speed using SVR," *IEEE Transactions On Industrial Electronics*, vol. 55, no. 3, pp. 1489-1490, 2008.

[18] P. Tsangaratos, I. Ilia, and I. Matiatos, "Spatial analysis of extreme rainfall values based on support vector Machines optimized by genetic algorithms," in *Spatial Modeling in GIS and R for Earth and Environmental Sciences*, H. R. Pourghasemi and C. Gokceoglu, Eds., Elsevier, Amsterdam, Netherlands, pp. 1–19, 2019.

[19] C. d. S. Chagas, W. de Carvalho Junior, S. B. Bhering, and B. Calderano Filho, "Spatial prediction of soil surface texture in a semiarid region using random forest and multiple linear regressions," *Catena*, vol. 139, pp. 232–240, 2016.

[20] P. T. Guo, M. F. Li, W. Luo, Q. F. Tang, Z. W. Liu, and Z. M. Lin, "Digital mapping of soil organic matter for rubber plantation at regional scale: an application of random forest plus residuals kriging approach," *Geoderma*, vol. 237-238, pp. 49–59, 2015.

[21] N. Bagalkot, A. Keprate, and R. Orderløkken, "Combining computational fluid dynamics and gradient boosting regressor for predicting force distribution on horizontal axis wind turbine," *Vibrations*, vol. 4, no. 1, pp. 248–262, 2021.

[22] A. Elnashar, H. Zeng, B. Wu et al., "Downscaling TRMM monthly precipitation using google Earth engine and google cloud computing," *Remote Sensing*, vol. 12, no. 23, p. 3860, 2020.

[23] S. Yadav and S. Shukla, "Analysis of K-fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality classification," in *Proceedings of the IEEE 6th International Conference on Advanced Computing (IACC)*, Bhimavaram, India, 2016.

[24] A. Neelim and T. Islam, *Climate Change in Bangladesh: a closer look to temperature and rainfall data*, 2010.

[25] M. A. A. Hoque, B. Pradhan, and N. Ahmed, "Assessing drought vulnerability using geospatial techniques in northwestern part of Bangladesh," *The Science of the Total Environment*, vol. 705, Article ID 135957, 2020.

[26] http://earth.google.com/.

Hindawi
Advances in Meteorology
Volume 2022, Article ID 7236527, 8 pages
https://doi.org/10.1155/2022/7236527

Research Article

The Prediction Algorithm and Characteristics Analysis of Kuroshio Sea Surface Temperature Anomalies

Dawei Shi,[1,2] Chao Li,[2,3] Zhu Zhu,[4] Runqing Lv,[3] Shengjie Chen,[3] and Yunfeng Zhu[1]

[1]*Lianyungang Meteorological Bureau, Lianyungang 222006, China*
[2]*Key Laboratory of Traffic Meteorology, China Meteorological Administration, Nanjing 210008, China*
[3]*Jiangsu Meteorological Observatory, Nanjing 210008, China*
[4]*Anhui Meteorological Observatory, Hefei 230000, China*

Correspondence should be addressed to Chao Li; qwerlc@163.com

Received 31 December 2021; Revised 8 April 2022; Accepted 11 April 2022; Published 27 April 2022

Academic Editor: Upaka Rathnayake

Based on 130 climate signal indexes provided by National Climate Center of China, this paper established a decision tree diagnostic prediction model for Spring Kuroshio Sea Surface Temperature (SST) from 1961 to 2015 (65 years) by using Chi-Squared Automatic Interaction Detector (CHAID) algorithm in data mining and obtained five rule sets to determine whether Spring Kuroshio SST is high or not. Considering the data of the 44 years from 1961 to 2004 as the training set of the model and the other years as the test set, the training accuracy of the model can reach to 95.45% and the test accuracy can reach to 81.82%. Three types of Spring Kuroshio SST are different in intensity and distribution. The results show that the prediction model of Spring Kuroshio SST based on CHAID algorithm has a high prediction accuracy, with the reasonable and effective model and the well-thought-out decision rules. Moreover, based on the results of decision classification, the SST anomalies correspond to different distribution characteristics of summer daily precipitation anomalies in eastern China, which can provide a new idea and method for climate prediction of regional summer precipitation.

1. Introduction

Kuroshio is famous for its high Sea Surface Temperature (SST), high salinity, fast current, and large flow. Geographically, the Kuroshio is composed of the Kuroshio in the source area, the Kuroshio in the East China Sea, and the Kuroshio in the south of Japan. The Kuroshio in the source area is located to the east of Luzon Island and Taiwan Island and to the west of 130°E [1]. Kuroshio is the main current communicating between the Pacific Ocean and the East China Sea, the South China Sea, which plays a very important role in the thermohaline current transport, atmospheric circulation, and relevant air sea interaction in the China Sea, and its seasonal and interannual characteristics are also closely related to the climate in China [2, 3]. Therefore, the research of Kuroshio has been one of the important international and national or regional research programs. In the World Climate Research Program, the World Ocean Circulation Experiment (WOCE) studied the characteristics of Kuroshio front and front vortex and analyzed the variation characteristics of Kuroshio path and great bend [4]. The Climate Variability and Predictability Programme (CLIVAR) focused on the role of air sea interaction of Kuroshio extension in the climate system. The Argo observed the thermohaline structure in the deep-sea area to support the study on the Kuroshio air sea heat exchange [5]. In addition to the above national research programs, a large number of Kuroshio research programs have been organized in China and the United States. China's Kuroshio Edge Exchange Process (KEEP) project studied the material exchange process between Kuroshio and the continental shelf of the East China Sea [6, 7]. The United States National Science Foundation (NSF) identified and quantified the dynamic and thermodynamic mechanisms of the interaction between Kuroshio extensions and countercurrent through the Kuroshio Extension Systems Study (KESS) project.

SST is one of the main indicators to characterize the thermal state of Kuroshio, and it can be used to study the air sea interaction in Kuroshio Area, the influence of Kuroshio variation on precipitation anomaly, the relationship between Kuroshio and El Nino-Southern Oscillation (ENSO), global warming, and climate change, which are the hot research issues concerned by a large number of scientists. At present, there have been a lot of research results. For example, in the study of Hosoda and Kawamura [8], the short-term abnormal changes of Kuroshio SST were mainly affected by atmospheric forcing. Wang et al. [9] believed that the interannual variation of SST in the sea area south of Japan and the Kuroshio extension area was mainly caused by ENSO, while the interdecadal variation was related to Pacific Decadal Oscillation (PDO). Numerous research results have shown that Kuroshio's warm current characteristics allow it to transport warm water from low latitude to high latitude and release heat into the atmosphere, thus having an important impact on the climate and atmospheric circulation in East Asia [10–15]. Therefore, the accurate prediction of Kuroshio SST anomalies is of great significance to the study of air sea interaction and climate anomalies in eastern China.

Some scholars have studied SST prediction using deep learning algorithms [16, 17]. However, the algorithm of deep learning is very complicated, which requires a lot of computing resources and the computation process cannot be well understood. In this paper, we hope to make a qualitative prediction of Sea Surface Temperature anomalies through a more lightweight algorithm.

With the continuous advancement of big data, cloud computing, and artificial intelligence technology, as well as the constant improvement in modern computer level, machine learning technique has been widely applied in many fields. Similarly, more and more scholars have applied machine learning technique to meteorological scientific research. Shi et al. [18, 19] used decision tree algorithm to establish a relatively accurate diagnosis or prediction model for road icing and extra strong fog disasters. Zhang et al. [20, 21] used machine learning to establish a more accurate classification prediction model for whether the typhoon path turns and whether the typhoon lands. Geng et al. [22] used the Finite Mixture Model (FMM) algorithm and the Classification and Regression Tree (CART) algorithm to predict the path classification and frequency of tropical cyclones landing in China and achieved good prediction results. According to David et al. [23], Random Forest (RF) algorithm was used to establish a prediction model of Mesoscale Convective System (MCS) based on radar data, satellite data, and model output data. However, there were no rich research achievements with applying machine learning technique to the Kuroshio SST. This paper can analyze the statistical characteristics of Kuroshio SST and establish a simple, scientific, and accurate diagnostic model by using the decision tree algorithm from the perspective of nonlinear algorithm. Through the decision classification results, the distribution characteristics of various types of precipitation in East Asia are analyzed, which provides a new idea and method for the climate prediction of SST in the Kuroshio Area.

2. Data and Methods

2.1. Data Source. This article will use the following three types of data:

(1) A set of 100 climate system indices compiled by the National Climate Center (NCC) from 1961 to 2015 (including 130 climate signal indices, including 88 atmospheric circulation indices, 26 SST indices, and 16 other indices) was used.

(2) The global monthly mean precipitation data from 1961 to 2015 provided by the Global Precipitation Climatology Center (hereinafter referred to as GPCC) has a spatial horizontal resolution of $1.0^\circ \times 1.0^\circ$ [24].

(3) The Comprehensive Ocean-Atmosphere Data Set (COADS) was used to integrate the SST dataset from Hadley Center, UK Met Office, with a spatial horizontal resolution of $1^\circ \times 1^\circ$.

2.2. CHAID Decision Tree Algorithm. Decision tree algorithm is a classical white box classification method in machine learning, suitable for dealing with complex nonlinear problems. This kind of algorithm usually segments the nodes in a recursive way, determines the segmentation threshold of the data according to the preset classification basis and separation excellence, and forms a decision tree when the data is segmented to the termination condition.

The CHAID algorithm, namely, the chi-square automatic cross-check algorithm, is a classification decision tree algorithm, invented by Kass in 1980 [25], and it can segment data according to the chi-square value. The algorithm takes the dependent variable as the root node and classifies different independent variables by calculating the chi-square value χ^2 of data classification. The formula is shown as follows:

$$\chi^2 = \sum_{i=1}^{k} \frac{(A_i - E_i)}{E_i}, \quad i = 1, 2, \ldots, k. \tag{1}$$

In formula (1), A_i is the horizontal observation frequency of i, E_i is the horizontal expected frequency of i, n is the total frequency, and p_i is the expected frequency of i, so that when $E_i = np_i$, k is the number of cells. When n is large, χ^2 statistical features obey $k - 1$ free chi-square distribution.

In this algorithm, classification attributes on each node are selected by the size of attribute chi-square, and the samples are split according to the attribute that can bring the largest chi-square, thus recursively splitting till the stop condition is reached.

The effect test of the algorithm model adopts the method of setting aside, which can use part of the data to train the model, namely, the training set, with another part of the independent data for testing the model, namely, the test set. The ratio of the number of correct training samples to the total number of training samples is the learning accuracy, and the ratio of the number of correct testing samples to the total number of testing samples is the testing accuracy.

2.3. Determination of the Key Area of Kuroshio SST. The common methods for determining key area include correlation, mean-square deviation distribution, and Empirical Orthogonal Function (EOF) analysis. Some studies have shown that external forcing factors, such as SST and snow cover, often lead to abnormal atmospheric circulation and then lead to changes in precipitation in East Asia [26]. In this paper, mean-square analysis of Spring SST in the Northwest Pacific Ocean was conducted, and it was found that (120 – 130°E, 22 – 32°N) was the large mean-square value area not only of Spring SST (as shown in Figure 1), but also in the Kuroshio areas, so that it was selected as the key area of Kuroshio SST (hereinafter referred to as Kuroshio Area).

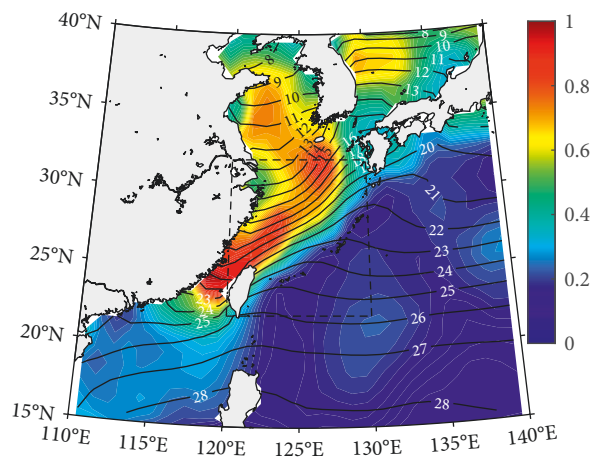


Figure 1: Distribution of climatic mean (isoline, unit °C) and mean-square error (colored area) of Spring SST.

3. Establishment of Diagnostic Model for Whether Spring Kuroshio SST Is High Based on CHAID Algorithm

3.1. Preprocessing of Experimental Data. The CHAID algorithm used in this paper is a classical labeled supervised machine learning algorithm. The most frequently used modeling strategy of the algorithm is the method of setting aside; that is, the total data sample is divided into two mutually exclusive parts: the training sample set and the test sample set. The training set data is used for establishing the decision tree model, and the test set data for testing the generality and robustness of the model. Generally, the training set accounts for 80% of the total sample, so that the test set data accounts for 20% of the total sample. First of all, we regarded the data from 1961 to 2004 as the training set of the model (80%), and the data samples from 2005 to 2015 as the test set (20%). In this paper, the standardized anomaly was used to judge whether Spring SST in Kuroshio Area is high. When the standardized SST anomaly in Kuroshio Area is greater than 1, it is considered that Spring SST in Kuroshio Area is high [27]. "Whether Spring SST in Kuroshio Area is high or not" can be abstracted into a binary classification question of yes or no. As shown in Table 1, according to the statistics of the data samples in the training set and the test set, there are 44 data samples in the training set, among which 7 samples have high SST, and 11 data samples in the test set, among which 4 samples have high SST in Kuroshio Area.

In this paper, climate signals in spring were selected as diagnostic factors to diagnose whether Spring SST in Kuroshio Area is high. By averaging the values in March, April, and May of each of the 100 climate system indices provided by BCC, 130 climate signal indices in spring have been obtained.

3.2. The Classification Diagnosis Model of Spring Kuroshio SST Based on CHAID Decision Tree Algorithm. Whether Spring Kuroshio SST is high was taken as the target variable of the model, with the 130 indices provided by National Climate Center (NCC) as the input variables of the model. The preprocessed training set was input into CHAID algorithm, and then the decision tree could be obtained through calculation (Figure 2).

The decision tree is intuitive in form and in line with the logical judgment thinking mode of human beings. By observing the decision tree, it can be found that the root node is the North American polar vortex intensity index; in other words, the most important factor for whether Spring SST in Kuroshio Area is high is the intensity of the Polar vortex in North America. In the decision tree model, every root node to a leaf node (T/F) can be abstracted into a decision rule in the form of "If... Then". Moreover, each rule in the decision tree can be summarized to form the decision rule set that is convenient for people to learn and use (see Table 2).

The rule set for judging whether Spring SST in Kuroshio Area is high or not was abstracted through the decision tree, and the decision tree model was established according to the training set data. The learning accuracy of the decision tree model reached 95.45%. Then, the decision rules were abstracted from root node to leaf node, and the learning accuracy of each rule can be obtained, which is convenient for reference with the actual situation. Finally, the generalization ability of the decision tree model was tested with the preprocessed test set data, with the test accuracy 81.82%. In conclusion, as shown in Table 3, this decision tree model has good classification effect and strong generalization ability and can provide a concise, understandable, and valuable reference for diagnosing whether Spring SST in Kuroshio Area is high.

4. Strong SST Model in Kuroshio Area

Based on the corresponding years of the three types (Type A, B, and D) of strong Spring SST in Kuroshio Area, their abnormal characteristics of Spring SST distribution (i.e., the anomaly value obtained from the average climate reduction of the type) were analyzed, respectively, to provide a scientific basis for the climate prediction of strong Spring SST in Kuroshio Area and its impact on precipitation in East Asia. The Spring SST distribution of cumulative and single anomalies of Types A, B, and D can be shown in Figure 3. As can be seen from Figure 3(a), in the spring of strong Kuroshio SST years, the entire

Table 1: Frequency distribution of whether Kuroshio SST is high in the total sample, training set sample, and test set sample.

	Total sample (1951–2015)	Training set sample (1951–2004)	Test set sample (2005–2015)
High SST	11	7	4
No high SST	44	37	7

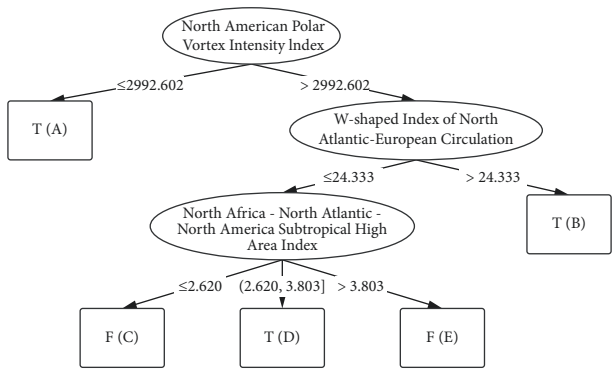


Figure 2: Diagnosis decision tree of whether Spring Kuroshio SST is high based on CHAID algorithm.

Table 2: Diagnostic rule set of whether Spring Kuroshio SST is high discovered by CHAID algorithm.

Decision rules	Whether the SST is high	Decision attributes	Learning accuracy
Rule A: If North American polar vortex intensity index ≤2992.602	YES	North American polar vortex intensity index	2/2 = 100.00%
Rule B: If North American polar vortex intensity index >2992.602 and W-shaped index of North Atlantic-European circulation >24.333	YES	North American polar vortex intensity index, W-shaped index of North Atlantic-European circulation	2/3 = 66.67%
Rule C: If North American polar vortex intensity index >2992.602 and W-shaped index of North Atlantic-European circulation ≤24.333 and north Africa-North Atlantic-North America subtropical high area index ≤2.62	NO	North American polar vortex intensity index, W-shaped index of North Atlantic-European circulation, North Africa-North Atlantic-North America subtropical high area index	20/ 20 = 100.00%
Rule D: If North American polar vortex intensity index >2992.602 and W-shaped index of North Atlantic-European circulation ≤24.333 and (North Africa-North Atlantic-North America subtropical high area index >2.62 and North Africa-North Atlantic-North America subtropical high area index ≤3.803)	YES	North American polar vortex intensity index, W-shaped index of North Atlantic-European circulation, North Africa-North Atlantic-North America subtropical high area index	2/2 = 100.00%
Rule E: If North American polar vortex intensity index >2992.602 and W-shaped index of North Atlantic-European circulation ≤24.333 and North Africa-North Atlantic-North America subtropical high area index >3.803	NO	North American polar vortex intensity index, W-shaped index of North Atlantic-European circulation, North Africa-North Atlantic-North America subtropical high area index	16/ 17 = 94.12%

Table 3: Annual distribution of Kuroshio SST anomalies according to various rules.

Rules	Years
Rule A (high SST)	1999, 2000, 2006, 2007, 2011
Rule B (high SST)	1972, 1997, 2002, 2009
Rule C (no high SST)	1961, 1962, 1963, 1964, 1965, 1966, 1967, 1968, 1969, 1973, 1975, 1976, 1977, 1978, 1985, 1986, 1987, 1990, 1993, 1994
Rule *D* (high SST)	1998, 2001
Rule *E* (no high SST)	1970, 1971, 1974, 1979, 1980, 1981, 1982, 1983, 1984, 1988, 1989, 1991, 1992, 1995, 1996, 2003, 2004, 2005, 2008, 2010, 2012, 2013, 2014, 2015

Note. Underlined years are classified incorrectly.

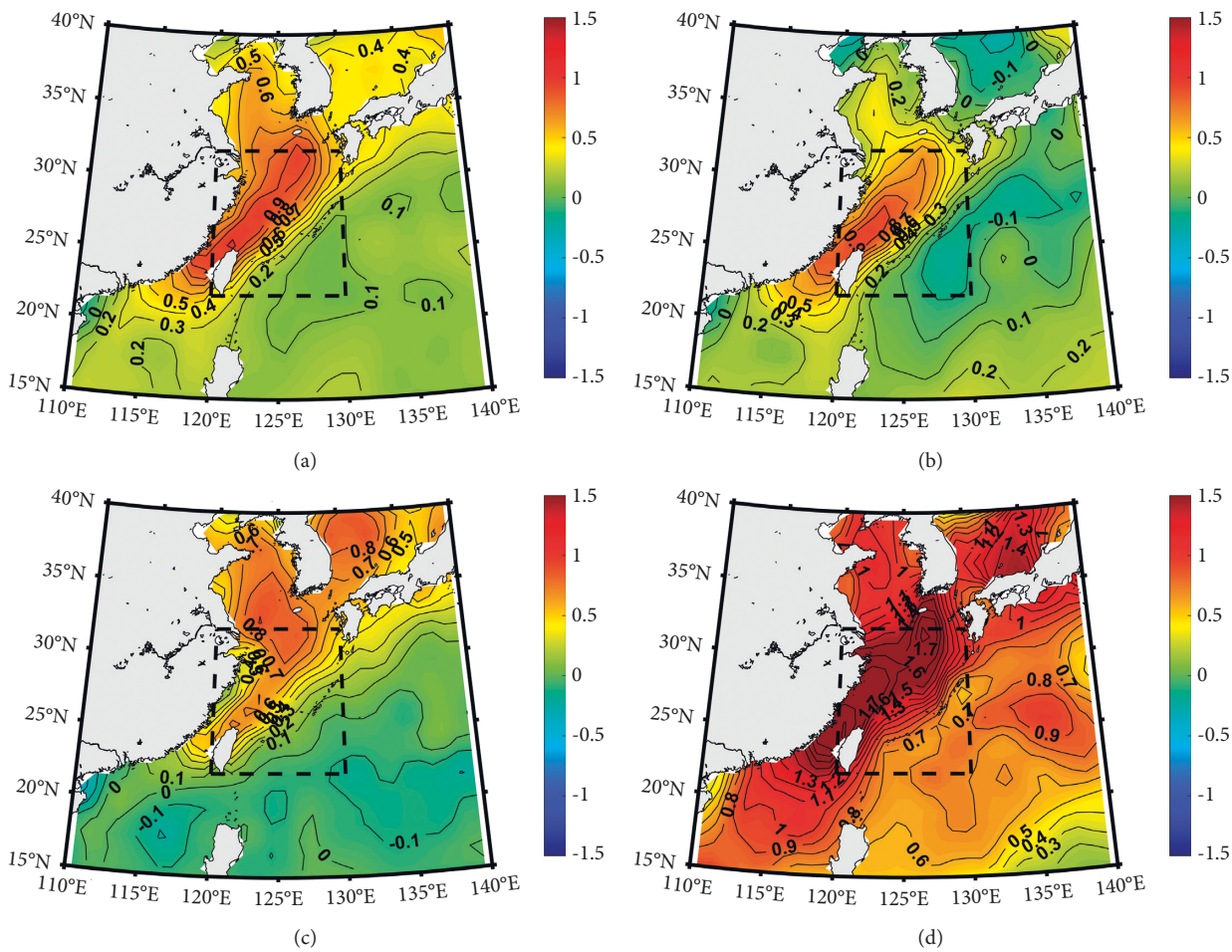


Figure 3: Distribution of Spring SST anomalies in Kuroshio Area. (a) Strong Annual Climate Mean; (b) Type A Anomaly; (c) Type B Anomaly; (d) Type D Anomaly.

Northwest Pacific Region presents consistent positive SST anomalies, and the Kuroshio Area is basically covered by positive SST anomalies greater than or equal to 0.5°C. The SST anomaly center of Type A is located in the Taiwan Strait, and its intensity reaches 0.9°C. The SST anomaly distribution is high in the west and low in the east in the Northwest Pacific Ocean, and the distribution in the Kuroshio Area is also "high in the west and low in the east." The distribution characteristics of high in the north and low in the south are shown in the SST anomalies of Type B in the Northwest Pacific and Kuroshio Area. In the area north of 20°N, the intensity of abnormally high SST can increase with the increase of latitude, but in the area south of 20°N, there is an anomalously low SST. The distribution characteristics of SST anomalies in Kuroshio Area are similar to those in the Northwest Pacific Ocean, showing a pattern of "high in the north and low in the south." The anomalous high SST area of Type D covers the whole Northwest Pacific Ocean. The SST in Kuroshio Area is abnormally higher than 0.7°C, with an SST anomaly center of 1.7°C, presenting a "uniformly high type" distribution. The anomalous high SST area of Type D covers the whole Northwest Pacific Ocean. The SST in Kuroshio Area is abnormally higher than 0.7°C, and there is an SST anomaly center of 1.7°C, presenting a "uniformly high" distribution. To sum up, in spring, the SST of Types A, B, and D is generally strong in Kuroshio Area, but the SST anomaly intensity and distribution characteristics of the three types are obviously different, deserving more attentions in the climate prediction of the SST in Kuroshio Area.

In order to further analyze the anomalous distribution characteristics of summer precipitation in eastern China when the SST in Kuroshio Area is abnormally warming, the summer daily precipitation distribution of cumulative and single anomalies of Type A, B, and D is shown in Figure 4. It can be seen from Figure 4 that, in the summer of strong Spring Kuroshio SST year, there is a negative anomaly area of daily precipitation in China east of 110°E and north of the Yellow River (35°N), while the south of the Yellow River is covered by an obvious positive anomaly, indicating that, in the summer of strong Spring Kuroshio SST years, there is less precipitation in the north of the Yellow River in eastern China and more precipitation in the south of the Yellow River. It can be seen from the anomaly distribution of Type A abnormal precipitation that the zero line of precipitation anomaly is

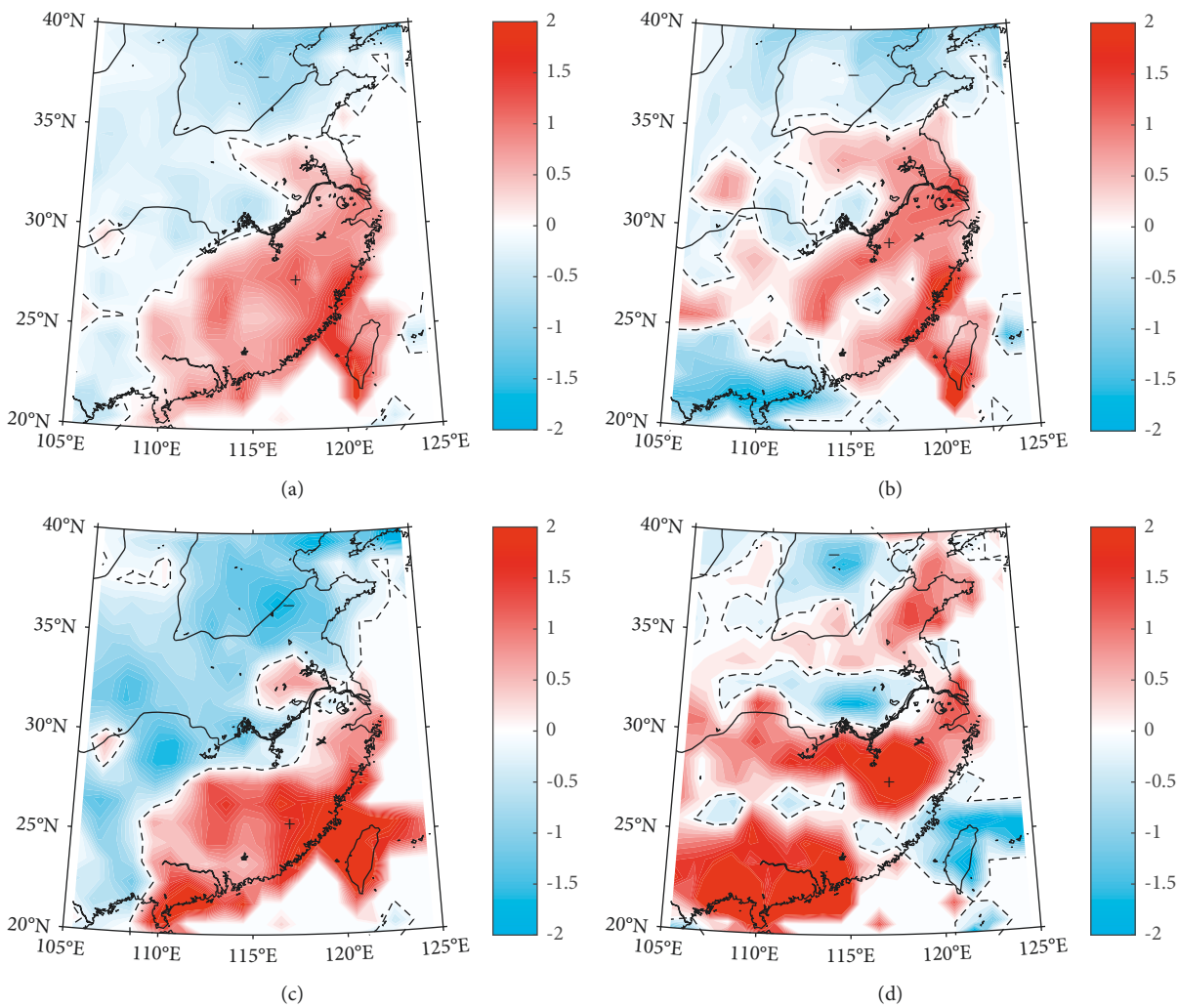


Figure 4: Distribution of summer daily precipitation anomalies in the Spring Kuroshio SST anomaly years. (a) Strong Annual Climate Mean; (b) Type A Anomaly. (c) Type B Anomaly; (d) Type D Anomaly.

located in the area of the Yellow River (35°N). The area north of the Yellow River is basically controlled by negative anomaly, while the area south of the Yellow River is covered by positive anomaly, indicating that, in a year of strong Spring SST of Type A, the summer precipitation is less in the area north of the Yellow River and more in the area south, with the distribution characteristics of "less in the north and more in the south" in the precipitation anomaly. The summer precipitation anomaly of Type B is also basically characterized by "less in the north and more in the south," but its zero line of precipitation anomaly, in the south of Type A's, is located along the Yangtze River. The north of the Yangtze River is a negative anomaly area, but the south is a positive anomaly area. It can be shown that, in a year of strong Spring SST of Type B, the summer precipitation is less in the north of the Yangtze River, but it is more in the south. It is worth noting that the summer precipitation of Types A and B is consistently more in the south of the Yangtze River, while the summer precipitation of Type D is slightly less in that area. In conclusion, when the Spring SST in Kuroshio Area is abnormally warm in the years of Types A, B, and D, the distributions of summer precipitation can present different characteristics, providing more reference for studying the impact of the Kuroshio SST anomalies on the precipitation in East Asia.

5. Summary and Discussion

In this paper, the CHAID algorithm is used to establish a multitree classification model to determine whether Spring SST in Kuroshio Area is high or not, and then the rule set of whether Spring SST in Kuroshio Area is high or not under different climatic backgrounds was obtained. According to the three rules of high SST in Kuroshio Area, the distribution characteristics of Kuroshio SST and the distribution of summer precipitation anomalies in eastern China were analyzed, respectively, thus drawing the following conclusions:

(1) With 130 circulation indices as input variables, the prediction model of whether Spring SST is high in Kuroshio Area was established by using CHAID algorithm, and then the classification rule set was obtained. The data of the 44 years from 1961 to 2004

were used as the training set of the model, and the remaining years as the test set. The training accuracy of the model for whether Spring Kuroshio SST is high reached 95.45%, and the test accuracy 81.82%.

(2) In spring, the SST of Types A, B, and D were all high in Kuroshio Area, but the intensity and distribution of abnormal high SST were different in the three types, which is worthy of attention in the diagnosis of Spring SST in Kuroshio Area.

(3) Although the Spring Kuroshio SST of Types A, B, and D were all abnormally high, there were significant differences in the distribution of summer daily precipitation anomalies in eastern China, which can provide more reference for studying the influence of Kuroshio SST anomalies on precipitation in East Asia.

With the advent of the era of big data, machine learning technique has been well applied in many fields. The accumulation of SST data and climate indices can open a window for the application of machine learning technique in precipitation prediction and provide a new way of statistical prediction. In this paper, the Spring SST in Kuroshio Area was regarded as the research object, and then the climate prediction of "whether SST is high or not" was carried out. Further research and discussion will be necessary for how to use machine learning technique to make a more refined prediction of Spring Kuroshio SST in its timescales and scope.

Data Availability

Some data source can be downloaded from https://cmdp.ncc-cma.net/Monitoring/cn_index_130.php and https://www.metoffice.gov.uk/hadobs/hadisst/data/download.html. Other research data used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest to report regarding the present study.

Acknowledgments

This paper was funded by 2019 Key Project of Jiangsu Meteorological Bureau (KZ201901) and 2021 Jiangsu Provincial Young Scientific and Technological Talents Project.

References

[1] H. Nitani, "Beginning of the Kuroshio," in *Kuroshio: Physical Aspect of the Japan Current*, pp. 129–163, University of Washington Press, Seattle, WA, USA, 1972.

[2] X. Lu, J. M. Song, H. M. Yuan, and N. Li, "Carbon exchange between Kuroshio and adjacent shelf waters," *Advances in Earth Science*, vol. 30, no. 002, pp. 214–225, 2015.

[3] Z. Y. Wang, *Study on Numerical Simulation of Kuroshio's Influence on China's Offshore Circulation*, National Marine Environmental Forecasting Center, China, 2012.

[4] Y. B. Fan and S. Z. Pu, "Research progress on global change in the field of China's marine science," *Advances in Earth Science*, vol. 3, no. 1, pp. 62–71, 1998.

[5] C. Y. Zhang, Z. G. Feng, and X. K. Zhang, "Analysis on research progress of Kuroshio," *World Sci-Tech R&D*, vol. 39, no. 3, pp. 239–249, 2017.

[6] W. Li, Q. Wang, and J. R. Ma, "Study on temporal and spatial distribution and formation mechanism of Kuroshio front east of Taiwan," *Hai Yang Tong Bao*, vol. 7, no. 4, pp. 401–408, 2011.

[7] W. Gtf, S. Y. Chao, and Y. C. Chung, "KEEP-exchange processes between the Kuroshio and the East China sea shield: preface," *Continental Shelf Research*, vol. 20, no. 4, pp. 331–334, 2000.

[8] K. Hosoda and H. Kawamura, "Seasonal variation of space/time statistics of short-term sea surface temperature variability in the Kuroshio region," *Journal of Oceanography*, vol. 61, no. 4, pp. 709–720, 2005.

[9] S. S. Wang, Y. P. Guan, and J. P. Huang, "Preliminary analysis on the correlation between the variation characteristics of SST and the atmospheric circulation in the Kuroshio and its extension area," *Acta Physica Sinica*, vol. 16, pp. 516–526, 2012.

[10] H. M. Xu, "Relationship between summer precipitation in south China and global SST," *Transactions of Atmospheric Sciences*, vol. 020, no. 003, pp. 392–399, 1997.

[11] Y. F. Li and Y. H. Ding, "SST, LST and summer abnormal precipitation in eastern China," *Climatic and Environmental Research*, vol. 7, no. 001, pp. 87–101, 2002.

[12] Z. X. Li and Z. B. Sun, "Numerical experiment of the influence of winter Kuroshio SSTA on the east Asian summer monsoon," *Transactions of Atmospheric Sciences*, vol. 29, no. 001, pp. 62–67, 2006.

[13] Y. K. Li, "Statistical analysis of Kuroshio SST, precipitation during flood season in China and upper air flow field in east Asia," *Journal of the Meteorological Sciences*, vol. 1, no. 03, pp. 263–275, 1989.

[14] R. Zhang, H. M. Xu, and B. C. Zhang, "Seasonal variation characteristics and causes of Kuroshio front in the East China Sea," *Journal of the Meteorological Sciences*, vol. 36, no. 002, pp. 203–211, 2016.

[15] A. Xie, H. M. Xu, and M. M. Xu, "Response characteristics of the atmosphere to the Kuroshio front under the background of prevailing wind on the sea surface," *Journal of the Meteorological Sciences*, vol. 34, no. 4, pp. 355–364, 2014.

[16] P. P. Sarkar, P. Janardhan, and P. Roy, "Prediction of sea surface temperatures using deep learning neural networks," *SN Applied Sciences*, vol. 2, no. 8, pp. 1458–1472, 2020.

[17] Y. G. Ham, J. H. Kim, and J. J. Luo, "Deep learning for multi-year ENSO forecasts," *Nature*, vol. 573, no. 7775, pp. 568–572, 2019.

[18] D. W. Shi, C. Li, Y. M. Shi et al., "Study on the localization diagnosis of extra heavy fog on the background of the fog weather based on machine learning algorithms," *Journal of Catastrophology*, vol. 33, no. 02, pp. 193–199, 2018.

[19] D. W. Shi, H. T. Geng, and C. Ji, "Construction and application of road icing forecast model based on CHAID decision tree algorithm," *Journal of the Meteorological Sciences*, vol. 35, no. 2, pp. 204–209, 2015.

[20] W. Zhang, Y. Leung, and J. C. L. Chan, "The analysis of tropical cyclone tracks in the western north pacific through data mining. Part I: tropical cyclone recurvature," *Journal of Applied Meteorology and Climatology*, vol. 52, no. 6, pp. 1394–1416, 2013a.

[21] W. Zhang, Y. Leung, and J. C. L. Chan, "The analysis of tropical cyclone tracks in the Western North Pacific through data mining. Part II: tropical cyclone landfall," *Journal of Applied Meteorology and Climatology*, vol. 52, no. 6, pp. 1417–1432, 2013b.

[22] H. Geng, D. Shi, W. Zhang, and C. Huang, "A prediction scheme for the frequency of summer tropical cyclone landfalling over China based on data mining methods," *Meteorological Applications*, vol. 23, no. 4, pp. 587–593, 2016.

[23] D. Ahijevych, J. O. Pinto, J. K. Williams, and M. Steiner, "Probabilistic forecasts of mesoscale convective system initiation using the random forest data mining technique," *Weather and Forecasting*, vol. 31, no. 2, pp. 581–599, 2016.

[24] Y. Zi, Y. L. Xu, and Y. F. Fu, "Comparison of climatic characteristics of precipitation observed by GPCP and Chinese stations," *Acta Meteorologica Sinica*, vol. 1, no. 65, pp. 63–74, 2007.

[25] Q. Huang, "Personal income analysis based on CHAID decision tree," *Mathematical Theory and Application*, vol. 29, no. 04, pp. 33–37, 2009.

[26] X. L. Chen, X. S. Shen, and H. P. Chen, "Analysis of the impact of land surface processes on the numerical prediction of heavy precipitation in the Huaihe River Basin in 2007," *Journal of Tropical Meteorology*, vol. 26, no. 6, pp. 667–679, 2010.

[27] Y. Huan and Y. Li, "Relationship between the synergy of Summer East Asian monsoon and South Asian monsoon and summer precipitation anomaly in southern China," *Plateau Meteorology*, vol. 37, no. 06, pp. 1563–1577, 2018.

Hindawi
Advances in Meteorology
Volume 2022, Article ID 1926854, 20 pages
https://doi.org/10.1155/2022/1926854

Research Article

Evaluation of Satellite Rainfall Products over the Mahaweli River Basin in Sri Lanka

Helani Perera,[1] Shalinda Fernando,[1] Miyuru B. Gunathilake,[2] T. A. J. G. Sirisena,[3] and Upaka Rathnayake[1]

[1]*Department of Civil Engineering, Faculty of Engineering, Sri Lanka Institute of Information Technology, Malabe, Sri Lanka*
[2]*Division of Hydrology and Aquatic Environment, Environment and Natural Resources, Norwegian Institute of Bioeconomy and Research, Ås, Norway*
[3]*Department of Water Engineering and Management, University of Twente, Enschede, Netherlands*

Correspondence should be addressed to Upaka Rathnayake; upaka.r@sliit.lk

Received 17 December 2021; Accepted 18 March 2022; Published 25 April 2022

Academic Editor: Antonio Donateo

The availability of accurate spatiotemporal rainfall data is of utmost importance for reliable predictions from hydroclimatological studies. Challenges and limitations faced due to the absence of dense rain gauge (RG) networks are seen especially in the developing countries. Therefore, alternative rainfall measurements such as satellite rainfall products (SRPs) are used when RG networks are scarce or completely do not exist. Noteworthy, rainfall data retrieved from satellites also possess several uncertainties. Hence, these SRPs should essentially be validated beforehand. The Mahaweli River Basin (MRB), the largest river basin in Sri Lanka, is the heart of the country's water resources contributing to a significant share of the hydropower production and agricultural sector. Given the importance of the MRB, this study explored the suitability of SRPs as an alternative for RG data for the basin. Daily rainfall data of six types of SRPs were extracted at 14 locations within the MRB. Thereafter, statistical analysis was carried out using continuous and categorical evaluation indices to evaluate the accuracy of SRPs. Nonparametric tests, including the Mann-Kendall and Sen's slope estimator tests, were used to detect the possibility of trends and the magnitude, respectively. Integrated MultisatellitE Retrievals for Global Precipitation Measurement (IMERG) outperformed among all SRPs, while Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN) products showed dire performances. However, IMERG also demonstrated underestimations when compared to RG data. Trend analysis results showcased that the IMERG product agreed more with RG data on monthly and annual time scales while Tropical Rainfall Measurement Mission Multisatellite Precipitation Analysis–3B42 (TRMM-3B42) agreed more on the seasonal scale. Overall, IMERG turned out to be the best alternative among the SRPs analyzed for MRB. However, it was clear that these products possess significant errors which cannot be ignored when using them in hydrological applications. The results of the study will be valuable for many parties including river basin authorities, agriculturists, meteorologists, hydrologists, and many other stakeholders.

1. Introduction

Water is being regarded as a remarkable substance due to its critical role related to human life. Reduced precipitation amounts can lead to prolonged droughts while excess precipitation rates can result in flooding. However, both extreme conditions disrupt the socioeconomic activities of people and harm the environment adversely. Water plays an important role in a country's economy in terms of irrigated agriculture, hydropower generation, provision of drinking water, and industrial use [1]. This marks the importance of proper water management for both social and economic benefits for a country. For proper management of water, accurate predictions from hydrological modeling are required. Hence, the role of spatially distributed precipitation data is crucial in this context [2].

Precipitation and evapotranspiration are the major components of the hydrologic cycle, and both these components carry significant shares of the hydrologic cycle [3]. Until the early 1980s, the only method used to obtain these

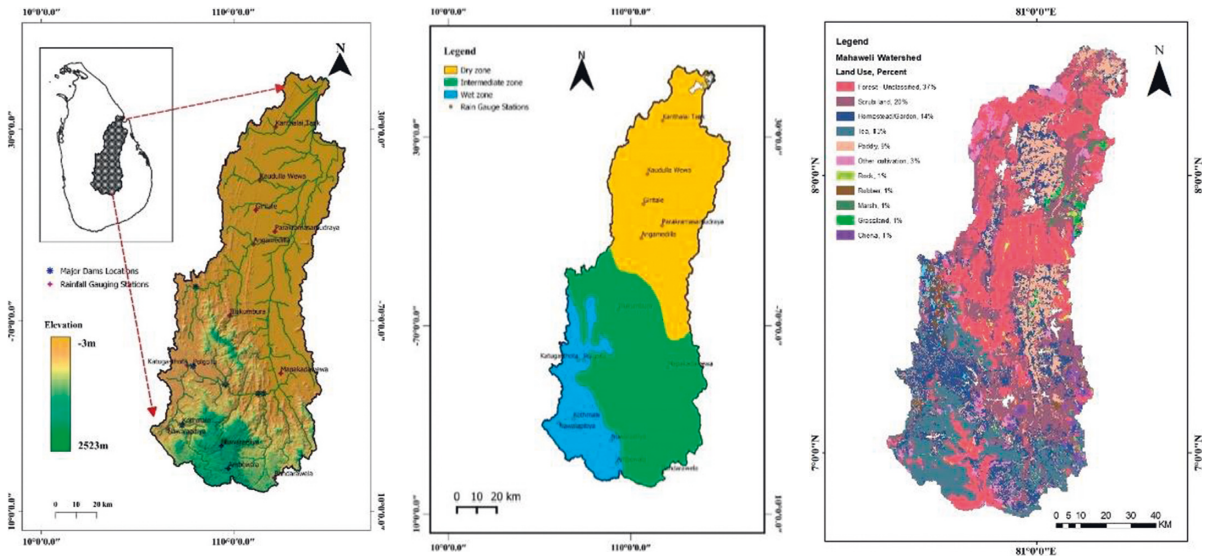
precipitation data was rain gauges (RGs) located across the world [4]. Although RG data are the most accurate method of measuring rainfall [5], there are significant challenges faced when maintaining rain gauge networks. These challenges are due to complex geographical features, harsh climatological environments, and high financial burdens [6]. Some of the reported shortcomings of RG data are missing rainfall days, data transmission dropouts, calibration errors of the RGs, and local wind effects affecting rainfall level estimations [7, 8]. Incidences during the period of war in countries also lead to collapsing of these rain gauges which ultimately resulted in no rainfall records in certain regions of the world [9].

With the continuous evolvement of science and technology, currently, several other methods are available to obtain rainfall data. They are radars, satellites, and reanalysis products. However, high costs for installment and maintenance of radar networks are unbearable for the developing world to adopt radar networks [6]. On the other hand, reanalysis products have significant biases [10]. Therefore, the satellite rainfall product (SRP) data have grabbed the attention of many researchers due to its ability to provide spatiotemporal rainfall data addressing most of the shortcomings faced when using RGs. The SRPs are attractive since they are available at no cost. In addition, the SRPs are available from fine spatial and temporal scales such as 0.10° and 30 minutes. For instance, both the near-real-time product of GSMaP and IMERG provides 30-minute resolution while the IMERG product provides rainfall in 0.10° resolution [11]. There are many SRPs available and most of them have been used in similar research studies as well. Some of them are Tropical Rainfall Measurement Mission (TRMM), Multisatellite Precipitation Analysis (TMPA) [12], Climate Prediction Center Morphing (CMORPH) [13], Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN) [14], and Multisource Weighted Ensemble Precipitation (MSWEP) [15]. But these products have been found to possess inherent systematic uncertainties and sampling errors which need to be clearly studied and identified before using them in hydrological applications [16, 17]. However, for flood and drought predictions, understanding global climate change, and determining the availability of water resources for agriculture and industrial purposes, the climatological data needs to be accurate. Therefore, comparing SRPs by benchmarking with RGs is essential to obtain a sound conclusion on their levels of uncertainty and thereafter to determine their suitability for many practical applications.

Following this requirement, many research studies have been conducted around the world to assess the accuracy and reliability of SRPs with respect to RG data [6, 18–20]. Many studies have demonstrated that TRMM (TMPA) is comparatively better than other SRPs since it used available gauge measurements in its calibration process [21, 22]. Higher efficiency [23] and higher accuracy [24] shown by TRMM 3B42 in research studies had proven its accuracy compared to its near-real-time product (TRMM B42RT). Research studies incorporating CMORPH had observed that it shows dependencies on rainfall estimates [25]. GPM (IMERG) when compared with TMPA in a study conducted in India showed better performance in capturing heavy rainfall events [26]. A study in Ethiopia that focused on CMORPH and PERSIANN-CCS discovered that almost 50% of underestimations can be expected from these SRPs and is cautious to be directly used in flood prediction models [16]. GSMaP-MVK data was proved to produce overestimates when compared with gauge data in a study conducted in Indonesia [27]. Although in the developed world, many studies have been carried out to examine the efficiency of SRPs as an alternative for RGs, the developing region has been least focused in this regard. In the context of Sri Lanka, only a few studies have been carried out to examine the efficiency of SRPs. A previous study was carried out for the Mundeni Aru Basin, Sri Lanka, to evaluate the applicability of SRPs in flood hazard mapping [28]. Two other studies attempted to determine the accuracy of SRPs in streamflow simulation for the Seethawaka watershed, Sri Lanka, using a hydrological model [29, 30].

The Mahaweli River Basin (MRB) in Sri Lanka covers approximately one-sixth of the landmass in the country (Figure 1). The MRB is imperative due to its major role played in hydropower generation and irrigation water supply for many parts of the country. The MRB is the main water supplier for agricultural activities in the eastern dry zone of the country irrigating more than 1,000 km^2 of land. The hydroelectricity produced from six dams of the MRB (Figure 1) supplies more than 40% of the country's electricity. More importantly, all three climatic zones of Sri Lanka fall within the MRB [31, 32]. Despite the vital importance of the MRB to Sri Lanka, the amount of research carried out in this study region is handful. Previously, many studies have been carried out in the subbasins of the MRB and only a very few have focused on the entire MRB. Some of the disciplines which investigated the entire MRB were hydrologic modeling [33], climate change impacts on paddy cultivation in dry zones in the MRB [34], assessing the streamflow variability and rainfall trend of climatic zones [35], El-Nino and La-Nina impact on streamflow [31], drought occurrence and atmospheric circulation [36], and projection of future climate [37]. Other research studies which focused on the subbasins of the MRB investigate the prediction of streamflow in ungauged catchments [38] and flood modeling [39]. Additionally, the Uma-Oya subbasin of the MRB has been an area of great interest for soil erosion. Some of the studies carried out in the Uma-Oya watershed focused on the assessment of the impact of siltation on hydropower generation [40], soil erosion studies [41, 42], land-use change impact on landslides [43], and land-use change impacts on river health [44, 45]. Several field-scale studies have also been conducted to investigate water pollution [46, 47] and Chronic Kidney Disease (CKD) [32].

The present study investigates the suitability of SRPs by comparing them with RGs and examines trends of SRPs and RG in the MRB of Sri Lanka. The accuracy of SRPs is evaluated against RGs through continuous evaluation and categorical indices. In addition, the trend analysis of precipitation products was carried out using the Mann-Kendall test and Sen's slope methods. The current study is the first of



(a) (b) (c)

Figure 1: Maps of MRB, Sri Lanka. (a) The Mahaweli River Basin, Sri Lanka. (b) Climate zones of the MRB. (c) Land use map of the MRB.

its kind which examines the efficiency of SRPs over the MRB. Hence, the present study carries a significant novelty to the research community, opening doors for the use of SRPs in the absence of RG data in the MRB. Stakeholders of the MRB including river basin authorities, agriculturists, meteorologists, hydrologists, and stakeholders (including government authorities and farmers) will be benefited from the results of the study.

2. Study Area and Datasets

2.1. Study Area. The MRB is the longest river in Sri Lanka extending to 335 km. The Mahaweli River starts from the SriPada Mountains in the central hills. The drainage area of the MRB is 10,448 km^2, covering one-sixth of the landmass in the country. Figure 1(a) illustrates the elevation map of MRB obtained by extracting Digital Elevation Models (DEMs) from the Shuttle Radar Topography Mission (SRTM) with a resolution of 30 m × 30 m. Depending on the climatic and geographical distribution of annual rainfall amount, three climatic regions (wet, intermediate, and dry zones) fall in the MRB (Figure 1(b)) [36].

The annual rainfall in the country is influenced by four monsoon seasons of Southwest Monsoon (SWM) from May to September, Northeast Monsoon (NEM) from December to February, first intermonsoon (FIM) during March and April, and second intermonsoon (SIM) during October and November. Having the wet zone of the river basin in the southwest part of the country and the dry zone in the northeast part, the annual rainfall on the MRB is dependent on all these four seasons with the main ones being SWM and NEM [48]. The river basin in three climatic zones can be clearly seen in Figure 2(a). The annual rainfall in this basin is estimated to be 2500 mm of which 900 mm is discharged back to the sea. This produces an annual average discharge of 8.4 billion m^3 in the basin [49]. Tea estates and forest areas can abundantly be seen upstream of the MRB, whereas cultivated areas are frequent downstream of MRB. The land-use types of MRB obtained from the Department of Surveys, Sri Lanka, with a resolution of 1 km × 1 km are given in Figure 1(c). Therefore, as illustrated in Figure 1(c), the upper catchment of the MRB is mainly occupied by agricultural lands (mostly tea cultivation) while the lower catchment is by forest cover.

2.2. Datasets

2.2.1. Rain Gauge Data. For the analysis of this study, based on the availability of rainfall data, 14 rain gauge stations in the wet, the intermediate, and the dry zones of the MRB were obtained from the Department of Meteorology, Sri Lanka. The selected 14 rain gauge stations with their location and period are listed in Table 1. The missing data percentage in the rain gauge stations was below 5%; therefore, rainfall data from neighboring stations were used to fill the missing values. The average rain gauge density in the Mahaweli catchment is 1.34 gauges per 1000 km^2. This is a significantly high value when compared with several other studies. Mu et al. [50] had an average rain gauge density of 0.53 gauges per 1000 km^2, Paca et al. [51] had 0.11 gauges per 1000 km^2, and Calvante et al. [52] had only 0.01 gauges per 1000 km^2.

2.2.2. Satellite Rainfall Products. Similar to the gauge measurements, satellite rainfall products were also obtained for the 14 station locations mentioned in Table 1. Six SRPs (Table 2) used in this analysis are Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN), Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks–Cloud Classification System (PERSIANN-CCS), Precipitation Estimation from Remotely Sensed Information

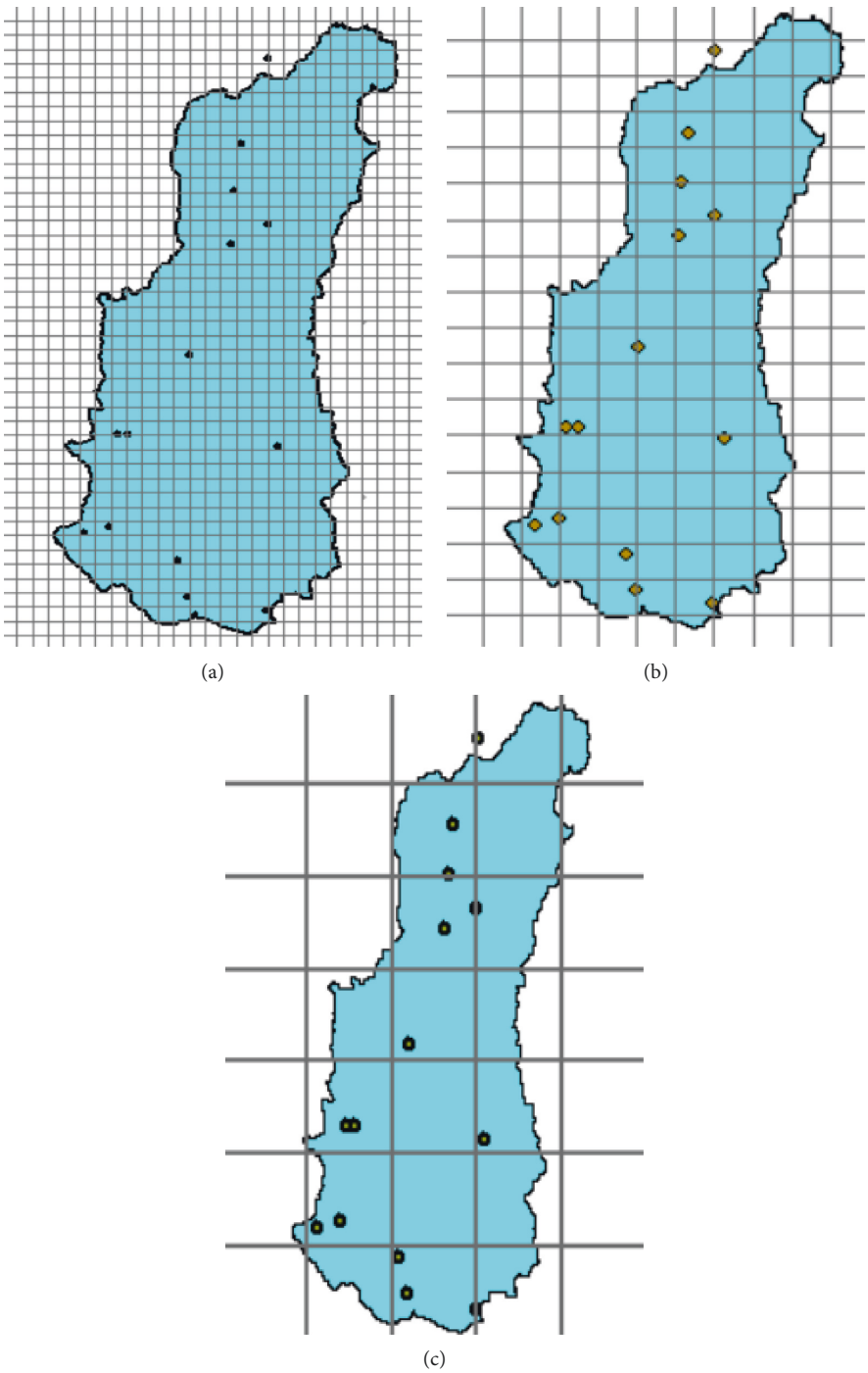


Figure 2: Grid layout and location of RGs. (a) For $0.04^\circ \times 0.04^\circ$. (b) For $0.10^\circ \times 0.10^\circ$. (c) For $0.25^\circ \times 0.25^\circ$.

using Artificial Neural Networks–Climate Data Record (PERSIANN-CDR), Integrated MultisatellitE Retrievals for GPM (IMERG) Version 6, Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis (TMPA) 3B42 (Version 7), Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis (TMPA) 3B42RT (Version 7), and Integrated MultisatellitE Retrievals for Global Precipitation Measurements (IMERG). The grid layout and location of RGs are demonstrated in Figure 3.

TRMM-3B42 and 3B42RT: TRMM 3B42/3B42RT satellite products were produced through a combined mission between the National Aeronautics and Space Association (NASA) and the Japan Aerospace Exploration Agency (JAXA) with the finest temporal resolution of 3 hours and a spatial resolution of $0.25^\circ \times 0.25^\circ$. TRMM 3B42 version 7 product is a research-grade satellite product and is one out of the two products of TRMM 3B42 retrievals. TRMM Combined Instrument (TCI) estimate employing TRMM Microwave Imager (TMI) and the Precipitation Radar (PR)

Table 1: Details of rain gauge stations.

Rain gauge stations	Latitude	Longitude	Timespan
Ambewala	6° 52′ 9.36″N	80° 47′ 44.34″E	1983–2016
Angamedilla	7° 51′ 18″N	80° 54′ 25.19″E	1983–2018
Bandarawela	6° 49′ 47.88″N	80° 59′ 52.71″E	1983–2016
Giritale	8° 0′ 16.10″N	80° 54′ 58.58″E	1983–2017
Illukumbura	7° 32′ 38.86″N	80° 48′ 6.43″E	1983–2015
Kanthalai tank	8° 22′ 16.47″N	81° 0′ 10.46″E	1987–2018
Katugasthota	7° 19′ 26.78″N	80° 37′ 13.96″E	1990–2019
Kaudulla wewa	8° 8′ 12.52″N	80° 56′ 1.86″E	1983–2017
Kothmale	7° 3′ 52.91″N	80° 35′ 54.79″E	1985–2018
Mapakadawewa	7° 17′ 21.55″N	81° 1′ 32.78″E	1983–2016
Nawalapitiya	7° 2′ 51″N	80° 32′ 3.99″E	1989–2017
Nuwaraeliya	6° 58′ 11.99″N	80° 46′ 11.99″E	1990–2019
Parakramasamudraya	7° 54′ 36.23″N	81° 0′ 1.78″E	1983–2018
Polgolla	7° 19′ 20.98″N	80° 38′ 45.94″E	1988–2018

Table 2: Details of SRPs used in the study.

Product	Data provider	Temporal coverage	Finest time resolution	Spatial resolution	Spatial coverage
TRMM-3B42 V7	NASA[1]	01/1998 to 12/2019	3 hours	0.25° × 0.25°	50°N-50°S
TRMM-3B42RT V7	NASA	03/2000 to 12/2019	3 hours	0.25° × 0.25°	60°N-560°S
IMERG V06	NASA	06/2000 to present	30 minutes	0.10° × 0.10°	90°N-90°S
PERSIANN	CHRS[1]	03/2000 to present	1 hour	0.25° × 0.25°	60°N-60°S
PERSIANN-CSS	CHRS	01/2003 to present	1 hour	0.04° × 0.04°	60°N-60°S
PERSIANN-CDR	CHRS	01/1983 to present	1 day	0.25° × 0.25°	60°N-60°S

[1] Center for Hydrometeorology and Remote Sensing (CHRS).
[2] National Aeronautics and Space Administration, USA (NASA).

data are the sources used to estimate precipitation through the TRMM satellite [12]. The spatial coverage of TRMM is between 50°N and 50°S [53]. This rainfall data is produced after two months from the observation time and is called a post-real-time product. The combined use of microwaves, infrared waves, and ground observed rainfall measurements makes it a better product in estimating rainfalls overall [19]. TRMM 3B42RT version 7 is a near-real-time product which is the remaining product of TRMM 3B42 retrievals.

IMERG: Integrated MultisatellitE Retrievals for Global Precipitation Measurements (IMERG) is another satellite product developed by NASA. The recently released version (V06 B) of IMERG Final Run was used in this study. This product has a spatial resolution of 0.10° × 0.10° and spatial coverage of 50°N-50°S. Moreover, the finest temporal resolution is 30 minutes. IMERG product is based on low-earth orbit satellites and geostationary satellites together with high spatiotemporal observed rainfall data. The algorithm used to produce IMERG is intended to intercalibrate, merge, and interpolate "all" satellite microwave precipitation estimates, together with microwave-calibrated infrared satellite estimates, precipitation gauge analyses, and potentially other precipitation estimators. The V06B product is considered a huge improvement over the superseded versions of IMERG since it was updated further in data processing, algorithms, and verification which effectively enhanced its detection accuracy [54, 55].

PERSIANN: Artificial Neural Network (ANN) algorithm is used in Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN) to estimate rainfall. In order to estimate precipitation through PERSIANN IR brightness temperature data from global geostationary satellites is obtained from Climate Prediction Center (CPC) and National Oceanic and Atmospheric Administration (NOAA). This product is available from 2000 to the present and the finest temporal coverage of 1 hour is available. The spatial resolution of the product is 0.25° with a spatial coverage of 60°N-60°S. In order to account for the high uncertainties in associating the relationships between precipitation and cloud-top brightness temperature due to cloud properties and atmospheric conditions, the ANN parameters are updated with rainfall estimates from low-orbital satellites whenever independent estimates of rainfall are available [14, 56].

PERSIANN-CCS: PERSIANN-Cloud Classification System (PERSIANN-CCS) is a real-time global high-resolution SRP. The ANN algorithm used in PERSIANN-CCS extracts local and regional cloud features from infrared geostationary satellite imagery. The information is extracted from the whole cloud patch and provides multiple infrared brightness temperature versus rainfall rate relationships for different cloud classification types enabling this product to generate variable rain rates at a given brightness temperature and variable rain/no-rain IR thresholds for different cloud types

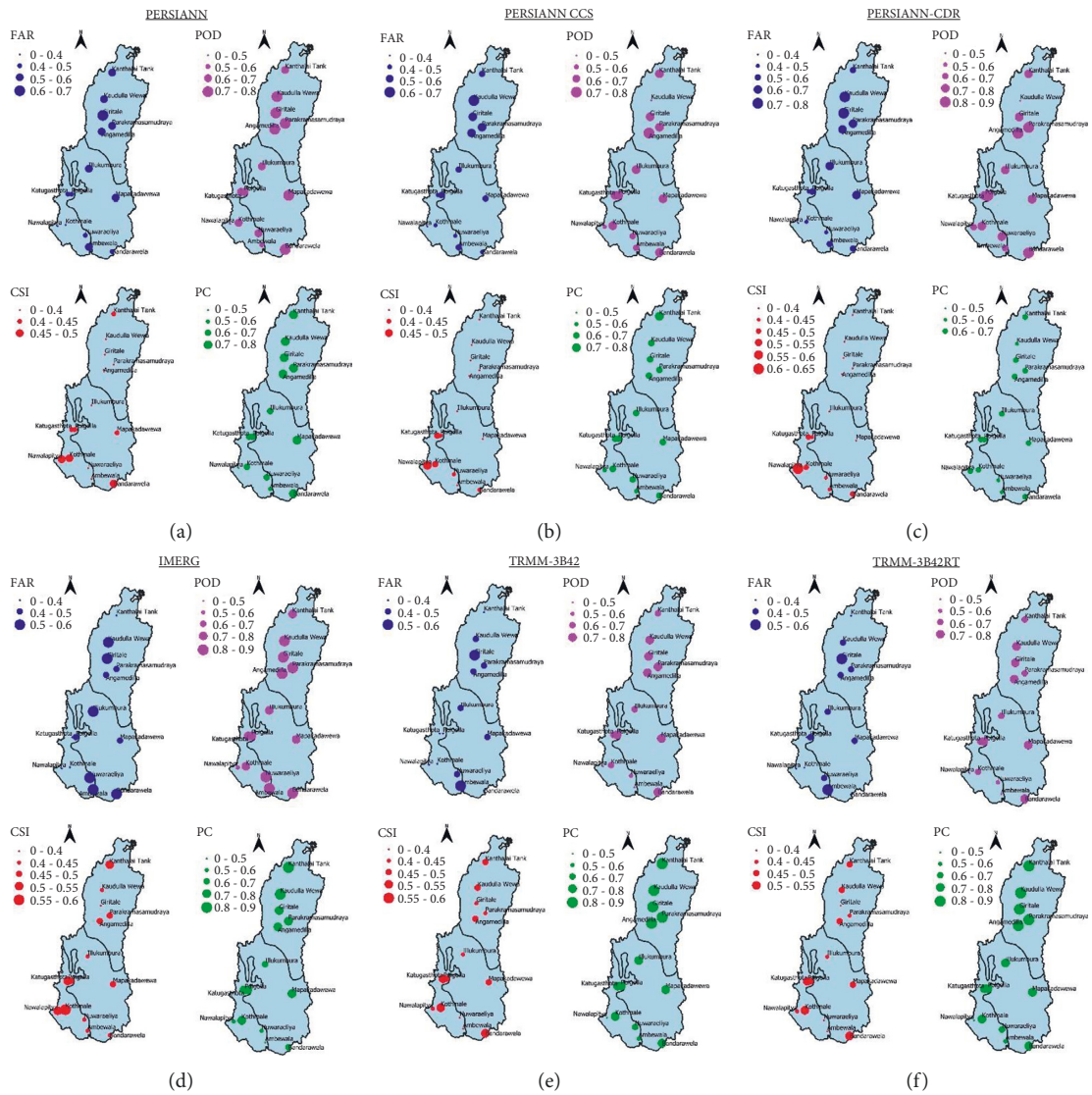


Figure 3: Categorical evaluation for light rainfall events. (a) For PERSIANN. (b) For PERSIANN-CCS. (c) For PERSIANN-CDR. (d) For IMERG. (e) For TRMM-3B42. (f) For TRMM-3B42RT.

[57, 58]. This product has a spatial resolution of $0.04^{\circ} \times 0.04^{\circ}$ and the finest temporal frequency of 1 hour is available. The PERSIANN-CCS data is available from 2003 to the present.

PERSIANN-CDR: PERSIANN-Climate Data Record (PERSIANN-CDR) developed using the PERSIANN algorithm was developed from the National Climatic Data (NCDC) CDR program of the National Oceanic and Atmospheric Administration (NOAA) [59]. The finest temporal resolution of this product is 1 day with a 60°N-60°S spatial coverage and a $0.25^{\circ} \times 0.25^{\circ}$ spatial resolution. Infrared brightness temperature CDR from Gridded Satellite (GridSat)-B1 of the International Satellite Cloud Climatology Project (ISCCP) is used as the input used to train the neural network of the PERSIANN model [19]. The PERSIANN-CDR data obtained were then bias-corrected using 2.5° monthly Global Precipitation Climatology Project (GPCP) data. This SRP was mainly developed with the aim of providing a long-term, high-resolution, and global precipitation dataset to be used in studies to determine changes and trends in daily precipitation, especially extreme precipitation events, due to climate change and natural variability [60].

2.2.3. Data Extraction. The six satellite products used in this study were extracted in different methods. PERSIANN group of products were directly obtained from the CHRS data portal as CSV files. IMERG and TRMM products were obtained as NetCDF (Network Common Data Form) files from the NASA GESDISC portal. Afterward, IMERG was extracted through the process of merging the files in Climate Data Operator (CDO) followed by the extraction using *R* coding in RStudio. The TRMM products were merged using a similar approach as in IMERG, but the extraction of the point rainfall data was done using MATLAB 9.6.

2.3. Overall Methodology. Statistical indices were used to determine the accuracy of the satellite data sets with respect

to RG data. Nonparametric tests were used to analyze the trend and to find the magnitude of the trends observed.

2.3.1. Categorical Evaluation Indices. To examine the detection accuracy of SRPs, there are four categorical indices, namely, the False Alarm Ratio (FAR), the Critical Success Index (CSI), the Probability of Detection (POD), and the Proportion Correct (PC). FAR gives the fraction of times the satellite detects rainfall events that did not actually occur. POD provides a quantitative measure of the number of times the satellite will detect rainfall events accurately. CSI provides the fraction of times the rain gauge and/or SRP data were correctly predicted and PC provides a quantitative measure of the accuracy of the detected rainfall [61]. The equations (1)–(4) associated with each of the indices are as given below.

False Alarm Ratio (FAR)

$$\mathbf{FAR} = \frac{\mathbf{T_F}}{\mathbf{T_H} + \mathbf{T_F}}. \tag{1}$$

Probability of Detection (POD)

$$\mathbf{POD} = \frac{\mathbf{T_H}}{\mathbf{T_H} + \mathbf{T_M}}. \tag{2}$$

Critical Success Index (CSI)

$$\mathbf{CSI} = \frac{\mathbf{T_H}}{\mathbf{T_H} + \mathbf{T_M} + \mathbf{T_F}}. \tag{3}$$

Proportion Correct (PC)

$$\mathbf{PC} = \frac{\mathbf{T_H} + \mathbf{T_C}}{\mathbf{T_H} + \mathbf{T_M} + \mathbf{T_F} + \mathbf{T_C}}, \tag{4}$$

where H represents the rainfall accurately detected (correct hits), **M** represents the rainfall not detected (missed data), **F** represents the precipitation that was falsely detected (false alarms), **C** represents the correct negatives, and $\mathbf{T_F}$, $\mathbf{T_H}$, $\mathbf{T_M}$, and $\mathbf{T_F}$ are the number of times each of the cases occurred, respectively.

The analysis using categorical indices was done for 1 mm/day and 10 mm/day thresholds representing light and heavy rainfall, respectively. The threshold to represent light and heavy rainfall was decided based on Table 3 [62] and the contingency table used for the analysis is shown in Table 4.

If the rainfall threshold $= x$ mm, the categorical indices were calculated based on the following conditions.

2.3.2. Continuous Evaluation Indices. Continuous evaluation indices were used to evaluate the performance of SRP with respect to RG data. To find out the extent to which these two datasets agree with each other (correlation), Pearson's correlation coefficient (CC) was used. To determine the absolute average magnitude of the error between the two datasets, Mean Absolute Error (MAE) was used. Root Mean Square Error (RMSE) which gives more relevance to the larger errors when compared with MAE was used again to determine the absolute average error magnitude. To determine the degree of overall underestimation or overestimation, Relative Bias (RB) was used [61, 64, 65]. Finally, to measure the goodness of fit between the observed and SREs which was initially developed by Gupta et al. [66] and further modified by Kling et al. [67], Kling Gupta Efficiency (KGE) was used. KGE is less sensitive to extreme rainfall and therefore can interpret the overall fitness of rainfall having different intensities [68]. Equations (5)–(9) associated with each of the indices are as given below.

Table 3: Rainfall intensities and thresholds.

Intensity class of rainfall	Rainfall thresholds per day
No/tiny rainfall	$P < 1$ mm
Light rainfall	1 mm $\le P < 2$ mm
Low moderate rainfall	2 mm $\le P < 5$ mm
High moderate rainfall	5 mm $\le P < 10$ mm
Heavy rainfall	$P \ge 10$ mm

Table 4: Contingency table for categorical indices [63].

Satellite events (x in mm)	Observation event (x in mm)	
	Yes ($P \ge x$)	No ($P < x$)
Yes ($P \ge x$)	Hits	False alarms
No ($P < x$)	Misses	Correct negatives

Pearson's correlation coefficient (CC)

$$\mathbf{CC} = \frac{\sum_{\mathbf{i}=1}^{\mathbf{n}} (\mathbf{G_i} - \overline{G})(\mathbf{S_i} - \overline{\mathbf{S}})}{\sqrt{\sum_{\mathbf{i}=1}^{\mathbf{n}} (\mathbf{G_i} - \overline{\mathbf{G}})^2} \mathbf{x} \sqrt{\sum_{i=1}^{n} (\mathbf{S_i} - \overline{\mathbf{S}})^2}}. \tag{5}$$

Mean Absolute Error (MAE)

$$\mathbf{MAE} = \frac{1}{\mathbf{n}} \sum_{\mathbf{i}=1}^{\mathbf{n}} |\mathbf{S_i} - \mathbf{G_i}|. \tag{6}$$

Root Mean Square Error (RMSE)

$$\mathbf{RMSE} = \sqrt{\frac{1}{\mathbf{n}} \sum_{\mathbf{i}=1}^{\mathbf{n}} (\mathbf{S_i} - \mathbf{G_i})^2}. \tag{7}$$

Relative Bias (RB)

$$\mathbf{RB} = \frac{(1/\mathbf{n}) \sum_{\mathbf{i}=1}^{\mathbf{n}} (\mathbf{S_i} - \mathbf{G_i})}{\overline{\overline{\mathbf{G}}}} \times 100\%. \tag{8}$$

Kling Gupta Efficiency.

$$\mathbf{KGE} = 1 - \sqrt{(\mathbf{CC} - 1)^2 + \left(\frac{\mathbf{S_d}}{\mathbf{G_d}}\right)^2 + \left(\frac{\overline{\mathbf{S}}}{\overline{\overline{\mathbf{G}}}}\right)^2}, \tag{9}$$

where $\mathbf{S_i}$, $\overline{\mathbf{S}}$, and $\mathbf{S_d}$ (*for SRP data*) are the $\mathbf{i^{th}}$ station, mean values of SRP data, and standard deviation of SRP data, respectively, $\mathbf{G_i}$, $\overline{\mathbf{G}}$, and $\mathbf{G_d}$ (*for Rain gauge data*) are the $\mathbf{i^{th}}$ station, mean values of gauge data, and standard deviation of gauge data, respectively, and **n** represents the total number of data considered.

The strength range of correlation coefficient (CC) was defined as interpreted by The Political Sciences Department at Quinnipiac University [69] and is given in Table 5.

Table 5: Correlation coefficient interpretation.

Correlation coefficient		Interpretation
+1	−1	Perfect
+0.9 to +0.7	−0.9 to −0.7	Very strong
+0.6 to +0.4	−0.6 to −0.4	Strong
+0.3	−0.3	Moderate
+0.2	−0.2	Weak
+0.1	−0.1	Negligible
0	0	None

2.3.3. Nonparametric Tests. To analyze the presence of trends in the RG and SRP data, the Mann-Kendall (MK) test was used. The Theil's & Sens Slope Estimator was used to quantify the trends.

Mann-Kendall (MK) Test. The MK test [70, 71] provides the significance of the trends that are observed in the rain gauge and satellite data. MK Test uses the hypothesis of H_0 for no trend scenario and H_1 when a trend is present in the datasets. The Mann-Kendall Statistic S is given by the following equation:

$$\mathbf{S} = \sum_{\mathbf{i}=1}^{\mathbf{n}-1} \sum_{\mathbf{j}=\mathbf{i}+1}^{\mathbf{n}} \operatorname{sgn}\left(\mathbf{x_j} - \mathbf{x_i}\right), \tag{10}$$

where $\operatorname{sgn}(\mathbf{x_j} - \mathbf{x_i}) = \begin{cases} +1 & > (\mathbf{x_j} - \mathbf{x_i}) \\ 0 & = (\mathbf{x_j} - \mathbf{x_i}) \\ -1 & < (\mathbf{x_j} - \mathbf{x_i}) \end{cases}.$

An increasing trend will be the outcome if S is having a very high positive value and decreasing trend if S is having a very low negative value. To compute the probability associated with the calculated S and the sample size to obtain a significance of the trend, [71] describes a normal approximation test incorporating the Mann-Kendall Statistic, S. A normalized test statistic Z is computed along with the probability associated with the Z value, $f_{(z)}$.

$$\mathbf{Z_c} = \begin{cases} \dfrac{\mathbf{S} - 1}{\sqrt{\mathbf{Var}(\mathbf{S})}}, & \mathbf{S} > 0, \\ 0, & \mathbf{S} = 0, \\ \dfrac{\mathbf{S} + 1}{\sqrt{\mathbf{Var}(\mathbf{S})}}, & \mathbf{S} < 0, \end{cases} \tag{11}$$

where
$\mathbf{Var}(\mathbf{S}) = (\mathbf{n}(\mathbf{n}-1)(2\mathbf{n}+5) - \sum_{\mathbf{i}=1}^{\mathbf{t}} \mathbf{t}_i(\mathbf{i})(\mathbf{i}-1)(2\mathbf{i}+5))/18$,
where $\mathbf{n}$ is the number of datasets, $\mathbf{t}$ is the number of tied groups, and $\mathbf{t_i}$ is the number of datasets in the $\mathbf{i^{th}}$ group.

The probability density function $f_{(z)}$ for a normal distribution with a mean of 0 and a standard deviation of 1 is given by

$$f(z) = \frac{1}{\sqrt{2\pi}} \mathbf{e}^{\mathbf{z}^2/2}. \tag{12}$$

Taking a 95% significance level, the trend will be determined to be decreasing if Z is negative and the probability is greater than 0.95. Similarly, the trend will be determined to be increasing if Z is positive and the probability is greater than 0.95. If the probability is lesser than 0.95, then it was concluded that no trend is present in the datasets [72].

Theil-Sens Slope Estimator Test. To quantify and obtain a magnitude of the trends observed from the MK test, Sens Slope Estimator was used [73]. Since the data sets used correspond to the same time intervals and upon arranging the data sets in ascending order with time, the slope of each time series data pair was calculated using the following equation:

$$\mathbf{Q_k} = \frac{\mathbf{X_j} - \mathbf{X_k}}{\mathbf{j} - \mathbf{i}}, \tag{13}$$

where $j > k$, X corresponds to a data value at a j/k time and $k = 1, 2, \ldots N$

Upon arranging the $\mathbf{N}$ values in ascending order, the median value of the Sens Slope $\mathbf{Q_i}$ will be calculated using the equation given below:

$$\mathbf{Q_i} = \begin{cases} \mathbf{Q}_{(\mathbf{n}+1)/2}, & \textbf{if N is odd}, \\ \dfrac{1}{2}\left(\mathbf{Q}_{\mathbf{n}/2} + \mathbf{Q}_{\mathbf{n}+2/2}\right), & \textbf{if N is even}. \end{cases} \tag{14}$$

3. Results and Discussion

3.1. Detection of Light Rainfall Events. The mean values calculated for each satellite product after the calculation of categorical indices with the 14 rainfall gauge stations yielded the following results. Figure 4 shows the graphical representation of the obtained results. The threshold used for the light rainfall representation was 1 mm/day. The best performance in terms of FAR was shown by TRMM-3B42 and TRMM-3B42RT products with a mean value of 0.40 and a stationwise variation of 0.01 to 0.54 while the worst performance was shown by PERSIANN-CDR with a mean of 0.54 and a stationwise variation of 0.01 and 0.75. With respect to POD and CSI, the best performance was shown by the IMERG rainfall with a mean of 0.79 and 0.47, respectively, as further proved by Moazami et al. [74]. Stationwise distribution of these indices was 0.47 to 0.88 in POD and 0.37 to 0.57 in CSI. The worst performance in POD was represented by TRMM-3B42 with a mean of 0.66, and in CSI, PERSIANN-CCS and CDR showed the worst performances giving a mean of 0.39. TRMM-3B42 performed best in PC with a mean of 0.73 and stationwise variation of 0.41 to 0.84 while PERSIANN-CDR showed the worst performance with a mean value of 0.62 and stationwise variation of 0.55 and 0.68. From the stationwise mean values, IMERG turned out to be the best, having high performances in both the number of times of accurate rainfall detection (POD) and correct rainfall prediction (CSI). PERSIANN-CDR proved to be the worst by performing the worst rainfall detection (PC), the number of times of correct rainfall prediction (CSI), and the highest fraction of false rainfall ratio (FAR). When comparing the categorical indices results with the climatic zone distribution of stations, IMERG stood out in

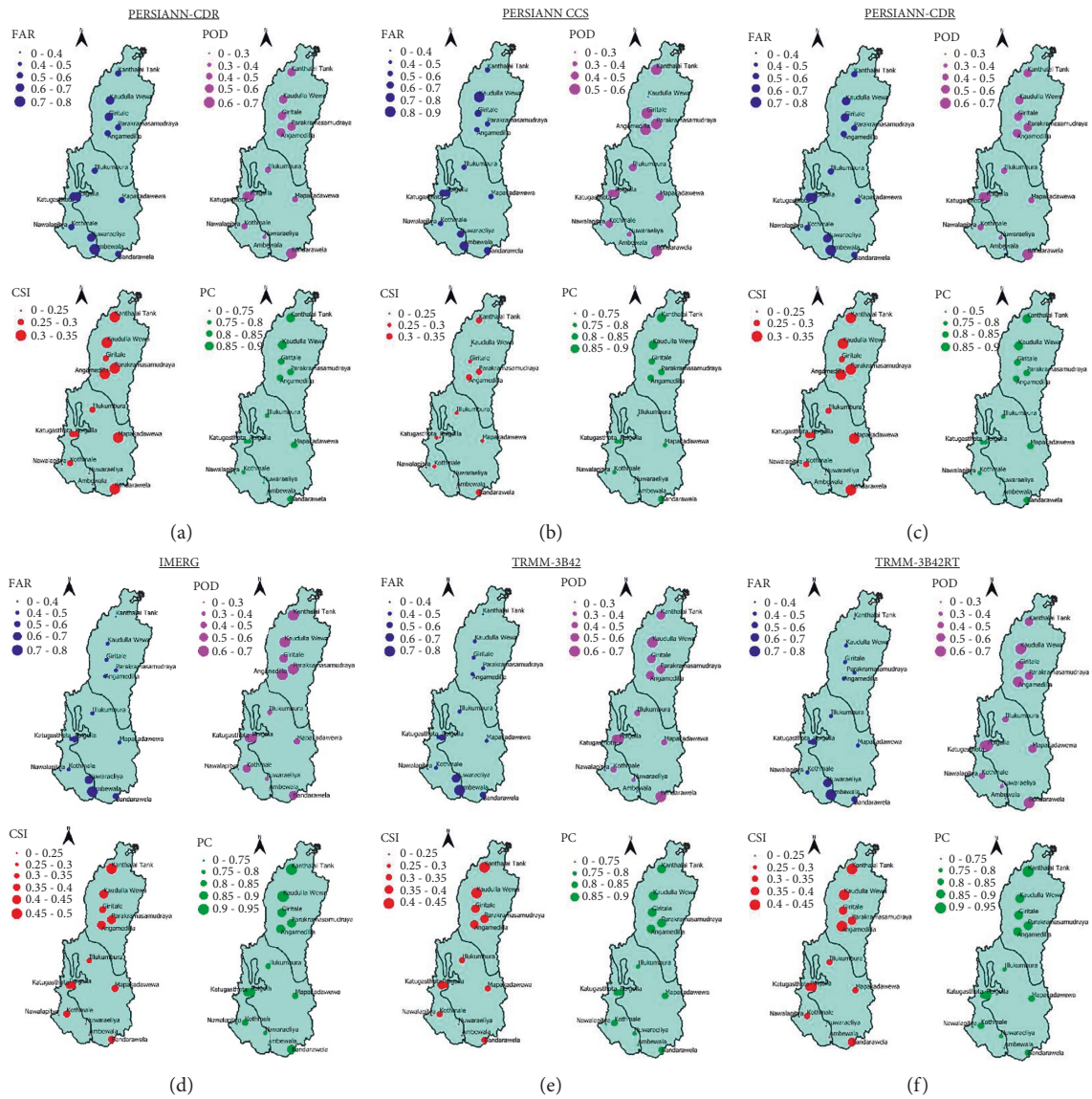


Figure 4: Categorical evaluation for heavy rainfall events. (a) For PERSIANN. (b) For PERSIANN-CCS. (c) For PERSIANN-CDR. (d) For IMERG. (e) For TRMM-3B42. (f) For TRMM-3B42RT.

POD and CSI having higher mean values in both dry and wet zones. TRMM-3B42 also performed well in both FAR and PC in wet and dry zones, respectively. PERSIANN-CDR performed badly in dry zones in both FAR and PC and CSI in the wet zone.

3.2. Detection of Heavy Rainfall Events. The results for categorical indices for each satellite product with a 10 mm/day threshold to represent heavy rainfall are summarized below. Figure 5 shows the graphical representation of the obtained results.

Best performance in FAR, POD, and CSI was shown by IMERG product with a mean value of 0.47, 0.5, and 0.35, respectively, which were in consistent with the range of values obtained in a study done in China incorporating different IMERG products [75]. The stationwise variation was 0.05 to 0.79, 0.18 to 0.64, and 0.12 to 0.46, respectively. The worst performance in FAR was shown by PERSIANN-CCS with a mean of 0.62 and stationwise variation of 0.06 and 0.84. In both POD and CSI, the worst performances were shown by PERSAINN-CDR with a mean value of 0.41 and 0.24, respectively. TRMM-3B42 and 3B42-RT performed best in PC with a mean of 0.81 and stationwise variation of 0.25 and 0.9 while PERSIANN-CCS showed the worst performance with a mean value of 0.77 and stationwise variation of 0.27 and 0.86. From the stationwise mean values, IMERG turned out to be the best, having high performances in having a lower fraction of false rainfall ratio (FAR), the number of times of accurate rainfall detection (POD), and reasonably accurate rainfall prediction (CSI). PERSIANN-CDR proved to be the worst by performing poorly in both the number of times of accurate rainfall detection (POD) and correct rainfall prediction (CSI). Similarly, PERSIANN-

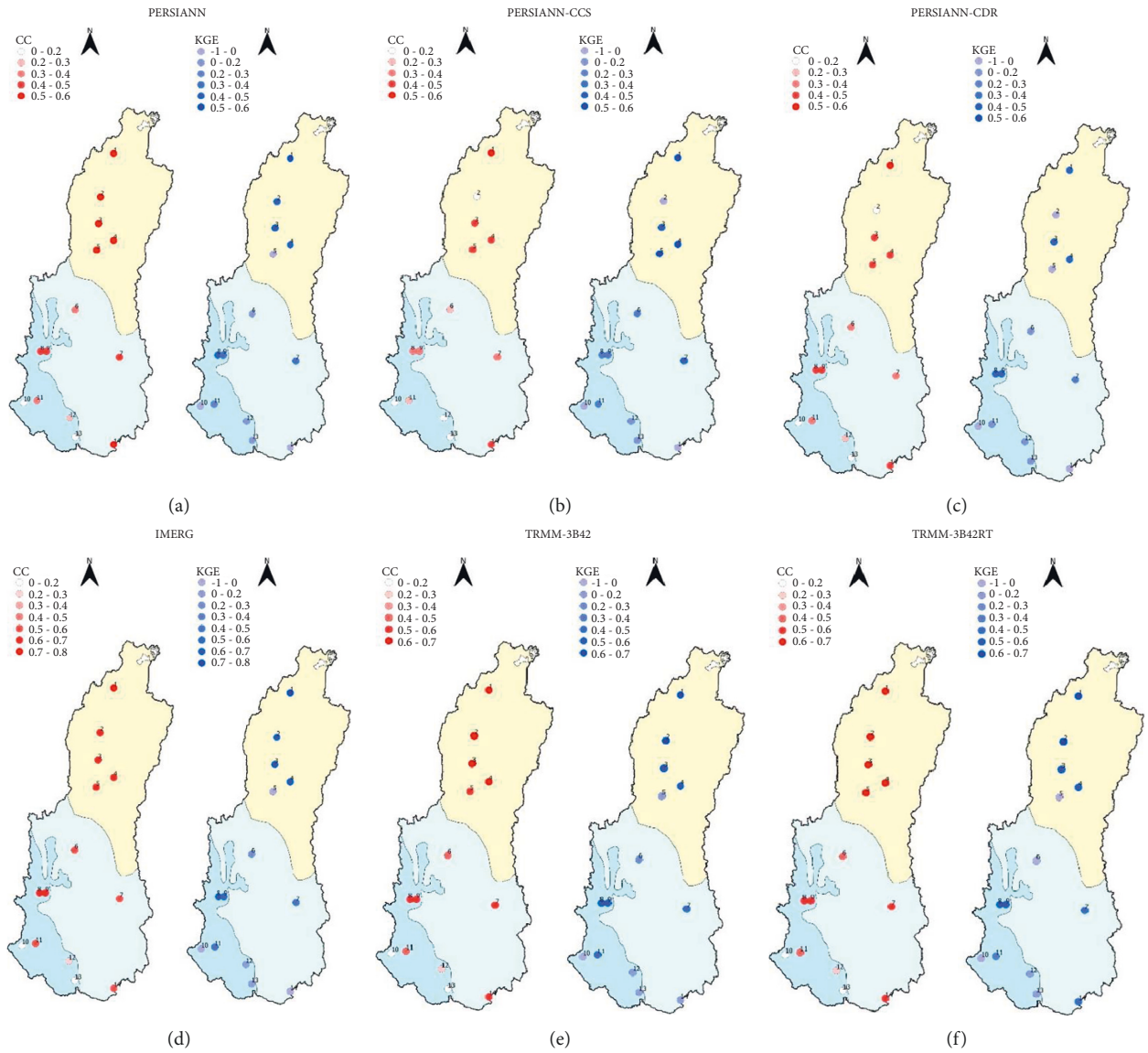


Figure 5: CC and KGE results for daily time scale. (a) For PERSIANN. (b) For PERSIANN-CCS. (c) For PERSIANN-CDR. (d) For IMERG. (e) For TRMM-3B42. (f) For TRMM-3B42RT.

CCS also proved to be the worst by performing relatively badly in both accurate rainfall detection (PC) and the number of times of correct rainfall prediction (CSI).

When comparing the climatic zone distribution of stations with the categorical indices, IMERG stood out in all four indices having higher mean values in dry zones. PERSIANN-CCS performed worst in both POD and CSI in wet zones and in FAR in dry zones. Likewise, PERSIANN-CDR also poorly performed in FAR and POD in dry and wet zones, respectively.

After comparing all the results obtained from light and heavy rainfall detection, it was clear that IMERG outstands the highest with a value of 0.9 in PC which is the highest out of all mean values recorded. It showed this higher accuracy mainly in the dry zones of the river basin. A study carried out on the performance evaluation of SRPs over varying climates and complex topographies has also proved this finding with the IMERG product [76]. PERSIANN-CCS and CDR were found to outstand the least having mean values in the range of 0 to 0.4 in both POD and CSI. This was further proved by Gadouali et al. [60], who performed a study on SRPs in Morocco. His results showed that PERSIANN-CDR was worst at detecting rainfalls. Another study which was performed on the evaluation and comparison of SRPs in Burkina Faso, West Africa, also proved the same with PERSIANN products [77].

3.3. Continuous Evaluation Indices. The correlation between the observed and satellite rainfall data was observed for both daily and monthly time scales. In both time scales, all CCs were in the range of $0 < CC < 1$, showing an overall positive correlation in all products. Out of the six satellite products, the highest median of 0.6 was observed in the IMERG

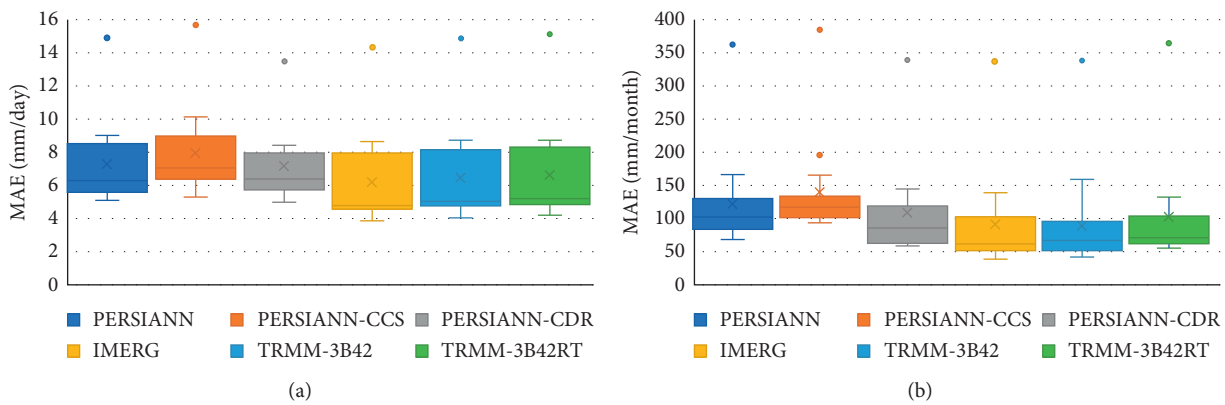


Figure 6: Results obtained from the MAE index. (a) For Daily. (b) For Monthly.

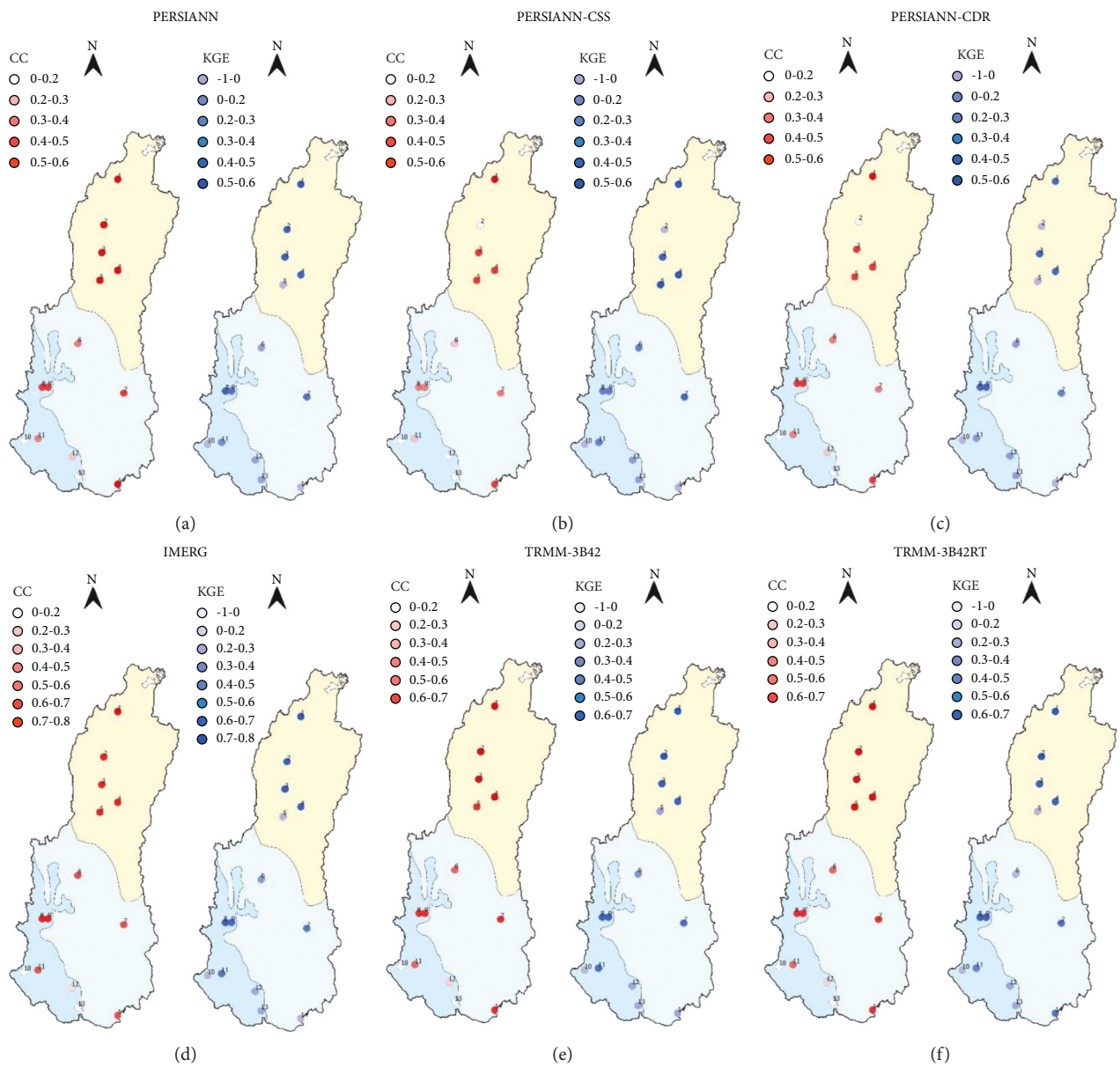


Figure 7: MK and Sen's Slope results for annual analysis. (a) For Observed rainfall data. (b) For PERSIANN. (c) For PERSIANN-CCS. (d) For PERSIANN-CDR. (e) For IMERG. (f) For TRMM-3B42. (g) For TRMM-3B42RT.

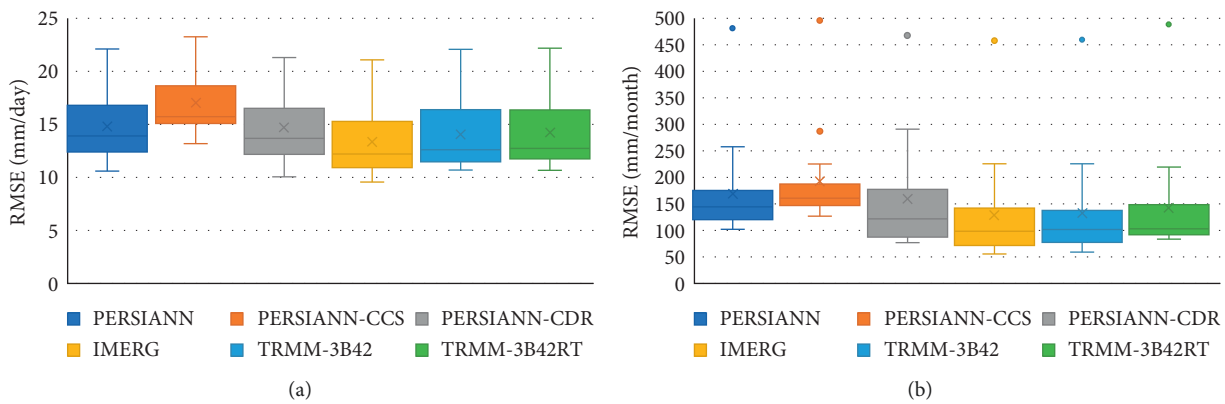


Figure 8: (a) For PERSIANN. (b) For PERSIANN-CCS. (c) For PERSIANN-CDR. (d) For IMERG. (e) For TRMM-3B42. (f) For TRMM-3B42RT.

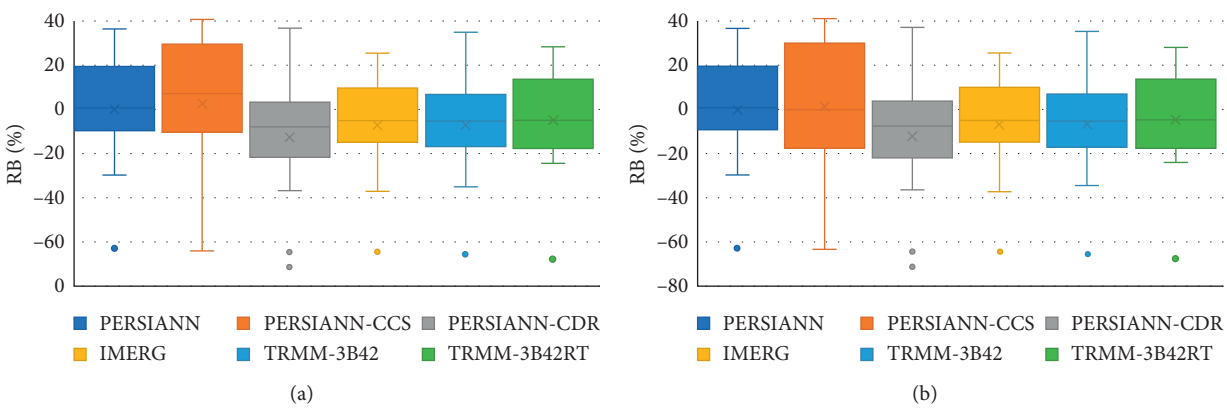


Figure 9: Results obtained from the RMSE index. (a) For Daily. (b) For Monthly.

product which interpreted a strong correlation. All the other products showed relatively poor correlation within the range of $0 < CC < 0.65$. In all 6 products, higher CC was observed mainly in dry zones while the lowest was identified in wet zones of the river basin. Studies conducted by Amjad et al. [76] in Turkey and Alijanian et al. [56] in Iran also showed that lower CC values are obtained for wet regions.

In the monthly time scale analysis, the highest median of 0.84 was shown by IMERG and TRMM-3B42 products. The highest positive values in both time scales were recorded from the dry zones of the river basin and the lower values from the wet zones. The order from strong positive to strong negative correlations was from IMERG and TRMM products to PERSIANN products in both time scales.

KGE shows a perfect fit when the value is close to 1 [68]. In both the daily and monthly analysis, IMERG showed the highest median value of 0.47 and 0.69 for daily and monthly scales, respectively. The worst fit of data was exhibited by PERSIANN-CDR on each time scale. All products showed relatively poor performance for daily data with all values less than 0.75. Most of the higher KGE values were observed in the dry zone and lower values were seen in the wet zones on the daily time scale. The goodness of fit within datasets was best with IMERG followed by TRMM products and PERSIANN products where the worst fit was observed. However, a better performance was observed by monthly values when compared with daily values.

Overall, both good correlation (CC) and fit between the datasets (KGE) were shown by the IMERG product further being proved by Adane et al. [78] from a similar study done in Northeastern Ethiopia. Poor correlation was mainly showcased by PERSIANN-CDR product. This poor performance can be a result of cloud-top IR observations as this product is mainly based on [15]. The graphical representation of the zonewise distribution of the results for the daily time scale from CC and KGE is shown in Figure 6 and the results of the same in monthly time scale are provided in Figure 7.

Both daily and monthly analyses with MAE, RMSE, and RB showed similar behavior (Figures 8–10, respectively). Out of the 6 products, the highest MAE was indicated by PERSIANN-CCS with a median of 7.09 mm/day and 110.01 mm/month in both daily and monthly analyses. On the other hand, IMERG turned out to have the lowest median MAE of 4.79 mm/day and 66.27 mm/month. The highest RMSE was reported by PERSIANN-CCS (median of 15.26 mm/day and 157.53 mm/month) and the lowest RMSE was from IMERG (median of 12.03 mm/day and 96.76 mm/month). All SRPs showed higher errors (MAE and RMSE) in

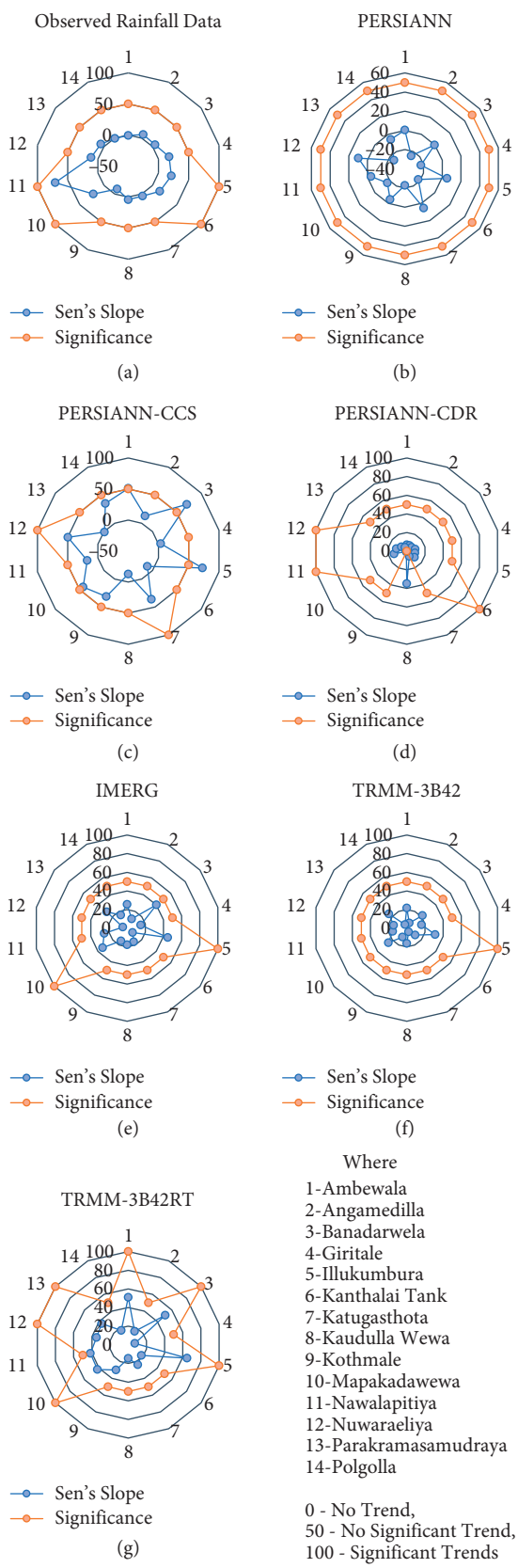


Figure 10: Results obtained from the RB index. (a) For Daily. (b) For Monthly.

the intermediate zone and lower errors in the dry zone of the river basin. This was further confirmed by a study done in Pakistan which stated that higher RMSE is prominent in regions having higher average annual rainfall [79]. The highest underestimation was shown by PERSIANN-CDR while the highest overestimation was from PERSIANN-CCS with medians of −7.55% and 7.79%, respectively. However, in monthly analysis, the highest underestimations were observed in PERSIANN-CDR and the highest overestimations were in PERSIANN product. All products showed higher underestimations in the intermediate zone and higher overestimations in the dry zone of the basin in both analyses.

Overall, it was noted that all SRPs are better at capturing rainfalls accurately with very less errors in the dry zone of the basin while the errors increased when moving from wet to the intermediate zone. IMERG was the best with very less errors in capturing accurate rainfalls when intercompared with the RG data. However, IMERG causes underestimations of rainfall data (median RB of −4.83% in both daily and monthly time scales). Comparatively, given that the IMERG product is set aside, TRMM products also showed better performance and PERSIANN products showed relatively the worst performances in all indices. The better performances of IMERG and TRMM products were further proved by Anjum et al. [80], who found that these products perform well with reference to the RG data at a monthly time scale. Also, overestimations were prominent only in PERISANN products while all other products produced underestimations.

From the performance shown in all the continuous evaluation indices, all products cannot efficiently reproduce the temporal variability in the observed rainfall gauge data on both daily and monthly time scales. Among the six evaluated products, IMERG followed by TRMM products performed well and PERSIANN products performed the worst in each index. Yang et al. [81] had yielded results similar to the present study. Their findings also showed that IMERG V06 and TRMM-3B42 V07 were in the best overall agreement with RG data in all temporal scales. Likewise, the categorical evaluation of these products is similar to what was observed from continuous evaluation indices. IMERG stood out here as well. The final run IMERG product, which is the product used in this study, has proved its higher performance in the Middle East as well [82]. It was also clear that these products possess significant errors which cannot be ignored when using them in hydrological applications. An approach for error correction would be required beforehand. Figures 8–10 show the box plot representation of the results for MAE, RMSE, and RB in both daily and monthly analyses, respectively.

3.4. Trend Analysis. Mann-Kendall trend test was used to identify any significant trends in the observed data and SRPs. Then, to quantify the trends obtained, Sen's slope estimator was used on the datasets.

Table 6: Monthly analysis of MK test and Sen's slope estimator significant trend results.

Station	Month	Kendall's Tau	*P* value (two-tailed)	Sen's slope (mm/month)	Remark
			Observed data		
Ambewala	August	0.305	0.011	5.747	Significant Trend
Bandarawela	April	0.244	0.042	3.842	Significant Trend
Giritale	February	0.254	0.036	2.067	Significant Trend
	December	0.257	0.033	8.471	Significant Trend
Kanthalai Tank	September	−0.321	0.010	−3.873	Significant Trend
Katugasthota	July	−0.255	0.048	−2.273	Significant Trend
Mapakadawewa	April	0.239	0.047	3.213	Significant Trend
	May	0.270	0.025	1.896	Significant Trend
Nawalapitiya	April	0.265	0.048	4.008	Significant Trend
	December	0.379	0.004	7.584	Significant Trend
Nuwaraeliya	March	0.347	0.007	2.732	Significant Trend
	November	0.265	0.040	4.300	Significant Trend
Polgolla	July	−0.290	0.022	−2.088	Significant Trend
			PERSIANN		
Station	Month	Kendall's Tau	*P* value (two-tailed)	Sen's slope (mm/month)	Remark
Kanthalai Tank	April	−0.427	0.011	−10.628	Significant Trend
Polgolla	May	0.357	0.033	12.650	Significant Trend
			PERSIANN-CCS		
Station	Month	Kendall's Tau	*P* value (two-tailed)	Sen's slope (mm/month)	Remark
Ambewala	October	0.452	0.032	21.659	Significant Trend
Bandarawela	October	0.487	0.020	29.042	Significant Trend
	December	0.487	0.020	12.667	Significant Trend
Kanthalai Tank	April	−0.400	0.031	−14.964	Significant Trend
Katugasthota	August	0.353	0.048	13.143	Significant Trend
Kothmale	February	0.477	0.010	5.667	Significant Trend
	May	0.500	0.007	21.639	Significant Trend
Nawalapitiya	May	0.498	0.010	21.667	Significant Trend
Nuwaraeliya	May	0.353	0.048	16.500	Significant Trend
Parakramasamudraya	April	−0.377	0.043	−12.300	Significant Trend
Polgolla	May	0.494	0.008	20.545	Significant Trend
			PERSIANN-CDR		
Station	Month	Kendall's Tau	*P* value (two-tailed)	Sen's slope (mm/month)	Remark
Ambewala	November	0.303	0.013	3.576	Significant Trend
	December	0.277	0.024	3.456	Significant Trend
Angamedilla	June	−0.301	0.011	−1.516	Significant Trend
	September	−0.324	0.006	−2.439	Significant Trend
	November	0.261	0.028	3.659	Significant Trend
Bandarawela	September	−0.248	0.039	−2.827	Significant Trend
	November	0.303	0.013	3.576	Significant Trend
	December	0.280	0.022	3.495	Significant Trend
Giritale	June	−0.241	0.045	−1.002	Significant Trend
	September	−0.298	0.013	−2.417	Significant Trend
	December	0.291	0.016	5.288	Significant Trend
Illukumbura	June	−0.337	0.006	−2.375	Significant Trend
	November	0.299	0.014	4.012	Significant Trend
	December	0.364	0.003	5.905	Significant Trend
Kanthalai Tank	February	0.259	0.041	2.085	Significant Trend
	February	0.287	0.017	7.254	Significant Trend
Kaudulla Wewa	June	0.480	<0.0001	2.343	Significant Trend
	July	0.383	0.001	2.264	Significant Trend
	August	0.340	0.005	2.558	Significant Trend
Kothmale	September	0.257	0.033	1.792	Significant Trend
	November	0.269	0.028	3.435	Significant Trend
Mapakadawewa	June	−0.265	0.030	−1.874	Significant Trend
	July	−0.242	0.047	−1.979	Significant Trend
	September	−0.280	0.022	−2.184	Significant Trend
	November	0.280	0.022	3.393	Significant Trend
	December	0.299	0.014	5.105	Significant Trend

Table 6: Continued.

Station	Month	Kendall's Tau	*P* value (two-tailed)	Sen's slope (mm/month)	Remark
			Observed data		
Parakramasamudraya	June	−0.250	0.034	−1.209	Significant Trend
	July	−0.247	0.037	−1.463	Significant Trend
	September	−0.331	0.005	−2.288	Significant Trend
Polgolla	March	0.255	0.048	3.045	Significant Trend
	July	−0.320	0.011	−1.998	Significant Trend
	December	0.264	0.040	3.727	Significant Trend
			IMERG		
Station	Month	Kendall's Tau	*P* value (two-tailed)	Sen's slope (mm/month)	Remark
Angamedilla	May	0.368	0.039	6.495	Significant Trend
Kothmale	May	0.412	0.017	7.677	Significant Trend
Nawalapitiya	May	0.353	0.048	6.468	Significant Trend
Nuwaraeliya	January	−0.392	0.019	−6.331	Significant Trend
Polgolla	May	0.438	0.011	6.920	Significant Trend
			TRMM-3B42		
Station	Month	Kendall's Tau	*P* value (two-tailed)	Sen's slope (mm/month)	Remark
Ambewala	October	0.373	0.031	6.604	Significant Trend
Angamedilla	June	−0.389	0.016	−4.339	Significant Trend
Bandarawela	October	0.373	0.031	6.604	Significant Trend
Giritale	June	−0.413	0.014	−2.258	Significant Trend
Illukumbura	June	−0.386	0.025	−4.779	Significant Trend
Kanthalai Tank	August	0.352	0.025	3.946	Significant Trend
Katugasthota	January	−0.316	0.040	−5.565	Significant Trend
Kaudulla Wewa	June	−0.387	0.018	−2.105	Significant Trend
Nuwaraeliya	January	−0.307	0.045	−4.132	Significant Trend
Parakramasamudraya	March	0.324	0.040	2.976	Significant Trend
Polgolla	May	0.314	0.046	6.070	Significant Trend
			TRMM-3B42RT		
Station	Month	Kendall's Tau	*P* value (two-tailed)	Sen's slope (mm/month)	Remark
Angamedilla	April	−0.415	0.013	−9.658	Significant Trend
	June	−0.373	0.031	−3.975	Significant Trend
Giritale	June	−0.439	0.015	−2.791	Significant Trend
Illukumbura	August	0.383	0.038	4.155	Significant Trend
Kanthalai Tank	April	−0.450	0.007	−4.802	Significant Trend
Kaudulla Wewa	April	−0.359	0.037	−7.386	Significant Trend
	June	−0.409	0.019	−2.120	Significant Trend
Kothmale	May	0.450	0.007	9.803	Significant Trend
Mapakadawewa	August	0.500	0.007	7.515	Significant Trend
Nawalapitiya	May	0.386	0.025	7.939	Significant Trend
Nuwaraeliya	February	0.357	0.033	6.566	Significant Trend
Parakramasamudraya	April	−0.427	0.011	−6.898	Significant Trend
	August	0.345	0.039	5.001	Significant Trend
Polgolla	May	0.450	0.007	12.377	Significant Trend

3.4.1. Mann-Kendall Test. The MK test that was performed for monthly, seasonal, and annual time scales of the observed rainfall gauge data showed significant increasing trends mostly in the intermediate zone of the river basin. Increasing trends in the seasonal analysis were mostly in the second intermonsoon season. In the annual and monthly analysis, the wet zone also showed significant increasing trends similar to the finding of Pawar and Rathnayake [83]. IMERG product agreed more with the trends observed in the rainfall gauge data in the monthly and annual time scales. TRMM-3B42 showed significant increasing trends during the second intermonsoon in the intermediate zone. In TRMM-3B42 and PERSIANN-CDR, from the monthly analysis, negative significant trends (dry zone) were mostly observed. In all three time scales, PERSIANN-CCS showed increasing trends in the wet zone. However, PERSIANN-CDR and TRMM-3B42RT showed mixed results while PERSIANN showed no significant trends in all three time scales.

3.4.2. Sen's Slope Estimator Test. Sen's slope for the trends obtained from the MK test indicated a slope >1.5 mm/month in the monthly analysis for the observed data. For the annual and seasonal analysis, the slopes were >20 mm/year and >9 mm/season, respectively. In the monthly analysis, IMERG indicated an increasing trend of >6 mm/month. High intensity increasing trends of >30 mm/year in the annual time scale were observed in IMERG and both TRMM

Table 7: Seasonal analysis of MK test and Sen's slope estimator significant trend results.

Station	Season	Kendall's Tau	*P* value (two-tailed)	Sen's slope (mm/season)	Remark
			Observed Data		
Illukumbura	SIM	0.273	**0.026**	9.941	Significant Trend
Mapakadawewa	SIM	0.265	**0.030**	11.116	Significant Trend
Nawalapitiya	FIM	0.328	**0.014**	7.904	Significant Trend
	NEM	0.512	**<0.0001**	15.453	Significant Trend
Nuwaraeliya	SIM	0.297	**0.021**	7.056	Significant Trend
			PERSIANN		
Station	Season	Kendall's Tau	*P* value (two-tailed)	Sen's slope (mm/season)	Remark
Kaudulla Wewa	FIM	−0.359	**0.037**	−14.233	Significant Trend
Parakramasamudraya	FIM	−0.345	**0.039**	−9.185	Significant Trend
			PERSIANN-CCS		
Station	Season	Kendall's Tau	*P* value (two-tailed)	Sen's alope (mm/season)	Remark
Illukumbura	NEM	0.462	**0.028**	24.222	Significant Trend
Katugasthota	SWM	0.412	**0.021**	38.830	Significant Trend
Kothmale	SWM	0.567	**0.002**	35.542	Significant Trend
Mapakadawewa	SIM	0.436	**0.038**	27.250	Significant Trend
Parakramasamudraya	FIM	−0.367	**0.048**	−17.655	Significant Trend
Polgolla	SWM	0.483	**0.01**	47.500	Significant Trend
			PERSIANN-CDR		
Station	Season	Kendall's Tau	*P* value (two-tailed)	Sen's slope (mm/season)	Remark
Ambewala	SWM	−0.246	**0.044**	−4.993	Significant Trend
Angamedilla	SIM	0.257	**0.030**	5.424	Significant Trend
	SWM	−0.287	**0.015**	−5.377	Significant Trend
Giritale	SIM	0.291	**0.016**	5.651	Significant Trend
	SWM	−0.262	**0.029**	−4.768	Significant Trend
Illukumbura	NEM	0.367	**0.010**	8.769	Significant Trend
	SIM	0.299	**0.014**	6.994	Significant Trend
	SWM	−0.428	**0.010**	−9.474	No Trend
Kanthalai Tank	NEM	0.290	**0.020**	9.922	Significant Trend
Kaudulla Wewa	NEM	0.375	**0.010**	13.545	Significant Trend
	SIM	0.355	**0.010**	8.153	Significant Trend
	SWM	0.550	**<0.0001**	11.862	Significant Trend
Mapakadawewa	SIM	0.242	**0.047**	5.474	Significant Trend
	SWM	−0.430	**0.010**	−8.966	Significant Trend
Nawalapitiya	NEM	0.266	**0.043**	7.347	Significant Trend
Parakramasamudraya	SIM	0.247	**0.037**	4.514	Significant Trend
	SWM	−0.247	**0.037**	−4.964	Significant Trend
Polgolla	NEM	0.260	**0.040**	8.054	Significant Trend
			IMERG		
Station	Season	Kendall's Tau	*P* value (two-tailed)	Sen's slope (mm/season)	Remark
Katugasthota	SWM	0.368	**0.023**	10.622	Significant Trend
Kothmale	SWM	0.404	**0.016**	14.497	Significant Trend
Nawalapitiya	SWM	0.346	**0.045**	12.768	Significant Trend
Polgolla	SWM	0.357	**0.033**	12.283	Significant Trend
			TRMM-3B42		
Station	Season	Kendall's Tau	*P* value (two-tailed)	Sen's slope (mm/season)	Remark
Ambewala	FIM	0.399	**0.021**	9.543	Significant Trend
	SIM	0.359	**0.037**	12.752	Significant Trend
Bandarawela	SIM	0.359	**0.037**	12.752	Significant Trend
Illukumbura	FIM	0.359	**0.037**	7.277	Significant Trend
	SIM	0.359	**0.037**	16.176	Significant Trend
Mapakadawewa	SIM	0.359	**0.037**	12.259	Significant Trend
			TRMM-3B42RT		
Station	Season	Kendall's Tau	*P* value (two-tailed)	Sen's slope (mm/season)	Remark
Ambewala	SIM	0.417	**0.024**	19.470	Significant Trend
Bandarawela	SIM	0.417	**0.024**	19.470	Significant Trend
Illukumbura	NEM	0.483	**0.009**	39.989	Significant Trend
Kanthalai Tank	FIM	−0.368	**0.028**	−4.435	Significant Trend

Table 7: Continued.

Station	Season	Kendall's Tau	*P* value (two-tailed)	Sen's slope (mm/season)	Remark
		Observed Data			
Kothmale	SWM	0.415	**0.013**	20.820	Significant Trend
Nawalapitiya	SWM	0.346	**0.045**	18.558	Significant Trend
Parakramasamudraya	SWM	0.415	**0.013**	18.072	Significant Trend

products. TRMM-3B42 showed increasing trends of >12 mm/season during the second intermonsoon season in the intermediate zone. The increasing trends shown by PERSIANN-CCS were >10 mm/year in the annual analysis. Figure 10 demonstrates the MK and Sen's slope estimator results obtained for the annual time scale. The results of the significant trends in the monthly and seasonal analysis are attached in Tables 6 and 7.

From the nonparametric analysis, the results concluded that the IMERG product agrees more with RG data in monthly and annual time scales while TRMM-3B42 agrees more in the seasonal analysis. These findings agree with a recent study by Hussein et al. [84]. Therefore, depending on these results, a careful choice of products for the different zones in the river basin is required. This showed that even in the same river basin, products behave differently with trend patterns depending on the climatic seasons and zones of the river basin.

4. Conclusions

In this research study, six SRPs (PERSIANN, PERSINN-CCS, PERSIANN-CDR, IMERG, TRMM-3B42, and TRMM-3B42RT) were evaluated against rainfall gauge data. Observed data at 14 locations spatially distributed in the three climatic zones of the MRB, Sri Lanka, were selected. Four categorical indices, FAR, POD, CSI, and PC, were used to determine the accuracy of rainfall detection and prediction of satellite products during light and heavy rainfall. IMERG product showed better performance while PERSIANN-CDR showed the worst performance in detecting and predicting rainfall during both these rainfall events. The accuracy of the SRPs was also evaluated using five continuous evaluation indices, CC, RMSE, MAE, RB, and KGE. Among the six evaluated products, in general, IMERG showed better performance while PERSIANN products showed poor performance. However, IMERG also caused underestimations of rainfall data. From the performance shown in all the continuous evaluation indices, all products cannot efficiently reproduce the temporal variability of RG data in both daily and monthly time scales. From the nonparametric tests done on the two datasets SRP and observed rainfall data to identify any significant trends, it was concluded that the IMERG product agrees more with observed rainfall data in monthly and annual time scales while TRMM-3B42 agrees more on the seasonal scale. In all three time scales, PERSIANN-CCS showed increasing trends in the wet zone. However, altogether, PERSIANN-CDR and TRMM-3B42RT showed mixed results while PERSIANN showed no significant trends in all three time scales. It can be concluded that a careful selection of global precipitation products is required prior to using them in any application. This showed that even in the same river basin, products behave differently with trend patterns depending on the climatic seasons and zones of the river basin. It was also clear that these products possess significant errors which cannot be ignored when using them in hydrological applications. However, in places of scarce rainfall data in the MRB, IMERG product proved to be a better choice overall. This study being the first of a kind that incorporated SRPs and observed rainfall data for the Mahaweli River Basin, Sri Lanka, is an immense contribution to many stakeholders and the research community as this basin carries significant importance to the country. This research study was subjected to limitations with the use of six research-based SRPs and 14 rain gauge stations. Therefore, further, it is recommended to carry out studies incorporating near-real-time products with more rain gauge stations to avoid point-pixel errors.

Data Availability

The climatic data used in this research study are available upon request for research purposes.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

The authors of this manuscript are grateful to all SRP communities for making the precipitation data freely available to the international research community. In addition, the authors would like to thank Mr. Jayanga Samarasinghe from Sri Lanka Institute of Information Technology, Sri Lanka, and Ms. Pavithra K. Baddewela from Transport and Logistic Management, University of Moratuwa, Sri Lanka, for their support in writing this paper. This research received no external funding.

References

[1] M. D. A. Qizi, *Water Resources and their Use in The National Economy*, Builders Of The Future, 2021.

[2] R. Loritz, M. Hrachowitz, M. Neuper, and E. Zehe, "The role and value of distributed precipitation data in hydrological models," *Hydrology and Earth System Sciences*, vol. 25, no. 1, pp. 147–167, 2021.

[3] R. Roca, L. V. Alexander, G. Potter et al., "FROGS: a daily $1^\circ \times 1^\circ$ gridded precipitation database of rain gauge, satellite and reanalysis products," *Earth System Science Data*, vol. 11, no. 3, pp. 1017–1035, 2019.

[4] E. C. Barrett, "Precipitation measurement by satellites: towards community algorithms," *Advances in Space Research*, vol. 13, no. 5, pp. 119–136, 1993.

[5] L. Brocca, P. Filippucci, S. Hahn et al., "SM2RAIN-ASCAT (2007-2018): global daily satellite rainfall data from ASCAT

soil moisture observations," *Earth System Science Data*, vol. 11, no. 4, pp. 1583–1601, 2019.

[6] M. B. Gunathilake, Y. V. Amaratunga, A. Perera, C. Karunanayake, A. S. Gunathilake, and U. Rathnayake, "Statistical evaluation and hydrologic simulation capacity of different satellite-based precipitation products (SbPPs) in the Upper Nan River Basin, Northern Thailand," *Journal of Hydrology: Regional Studies*, vol. 32, Article ID 100743, 2020.

[7] G. J. Ciach, "Local random errors in tipping-bucket rain gauge measurements," *Journal of Atmospheric and Oceanic Technology*, vol. 20, no. 5, pp. 752–759, 2003.

[8] G. Villarini, P. Mandapaka, W. Krajewski, and R. Moore, "Rainfall and sampling uncertainties: a rain gauge perspective," *Journal of Geophysical Research*, vol. 113, 2008.

[9] C. Karunanayake, M. B. Gunathilake, and U. Rathnayake, "Inflow forecast of iranamadu reservoir, Sri lanka, under projected climate scenarios using artificial neural networks," *Applied Computational Intelligence and Soft Computing*, vol. 2020, Article ID 8821627, 11 pages, 2020.

[10] K. E. Trenberth, A. Dai, R. M. Rasmussen, and D. B. Parsons, "The changing character of precipitation," *Bulletin of the American Meteorological Society*, vol. 84, no. 9, pp. 1205–1218, 2003.

[11] J. Shi, F. Yuan, C. Shi et al., "Statistical evaluation of the latest GPM-Era IMERG and GSMaP satellite precipitation products in the Yellow River source region," *Water*, vol. 12, no. 4, p. 1006, 2020.

[12] G. J. Huffman, D. T. Bolvin, E. J. Nelkin et al., "The TRMM Multisatellite Precipitation Analysis (TMPA): quasi-global, multiyear, combined-sensor precipitation estimates at fine scales," *Journal of Hydrometeorology*, vol. 8, no. 1, pp. 38–55, 2007.

[13] R. Joyce, J. Janowiak, P. Arkin, and P. Xie, "CMORPH: a method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution," *Journal of Hydrometeorology*, vol. 5, no. 3, pp. 487–503, 2004.

[14] S. Sorooshian, K.-L. Hsu, X. Gao, H. V. Gupta, B. Imam, and D. Braithwaite, "Evaluation of PERSIANN system satellite-based estimates of tropical rainfall," *Bulletin of the American Meteorological Society*, vol. 81, no. 9, pp. 2035–2046, 2000.

[15] H. E. Beck, A. I. J. M. Van Dijk, V. Levizzani et al., "MSWEP: 3-hourly 0.25° global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data," *Hydrology and Earth System Sciences*, vol. 21, no. 1, pp. 589–615, 2017.

[16] M. M. Bitew and M. Gebremichael, "Evaluation through independent measurements: complex terrain and humid tropical region in Ethiopia," *Satellite Rainfall Applications for Surface Hydrology*, Spriger, Berlin, Germany, pp. 205–214, 2009.

[17] E. N. Anagnostou, V. Maggioni, E. I. Nikolopoulos, T. Meskele, F. Hossain, and A. Papadopoulos, "Benchmarking high-resolution global satellite rainfall products to radar and rain-gauge rainfall estimates," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 4, pp. 1667–1683, 2010.

[18] S. Jiang, L. Ren, B. Yong, X. Yang, and L. Shi, "Evaluation of high-resolution satellite precipitation products with surface rain gauge observations from laohahe basin in Northern China," *Water Science and Engineering*, vol. 3, 2010.

[19] A. Mondal, V. Lakshmi, and H. Hashemi, "Intercomparison of trend analysis of multisatellite monthly precipitation products and gauge measurements for river basins of India," *Journal of Hydrology*, vol. 565, pp. 779–790, 2018.

[20] M. B. Gunathilake, M. Zamri, T. P. Alagiywanna et al., "Hydrologic utility of satellite-based and gauge-based gridded precipitation products in the Huai Bang Sai Watershed of Northeastern Thailand," *Hydrology*, vol. 8, no. 4–165, pp. 1–21, 2021.

[21] P. Xie and P. A. Arkin, "Analyses of global monthly precipitation using gauge observations, satellite estimates, and numerical model predictions," *Journal of Climate*, vol. 9, no. 4, pp. 840–858, 1996.

[22] A. V. Mehta and S. Yang, "Precipitation climatology over Mediterranean Basin from ten years of TRMM measurements," *Advances in Geosciences*, vol. 17, pp. 87–91, 2008.

[23] A. Milewski, R. Elkadiri, and M. Durham, "Assessment and comparison of TMPA satellite precipitation products in varying climatic and topographic regimes in Morocco," *Remote Sensing*, vol. 7, no. 5, pp. 5697–5717, 2015.

[24] K. Tong, F. Su, D. Yang, and Z. Hao, "Evaluation of satellite precipitation retrievals and their potential utilities in hydrologic modeling over the Tibetan Plateau," *Journal of Hydrology*, vol. 519, pp. 423–437, 2014.

[25] B. Kumar, K. C. Patra, and V. Lakshmi, "Daily rainfall statistics of TRMM and CMORPH: a case for trans-boundary Gandak river basin," *Journal of Earth System Science*, vol. 125, no. 5, pp. 919–934, 2016.

[26] S. Prakash, A. K. Mitra, D. S. Pai, and A. AghaKouchak, "From TRMM to GPM: how well can heavy rainfall be detected from space?" *Advances in Water Resources*, vol. 88, pp. 1–7, 2016.

[27] F. Fatkhuroyan, T. Wati, A. Sukmana, and R. Kurniawan, "Validation of satellite daily rainfall estimates over Indonesia," *Forum Geografi*, vol. 32, no. 2, pp. 170–180, 2018.

[28] S. Yoshimoto and G. Amarnath, "Applications of satellite-based rainfall estimates in flood inundation modeling-A case study in Mundeni Aru river basin, Sri Lanka," *Remote Sensing*, vol. 9, no. 10, p. 998, 2017.

[29] M. B. Gunathilake, C. Karunanayake, A. S. Gunathilake et al., "Hydrological models and artificial neural networks (ANNs) to simulate streamflow in a tropical catchment of Sri Lanka," *Applied Computational Intelligence and Soft Computing*, vol. 2021, Article ID 6683389, 9 pages, 2021.

[30] M. B. Gunathilake, T. Senerath, T. Senerath, and U. Rathnayake, "Artificial neural network based PERSIANN data sets in evaluation of hydrologic utility of precipitation estimations in a tropical watershed of Sri Lanka," *AIMS Geosciences*, vol. 7, no. 3, pp. 478–489, 2021.

[31] L. Zubair, "El Niño-southern oscillation influences on the Mahaweli streamflow in Sri Lanka," *International Journal of Climatology*, vol. 23, no. 1, pp. 91–102, 2003.

[32] S. Diyabalanage, S. Abekoon, I. Watanabe et al., "Has irrigated water from Mahaweli river contributed to the kidney disease of uncertain etiology in the dry zone of Sri Lanka?" *Environmental Geochemistry and Health*, vol. 38, no. 3, pp. 679–690, 2016.

[33] H. Selvarajah, T. Koike, M. Rasmy et al., "Development of an integrated approach for the assessment of climate change impacts on the hydro-meteorological characteristics of the Mahaweli river basin, Sri Lanka," *Water*, vol. 13, no. 9, p. 1218, 2021.

[34] S. Withanachchi, S. Köpke, C. Withanachchi, R. Pathiranage, and A. Ploeger, "Water resource management in dry zonal paddy cultivation in Mahaweli River basin, Sri Lanka: an analysis of spatial and temporal climate change impacts and traditional knowledge," *Climate*, vol. 2, no. 4, pp. 329–354, 2014.

[35] S. Shelton and Z. Lin, "Streamflow variability in Mahaweli River basin of Sri Lanka during 1990-2014 and its possible mechanisms," *Water*, vol. 11, no. 12, p. 2485, 2019.

[36] Z. Lin and S. Shelton, "Interdecadal change of drought characteristics in Mahaweli river basin of Sri Lanka and the associated atmospheric circulation difference," *Frontiers of Earth Science*, vol. 8, 2020.

[37] N. Imbulana, S. Gunawardana, S. Shrestha, and A. Datta, "Projections of extreme precipitation events under climate change scenarios in Mahaweli river basin of Sri Lanka," *Current Science*, vol. 114, no. 07, pp. 1495–1509, 2018.

[38] T. R. W. S. Dhanapala and H. A. H. Jayasena, "Hydrologic model for selected sub-catchments in the Mahaweli basin of Sri Lanka," in *Proceedings of the National Conference on Status and Future Direction of Water Research in Sri Lanka*, vol. 17, International Water Management Institute, Colombo, Sri Lanka, 1998.

[39] U. Rathnayake, S. B. Weerakoon, K. D. W. Nandalal, and U. Rathnayake, "Flood modeling in the Mahaweli river reach from kothmale to polgolla," *International conference on Mitigation of the risk of natural hazards, Peradeniya, Sri Lanka*, vol. 8, 2007.

[40] E. P. N. Udayakumara and U. A. D. P. Gunawardena, *Reducing Siltation and Increasing Hydropower Generation from the Rantambe Reservoir, Sri Lanka*, South Asian Network for Development and Environmental Economics (SANDEE), Paper, Kathmandu, Nepal, 2016.

[41] B. A. R. H. Dias, E. P. N. Udayakumara, J. M. C. K. Jayawardana, S. Malavipathirana, and D. A. T. W. K. Dissanayake, "Assessment of soil erosion in uma oya catchment, Sri Lanka," *Journal of Environmental Professionals Sri Lanka*, vol. 8, no. 1, 2019.

[42] H. Gunatilake and G. Vieth, "Estimation of on-site cost of soil erosion: a comparison of replacement and productivity change methods," *Journal of Soil and Water Conservation*, vol. 55, pp. 197–204, 2000.

[43] T. Hewawasam, "Effect of land use in the upper Mahaweli catchment area on erosion, landslides and siltation in hydropower reservoirs of Sri Lanka," *Journal of the National Science Foundation of Sri Lanka*, vol. 38, no. 1, 2010.

[44] J. M. C. K. Jayawardana, W. D. T. M. Gunawardana, E. P. N. Udayakumara, and M. Westbrooke, "Land use impacts on river health of Uma Oya, Sri Lanka: implications of spatial scales," *Environmental Monitoring and Assessment*, vol. 189, no. 4, p. 192, 2017.

[45] R. N. N. Weerasinghe, J. M. C. K. Jayawardana, and E. P. N. Udayakumara, "Quantitative assessment of soil erosion and its association with river health in Uma Oya watershed in Sri Lanka," in *Proceedings of the ISAE 2016*, pp. 137–140, Niagara-On-The-Lake, Canada, October 2016.

[46] P. Aravinna, N. Priyantha, A. Pitawala, and S. K. Yatigammana, "Use pattern of pesticides and their predicted mobility into shallow groundwater and surface water bodies of paddy lands in Mahaweli river basin in Sri Lanka," *Journal of Environmental Science and Health, Part B*, vol. 52, no. 1, pp. 37–47, 2016.

[47] C. B. Dissanayake and S. V. R. Weerasooriya, "The environmental chemistry of Mahaweli river, Sri Lanka," *International Journal of Environmental Studies*, vol. 28, pp. 207–223, 1986.

[48] B. A. Malmgren, R. Hulugalla, Y. Hayashi, and T. Mikami, "Precipitation trends in Sri Lanka since the 1870s and relationships to El Niño-southern oscillation," *International Journal of Climatology*, vol. 23, no. 10, pp. 1235–1252, 2003.

[49] J. M. R. S. Bandara, H. V. P. Wijewardena, Y. M. A. Y. Bandara, R. G. P. T. Jayasooriya, and H. Rajapaksha, "Pollution of river mahaweli and farmlands under irrigation by cadmium from agricultural inputs leading to a chronic renal failure epidemic among farmers in NCP, Sri Lanka," *Environmental Geochemistry and Health*, vol. 33, no. 5, pp. 439–453, 2011.

[50] Y. Mu, T. Biggs, and S. S. P. Shen, "Satellite-based precipitation estimates using a dense rain gauge network over the southwestern brazilian amazon: implication for identifying trends in dry season rainfall," *Atmospheric Research*, vol. 261, Article ID 105741, 2021.

[51] V. H. D. M. Paca, G. Espinoza-Dávalos, D. Moreira, and G. Comair, "Variability of trends in precipitation across the amazon river basin determined from the CHIRPS precipitation product and from station records," *Water*, vol. 12, no. 5, p. 1244, 2020.

[52] R. B. L. Cavalcante, D. B. d. S. Ferreira, P. R. M. Pontes, R. G. Tedeschi, C. P. W. da Costa, and E. B. de Souza, "Evaluation of extreme rainfall indices from CHIRPS precipitation estimates over the Brazilian Amazonia," *Atmospheric Research*, vol. 238, Article ID 104879, 2020.

[53] G. J. Huffman, R. F. Adler, D. T. Bolvin, and E. J. Nelkin, "The TRMM multi-satellite precipitation analysis (TMPA)," in *Satellite Rainfall Applications for Surface Hydrology*, pp. 3–22, Springer, Dordrecht, Netherlands, 2010.

[54] G. J. Huffman, D. T. Bolvin, D. Braithwaite et al., *Algorithm Theoretical Basis Document (Atbd) Version 06 Nasa Global Precipitation Measurement (Gpm) Integrated Multi-Satellite Retrievals for Gpm (Imerg)*, National Aeronautics and Space Administration, Washington, DC, USA, 2019.

[55] G. J. Huffman, *NASA Global Precipitation Measurement (GPM) Integrated Multi Satellite Retrievals for GPM (IMERG)*, NASA, Washington, DC, USA, 2020.

[56] M. Alijanian, G. R. Rakhshandehroo, A. K. Mishra, and M. Dehghani, "Evaluation of satellite rainfall climatology using CMORPH, PERSIANN-CDR, PERSIANN, TRMM, MSWEP over Iran," *International Journal of Climatology*, vol. 37, no. 14, pp. 4896–4914, 2017.

[57] K.-L. Hsu, A. Behrangi, B. Imam, and S. Sorooshian, "Extreme precipitation estimation using satellite-based PERSIANN-CCS algorithm," in *Satellite Rainfall Applications for Surface Hydrology*, pp. 49–67, Springer, Dordrecht, Netherlands, 2010.

[58] Y. Hong, D. Gochis, J.-t. Cheng, K.-l. Hsu, and S. Sorooshian, "Evaluation of PERSIANN-CCS rainfall measurement using the NAME event rain gauge network," *Journal of Hydrometeorology*, vol. 8, no. 3, pp. 469–482, 2007.

[59] H. Ashouri, K.-L. Hsu, S. Sorooshian et al., "PERSIANN-CDR: daily precipitation climate data record from multi-satellite observations for hydrological and climate studies," *Bulletin of the American Meteorological Society*, vol. 96, no. 1, pp. 69–83, 2015.

[60] F. Gadouali and M. Messouli, "Evaluation of multiple satellite-derived rainfall products over Morocco," *International Journal of Horticultural Science and Technology*, vol. 10, no. 1, pp. 72–89, 2020.

[61] M. N. Anjum, Y. Ding, D. Shangguan, M. W. Ijaz, and S. Zhang, "Evaluation of high-resolution satellite-based real-time and post-real-time precipitation estimates during 2010 extreme flood event in swat river basin, hindukush region," *Advances in Meteorology*, vol. 2016, Article ID 2604980, 8 pages, 2016.

[62] M. S. Nashwan, S. Shahid, and X. Wang, "Assessment of satellite-based precipitation measurement products over the hot desert climate of Egypt," *Remote Sensing*, vol. 11, no. 5, p. 555, 2019.

[63] A. B. Ayub, F. Tangang, L. Juneng, M. L. Tan, and J. X. Chung, "Evaluation of gridded precipitation data sets in Malaysia," *Remote Sensing*, vol. 12, no. 4, p. 613, 2020.

[64] G. Wei, H. Lü, W. T Crow, Y. Zhu, J. Wang, and J. Su, "Evaluation of satellite-based precipitation products from IMERG V04A and V03D, CMORPH and TMPA with gauged rainfall in three climatologic zones in China," *Remote Sensing*, vol. 10, no. 1, p. 30, 2018.

[65] F. Chen and X. Li, "Evaluation of IMERG and TRMM 3B43 monthly precipitation products over mainland China," *Remote Sensing*, vol. 8, no. 6, p. 472, 2016.

[66] H. V. Gupta, H. Kling, K. K. Yilmaz, and G. F. Martinez, "Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling," *Journal of Hydrology*, vol. 377, no. 1-2, pp. 80–91, 2009.

[67] H. Kling, M. Fuchs, and M. Paulin, "Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios," *Journal of Hydrology*, vol. 424-425, pp. 264–277, 2012.

[68] D. E. Radcliffe and R. Mukundan, "PRISMvs. CFSR precipitation data effects on calibration and validation of SWAT models," *Journal of the American Water Resources Association*, vol. 53, no. 1, pp. 89–100, 2017.

[69] H. Akoglu, "User's guide to correlation coefficients," *Turkish journal of emergency medicine*, vol. 18, no. 3, pp. 91–93, 2018.

[70] H. B. Mann, "Nonparametric tests against trend," *Journal of the econometric society*, vol. 13, no. 3, pp. 245–259, 1945.

[71] M. G. Kendall, *Rank Correlation Methods*, Griffin, London, UK, 4th edition, 1948.

[72] HydroGeoLogic Inc, *Final 2004 Annual Groundwater Monitoring Report and Quarterly Groundwater Monitoring Report, Quarter 4, 2004, Operable Unit 1 Fritzsche Army Airfield Fire Drill Area Former fort Ord California*, HydroGeoLogic Inc, Huntsville, AL, USA, 2005.

[73] S. I. Sridhar and A. Raviraj, "Statistical trend analysis of rainfall in Amaravathi river basin using Mann-Kendall test," *Current World Environment*, vol. 12, no. 1, pp. 89–96, 2017.

[74] S. Moazami and M. R. Najafi, "A comprehensive evaluation of GPM-IMERG V06 and MRMS with hourly ground-based precipitation observations across Canada," *Journal of Hydrology*, vol. 594, Article ID 125929, 2021.

[75] C. Yu, D. Hu, Y. Di, and Y. Wang, "Performance evaluation of IMERG precipitation products during typhoon Lekima (2019)," *Journal of Hydrology*, vol. 597, Article ID 126307, 2021.

[76] M. Amjad, M. T. Yilmaz, I. Yucel, and K. K. Yilmaz, "Performance evaluation of satellite- and model-based precipitation products over varying climate and complex topography," *Journal of Hydrology*, vol. 584, Article ID 124707, 2020.

[77] M. Dembélé and S. J. Zwart, "Evaluation and comparison of satellite-based rainfall products in burkina faso, West Africa," *International Journal of Remote Sensing*, vol. 37, no. 17, pp. 3995–4014, 2016.

[78] G. B. Adane, B. A. Hirpa, C.-H. Lim, and W.-K. Lee, "Evaluation and comparison of satellite-derived estimates of rainfall in the diverse climate and terrain of central and northeastern Ethiopia," *Remote Sensing*, vol. 13, no. 7, p. 1275, 2021.

[79] M. Masood, A. S. Shakir, A. H. Azhar, G. Nabi, and fnm Habib-u-Rehman, "Assessment of real time, multi-satellite precipitation products under diverse climatic and topographic conditions," *Asia-Pacific Journal of Atmospheric Sciences*, vol. 56, no. 4, pp. 577–591, 2020.

[80] M. N. Anjum, Y. Ding, D. Shangguan et al., "Performance evaluation of latest integrated multi-satellite retrievals for global precipitation measurement (IMERG) over the northern highlands of Pakistan," *Atmospheric Research*, vol. 205, pp. 134–146, 2018.

[81] M. Yang, Z. Li, M. N. Anjum, and Y. Gao, "Performance evaluation of version 5 (V05) of integrated multi-satellite retrievals for global precipitation measurement (IMERG) over the tianshan Mountains of China," *Water*, vol. 11, no. 6, p. 1139, 2019.

[82] M. T. Mahmoud, M. A. Al-Zahrani, and H. O. Sharif, "Assessment of global precipitation measurement satellite products over Saudi Arabia," *Journal of Hydrology*, vol. 559, pp. 1–12, 2018.

[83] U. Pawar and U. Rathnayake, "Spatiotemporal rainfall variability and trend analysis over mahaweli basin, Sri Lanka," *Arabian Journal of Geosciences*, vol. 15, no. 320, pp. 1–16, 2022.

[84] K. A. Hussein, T. S. Alsumaiti, D. T. Ghebreyesus, H. O. Sharif, and W. Abdalati, "High-resolution spatiotemporal trend analysis of precipitation using satellite-based products over the United Arab Emirates," *Water*, vol. 13, no. 17, p. 2376, 2021.

Hindawi
Advances in Meteorology
Volume 2022, Article ID 6819539, 14 pages
https://doi.org/10.1155/2022/6819539

Research Article

Estimation of Potential Evapotranspiration across Sri Lanka Using a Distributed Dual-Source Evapotranspiration Model under Data Scarcity

Udara Senatilleke,[1] Himasha Abeysiriwardana,[1] Randika K. Makubura,[1] Faisal Anwar,[2] and Upaka Rathnayake[1]

[1]*Department of Civil Engineering, Faculty of Engineering, Sri Lanka Institute of Information Technology, Malabe, Sri Lanka*
[2]*Department of Civil Engineering, Curtin University, GPO Box U1987, Perth, WA 6845, Australia*

Correspondence should be addressed to Upaka Rathnayake; upaka.r@sliit.lk

Received 19 December 2021; Accepted 4 April 2022; Published 23 April 2022

Academic Editor: Marina Baldi

Evapotranspiration estimations are not common in developing countries though most of them have water scarcities for agricultural purposes. Therefore, it is essential to estimate the rates of evapotranspiration based on the available climatic parameters. Proper estimations of evapotranspiration are unavailable to Sri Lanka, even though the country has a significant agricultural contribution to its economy. Therefore, the Shuttleworth–Wallace (S-W) model, a process-based two-source potential evapotranspiration (PET) model, is implemented to simulate the spatiotemporal distribution of PET, evaporation from soil (ETs), and transpiration from vegetation canopy (ETc) across the total landmass of Sri Lanka. The country was divided into a grid with $6km \times 6km$ cells. The meteorological data, including rainfall, temperature, relative humidity, wind speed, net solar radiation, and pan evaporation, for 14 meteorological stations were used in this analysis. They were interpolated using Inverse Distance Weighting (IDW), Universal kriging, and Thiessen polygon methods as appropriate so that the generated thematic layers were fairly closer to reality. Normalized Difference Vegetation Index (NDVI) and soil moisture data were retrieved from publicly available online domains, while the threshold values of vegetation parameters were taken from the literature. Notwithstanding many approximations and uncertainties associated with the input data, the implemented model displayed an adequate ability to capture the spatiotemporal distribution of PET and its components. A comparison between predicted PET and recorded pan evaporations resulted in a root mean square error (RMSE) of 0.75 mm/day. The model showed high sensitivity to Leaf Area Index (LAI). The model revealed that both spatial and temporal distribution of PET is highly correlated with the incoming solar radiation fluxes and affected by the rainfall seasons and cultivation patterns. The model predicted PET values accounted for 80–90% and 40–60% loss of annual mean rainfall, respectively, in the drier and wetter parts of the country. The model predicted a 0.65 ratio of annual transpiration to annual evapotranspiration.

1. Introduction

Evapotranspiration (ET) is an integral part of the global hydrological cycle and regional water budget, as ET represents the loss of the surface and soil water to the atmosphere as water vapor by the combined actions of the two processes: evaporation from the surface water bodies; bare soil and other surfaces that intercept rainwater, and transpiration from plants [1]. The primary driver of ET is solar radiation, which provides the latent heat requirement for water vaporization [2]. In addition, several other factors such as wind speed, humidity, air temperature, soil type, crop type, land use type [2, 3], and rainfall [4] affect the rate of ET. Because of the large number of influencing factors and their heterogeneity over a watershed, accurate estimation of ET is challenging, specifically under data scarcities. Thus, estimation of potential evapotranspiration (PET)—the potential amount of water that could evaporate and transpire from a

vegetated landscape with unlimited water supply to the surface [5], has been mostly utilized instead of actual ET estimation. [1, 6].

A few dozen empirical and semiempirical PET models, such as Penman [2], Thornthwaite [7], Priestley-Taylor [8], Monteith [9], and Shuttleworth and Wallace [10], have been developed over the last few decades. These models are based on different sets of assumptions and inherit various limitations; hence, the results generated are highly inconsistent [6]. Some of the PET models, such as Penman [2] and Priestley-Taylor [8], estimate potential evaporation over water surfaces but do not account for transpiration from vegetation cover. Monteith [9] developed the Penman–Monteith (P-M) model, one of the widely used models to estimate PET, assuming the vegetation canopy as a single uniform cover or 'big-leaf,' thus accounting for the transpiration process. One of the drawbacks of the P-M model is that it neglects sparse vegetation. Several researchers, i.e., Shuttleworth and Wallace (S-W) (1985), Choudhury and Monteith [11], Mo et al. [12], have developed PET models by extending the P-M model by incorporating sparse canopy. These extended models assume two-source (the crop and the substrate soil) schemes and balance the energy exchanged between soil, canopy, airspace between soil and canopy, and the atmosphere above the canopy. It has been found that two-source models better predict PET than "big leaf" models [13, 14]. Shuttleworth and Wallace [10] and Choudhury and Monteith (1988) have used a resistance network and estimated PET as the summation of transpiration from vegetation and evaporation from substrate soil. Mo et al. [12] have modified these two models by incorporating evaporation from intercepted storage.

In addition, the usage of satellite data and remote sensing techniques were used in the estimation of potential evapotranspiration [15–18]. These techniques were highly useful for remote areas with meteorological data scarcities. As many of the models were originally developed for specific regions, the empirical relationships between evaporation and influencing factors may not be necessarily the same for other regions [19]. Further, not all meteorological data required by models have often been measured at all meteorological stations [1, 19]. Under these conditions, it is of considerable interest to evaluate PET models for their applicability and reliability in different regions and climatic zones.

Although many such attempts have been taken worldwide at different regional scales [1, 6, 12, 20], comprehensive studies aiming to evaluate or develop PET models are still scant in Sri Lanka. Sri Lanka is characterized by high temperatures, high humidity, and unevenly distributed (both temporally and spatially) rainfall (Imbulana et al., [21]) and thus, considered mild to semiarid. Earlier studies have shown ET has accounted for 30% to more than 60% loss of total rainfall in semiarid regions [22]. Hence, ET can be considered the major water loss pathway in Sri Lanka's water budget. This emphasizes the need for accurate quantification of ET, especially in the dry zone, the semiarid region of Sri Lanka, for sustainable water management and efficient irrigation. In addition, the meteorological data scarcity (other than rainfall) in the country is high and the available data are expensive. Therefore, a high necessity can be identified to develop accurate models to estimate ET.

Both P-M and S-W models have been widely used [20]. However, considering the limitations in the P-M model, this study employs the S-W model, which has not been applied in the Sri Lankan context. While understanding the research gap, this paper aims to (1) apply the S-W model for simulation of evapotranspiration over the entire country and (2) illustrate the temporal and spatial variations of evapotranspiration over the entire country.

2. Study Area

Sri Lanka (refer to Figure 1) is a tropical island in the Indian Ocean, located between 5°N and 10°N latitudes and 79°E to 82°E longitudes, with a total geographical area of 65,610 km^2, comprising 62,705 km^2 area of land and 2,905 km^2 area of water. These water bodies comprise 103 distinct natural river basins and an extensive network of tanks and reservoirs (about 13,000). Approximately two-thirds of the country's landmass is low lands with elevations less than 100 m above the mean sea level. Highlands, elevations varying from 100 m to 2500 m approximately (highest mountain peak 2525 m), lie in the country's central part (Imbulana et al., [21].

The only precipitation method, rainfall, has an unequal spatial and temporal distribution with a mean annual rainfall of 1861 mm, while the rainfall distribution is governed by the two major monsoon seasons: southwest monsoon (SWM) from March to September and northeast monsoon (NEM) from December to February. The country is divided into three major climatic zones, i.e., the wet zone, intermediate zone, and dry zone, based on the rainfall received and distribution. The wet zone is separated by the 2000 m annual average rainfall isohyet. In the Wet zone, rainfall ranges from 2000 mm to over 5000 mm, with an annual average rainfall of about 2,400 mm. But in the dry zone, the annual average rainfall is about 1450 mm with a minimum of lower than 1000 mm. In addition to two major monsoons and intermonsoon rains, tropical depressions that originate in the Bay of Bengal frequently enter Sri Lanka resulting in extreme rainfall events which sometimes may exceed 500 mm/day. Rainfall by all the methods counts for a mean rainfall of 1861 mm over the country per annum. Nearly 35%–45% of annual rainfall contributes to annual surface runoff. However, in most dry zone river basins, the runoff percentage is less than 35%, with the rest of the rainfall lost as evaporation and groundwater recharge.

Mean annual temperature in lowlands and highlands varies between 26.5 and 28.5 °C and 14.7–17.1 °C, respectively. Pan evaporation values show considerable temporal and spatial variations—varying between 1900 and 795 mm/year, with higher values recorded in the hotter dry zone. The climate of the country is characterized by high relative humidity, generally ranging between 75% and 95%. (Imbulana et al., [21]).

As it was stated in the introduction, the measured meteorological data are expensive in Sri Lanka and also there is a scarcity of measured data. One of the major limitations

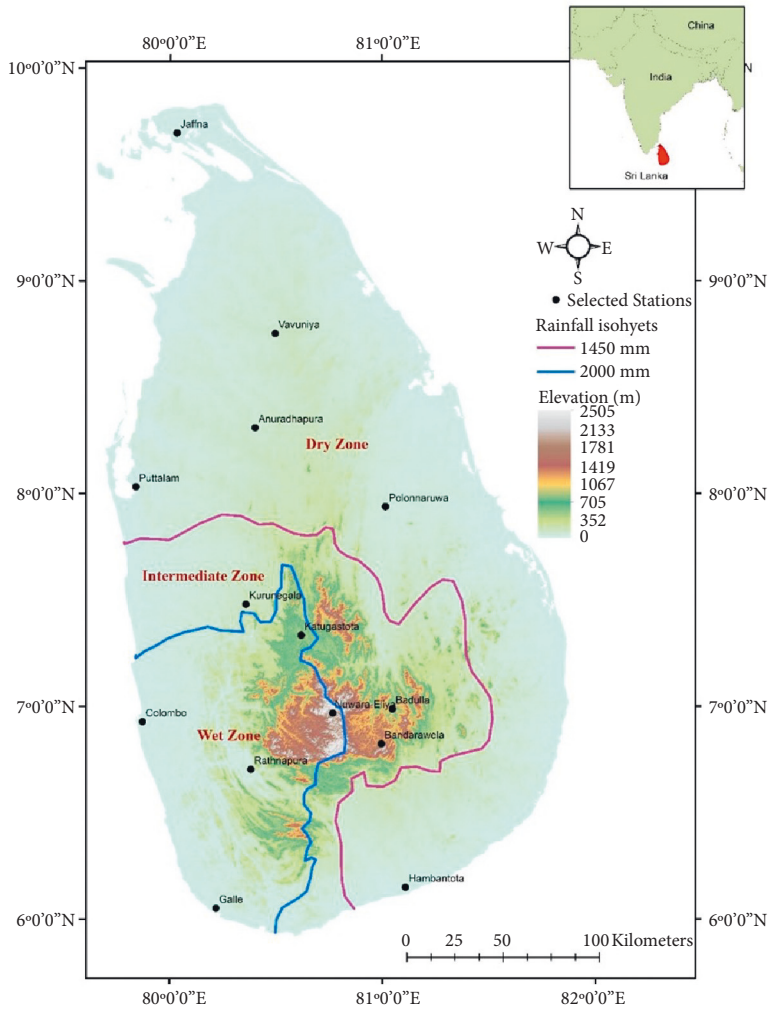


Figure 1: Map of Sri Lanka indicating three major climatic zones and topographic zonation (created by the authors).

was obtaining solar radiation data, as solar radiation is not measured at all 22 main meteorological stations that the Department of Meteorology of Sri Lanka is maintaining across the country. Therefore, only 14 stations were selected (refer to Figure 1) as they mostly cover all climatic (wet, intermediate, and dry) and topographic (hilly areas and lowlands) zones of the country.

3. Methodology

3.1. Evapotranspiration Model

3.1.1. Basic Governing Equations to the Evapotranspiration Model. In S-W model, total PET is computed as the summation of two major evapotranspiration components: soil evapotranspiration and transpiration from the dry canopy [10]:

$$E_t = E_s + E_c, \tag{1}$$

where E_t is the total PET (mm), E_s is the evaporation from soil (mm), and E_c is the transpiration from the dry canopy (mm). E_c and E_s can be expressed as follows [10]:

$$E_c = \frac{1}{\lambda} \cdot \frac{\Delta R_{nc} + \left(\rho C_p D_0 / r_{ac}\right)}{\Delta + \gamma\left(1 + \left(r_c / r_{ac}\right)\right)}, \tag{2}$$

$$E_c = \frac{1}{\lambda} \cdot \frac{\Delta\left(R_{ns} - G\right) + \left(\rho C_p D_0 / r_{as}\right)}{\Delta + \gamma\left(1 + \left(r_s / r_{as}\right)\right)}, \tag{3}$$

where λ is the latent heat of vaporization (MJ kg $^{-1}$), Δ is the slope of saturation vapour pressure curve (kPa ^{o}C $^{-1}$), R_{nc} is the net radiation absorbed by the canopy (MJ m $^{-2}$), ρ is the air density (kg m^{-3}), C_p is the air specific heat at constant pressure (=1.013 × 10^{-3} MJ kg $^{-1}$^{o}C $^{-1}$), D_0 is the water vapour deficit at the canopy height (kPa), r_{ac} is the bulk boundary-layer resistance of the canopy (s m^{-1}), γ is the psychrometric constant (kPa ^{o}C $^{-1}$), r_c is the canopy resistance (s m^{-1}), R_{ns} is the net radiation at the substrate surface (MJ m $^{-2}$), G is the soil heat flux (MJ m $^{-2}$), r_{as} is the aerodynamic resistance between the soil surface and canopy air space (s m^{-1}), and r_s is the soil resistance (s m^{-1}). D_0 can be expressed as follows [10]:

$$D_0 = D + \left\{\Delta\left(R_n - G\right) - (\Delta + \gamma)\lambda E\right\} \frac{r_a}{\rho C_p}, \tag{4}$$

where D is the water vapor deficit at the reference height (kPa), R_n is the net incoming radiation (kPa), and r_a is the aerodynamic resistance between canopy source and reference level (s m^{-1}). By substituting equations (2)–(4) in equation (1) and manipulating, Shuttleworth and Wallace [10] have derived the following equation for the PET model.

$$\lambda E_t = C_c PM_c + C_s PM_s. \quad (5)$$

More explanations of these parameters are given in Appendix equations (A.1)–(A.7). Chow et al. [23] have also expressed some of the parameters, and these are given in Appendix equations (A.8)–(A.12).

3.1.2. Estimation of Net Radiation ($\mathbf{R_n}$). Estimation of radiation is important. The net radiation (R_n) received at the Earth surface can be subdivided (refer to equation (6)) into the net radiation absorbed by the canopy (R_{nc}) and the net radiation absorbed by the soil (R_{ns}) [12].

$$R_n = R_{nc} + R_{ns}. \quad (6)$$

Shuttleworth and Wallace [10] have expressed the relationship of R_n and R_{ns} as in equation (7), where C_r is the extinction coefficient of the vegetation for net radiation. All radiation terms are in MJ m^{-2}. C_r was taken as 0.5 [12, 20]. C_r was taken as 0.7 [10].

$$R_{ns} = R_n \exp(-C_r \mathrm{LAI}). \quad (7)$$

LAI is the Leaf Area Index, a dimensionless parameter that characterises vegetation cover. There are various definitions of LAI; hence, different sets of equations can be found in the literature to estimate LAI. In our study, we utilized the method followed by Zhou et al. [20], which is given in Appendix equations (A.13)–(A.15). In addition, the maximum LAI values for different vegetation types are given in Table 1. Numbers 1–3 are for the tall vegetation, while the others are for the shorter vegetation. The complete table is given in the Appendix as Table 2.

3.1.3. Estimation of Water Vapour Deficit ($\mathbf{D}$). Estimation of water vapor deficit at the reference height (D) can be found in the following equation [20]:

$$D = e_s - e_a, \quad (8)$$

where e_s and e_a are the saturation vapor pressure (kPa) and the ambient vapor pressure (kPa), respectively, and the expressions for e_s and e_a are given in Appendix equations (A.16) and (A.17) [23].

3.1.4. Estimation of Aerodynamic Resistance between Canopy Source and Reference Level ($\mathbf{r_a}$). The aerodynamic resistance between the canopy source and reference level was calculated using the following equation (9), which was adapted by Shuttleworth and Gurney [24]:

Table 1: Maximum LAI values for different vegetation types [20].

Code	Land use type	Classification as per the literature	$\mathbf{LAI_{max}}$
1	Coconut	Evergreen needle leaf forests	5.5
2	Rubber	Evergreen broadleaf forests	7
3	Forest, unclassified	Mixed forests	5.7
4	Homesteads/ garden	Open shrub lands	3
5	Shrublands		3
6	Tea		3
7	Grasslands	Grasslands	1.8
8	Marshy lands	Permanent wetlands	6
9	Chena	Croplands	7
10	Other cultivations		7
11	Paddy		7
12	Urban and built-up	Urban and built-up	0
13	Barren land	Barren or sparsely vegetated	0.3
14	Water bodies	Water bodies	0

$$r_a = \frac{1}{ku_*}\ln\left(\frac{z_a - d_0}{h_c - d_0}\right) + \frac{h_c}{\eta K_h}\left[\exp\left\{\eta\frac{1-\left(Z_0 + d_p\right)}{h_c}\right\} - 1\right], \quad (9)$$

where k is the von Karman's constant ($k = 0.41$), u_* is the friction velocity (m s^{-1}), z_a is the reference height (m), d_0 is the zero-plane displacement of the canopy (m), h_c is the canopy height (m) (refer to Table 2), η is the eddy diffusivity decay constant of the vegetation, K_h is the eddy diffusion coefficient at the top of the canopy (m^2 s^{-1}), Z_0 is the "preferred" roughness length (m), and d_p is the "preferred" zero plane displacement. The equations used to compute each term are given in Appendix equations (A.18)–(A.27).

3.1.5. Estimation of Bulk Boundary-Layer Resistance of the Canopy ($\mathbf{r_{ac}}$). Shuttleworth and Gurney [24] have estimated bulk boundary layer resistance of the canopy by assuming that energy transfer only occurs by molecular diffusion through a laminar layer around leaves and using the following equation, which was used in our model:

$$r_{ac} = \frac{100}{\eta}\cdot\sqrt{\left(\frac{l}{u_h}\right)}\cdot\left[1 - \exp\frac{-\eta}{2}\right]^{-1}\cdot\frac{1}{2LAI}, \quad (10)$$

where l is the canopy characteristic leaf width (m), and u_h is the wind speed at the top of the canopy (m s^{-1}). u_h was computed using equation (11) [23], where u^* is the shear velocity (m s^{-1}), which was calculated assigning wind speed values recorded at the reference height and reference height, respectively, for u_h and h_c. Then u_h was calculated, assigning the respective u^* values. l was calculated using equation (12), where $l_{\max}$ is the maximum leaf width (m) (refer to Table 2).

Table 2: Comparison of averaged annual PET (simulated), pan evaporation, and rainfall at the fourteen selected stations.

Station	Cell number	Climate zone	PET (mm/yr)	Pan evaporation (mm/yr)	Rainfall (mm/yr)	PET/rainfall (%)	PanEvap./rainfall (%)
Jaffna	15	Dry zone	1430.16	1414.45	1297.77	110.2	109.0
Vavuniya	238		1320.01	1252.80	1555.85	84.8	80.5
Anuradhapura	461		1506.76	1250.36	1624.76	92.7	77.0
Puttalam	602		1516.56	1503.50	1224.44	123.9	122.8
Polonnaruwa	690		1671.67	1488.14	1784.15	93.7	83.4
Hamabantota	1786	Intermediate zone	1585.07	1524.47	1075.55	147.4	141.7
Kurunegala	954		1552.39	1416.71	2075.27	74.8	68.3
Badulla	1303		955.34	866.58	1896.46	50.4	45.7
Bandarawela	1412		1095.88	982.44	1650.80	66.4	59.5
Katugastota	1069		1233.44	1143.91	1934.22	63.8	59.1
Nuwara-eliya	1298	Wet zone	968.67	889.84	1846.58	52.5	48.2
Colombo	1319		1041.39	1319.97	2493.04	41.8	52.9
Rathnapura	1473		1025.53	921.47	4072.70	25.2	22.6
Galle	1809		1058.68	963.87	2450.26	43.2	39.3

$$\frac{u_h}{u^*} = \frac{1}{k}\ln\left(\frac{h_c}{z_0}\right), \tag{11}$$

$$l = \begin{cases} l_{\max}, & \text{for perennial vegetation,} \\ l_{\max}[1 - \exp(-0.6LAI)], & \text{for annual vegetation.} \end{cases} \tag{12}$$

3.1.6. Estimation of Aerodynamic Resistance between the Soil Surface and Canopy Air Space ($\mathbf{r_{as}}$). Estimation of the aerodynamic resistance between the soil surface and canopy air space requires complex formulations as r_{as} is affected by many factors. r_{as} can be calculated using (notations were previously defined)

$$r_{as} = \frac{h_c \exp(\eta)}{\eta K_h}\left[\exp\left(\frac{-\eta z_{og}}{h_c}\right) - \exp\left(\frac{-\eta\left(Z_0 + d_p\right)}{h_c}\right)\right]. \tag{13}$$

3.1.7. Estimation of Canopy Resistance ($\mathbf{r_c}$). The canopy resistance was estimated using the Jarvis canopy resistance model [20, 25]. Jarvis model is given by the following equation:

$$r_c = \frac{r_{st\min}}{\mathrm{LAI}_e\left[f(R_n)f(T_k)f(D)f(\theta)\right]}, \tag{14}$$

where $r_{st\min}$ is the minimum stomatal resistance (s m^{-1}) (refer to Table 2) and LAI_e is the effective LAI. Expressions for LAI_e, $f(R_n)$, $f(D)$, $f(T_k)$, and $f(\theta)$ are given in Appendix equations (A.28)–(A.32).

3.1.8. Estimation of Soil Resistance ($\mathbf{r_s}$). Due to the complexities in acquiring data required for accurate assessments of the soil resistance, it was set $r_s = 0sm^{-1}$ at saturation point and $r_s = 200sm^{-1}$ at wilting point, and this was suggested by Shuttleworth and Wallace [10]. Then interpolate between two extremes to estimate r_s at average soil moisture content.

*3.1.9. Estimation of Soil Heat Flux (**G**).* Different methods are available in the literature to estimate G. Many researchers consider G is 30% of the R_n [20]; however, Mo et al. [12] have suggested the following equation, which was used in this study:

$$G = 0.183R_n \exp(-0.299\mathrm{LAI}). \tag{15}$$

3.2. Data Collection and Preprocessing

3.2.1. Topographic Data. Elevation data are required to calculate the atmospheric pressure (equation (A.12)), which is then used to compute ρ (air density-equation (A.10)). Elevation values were extracted from a digital elevation model (DEM) with a spatial resolution of $30m \times 30m$, which was obtained from the Department of Survey, Sri Lanka. The DEM resolution was adjusted in ArcGIS to match the model grid size ($6km \times 6km$) of Figure 2(a).

3.2.2. Land Use and Vegetation Data. Land cover data were obtained from the Department of Survey, Sri Lanka. The raw data obtained were in vector format. The raw data were then converted to raster format with a resolution of $6km \times 6km$. There are altogether 46 land use classes in the original vector file. They were recategorized into ten classes, as shown in Table 1 and Figure 2(b). Further, the vegetation cover was subcategorised as short-tall and annual-perennial (Table 1). Collecting the land cover threshold parameters required by the model, i.e., $\mathrm{LAI}_{\max}$, $l_{\max}$, h_c, F_{cl}, $r_{st\min}$, $\mathrm{NDVI}_{98\%}$, and z_{og} was the greatest challenge, as estimation of said parameters was extremely difficult. Therefore, aforesaid parameters were gathered from Zhou et al. [20] and are given in Table 1.

3.2.3. NDVI. Remotely sensed Landsat images from 2009 to 2019, with a spatial resolution of $30m \times 30m$, were acquired from USGS [26]. A total of 9 Landsat tiles were required to cover entire Sri Lanka. However, due to the high dense cloud cover (>5%), Landsat images from a previous or a subsequent month were utilized for some months. This may result in some errors in the actual condition, thus considered a

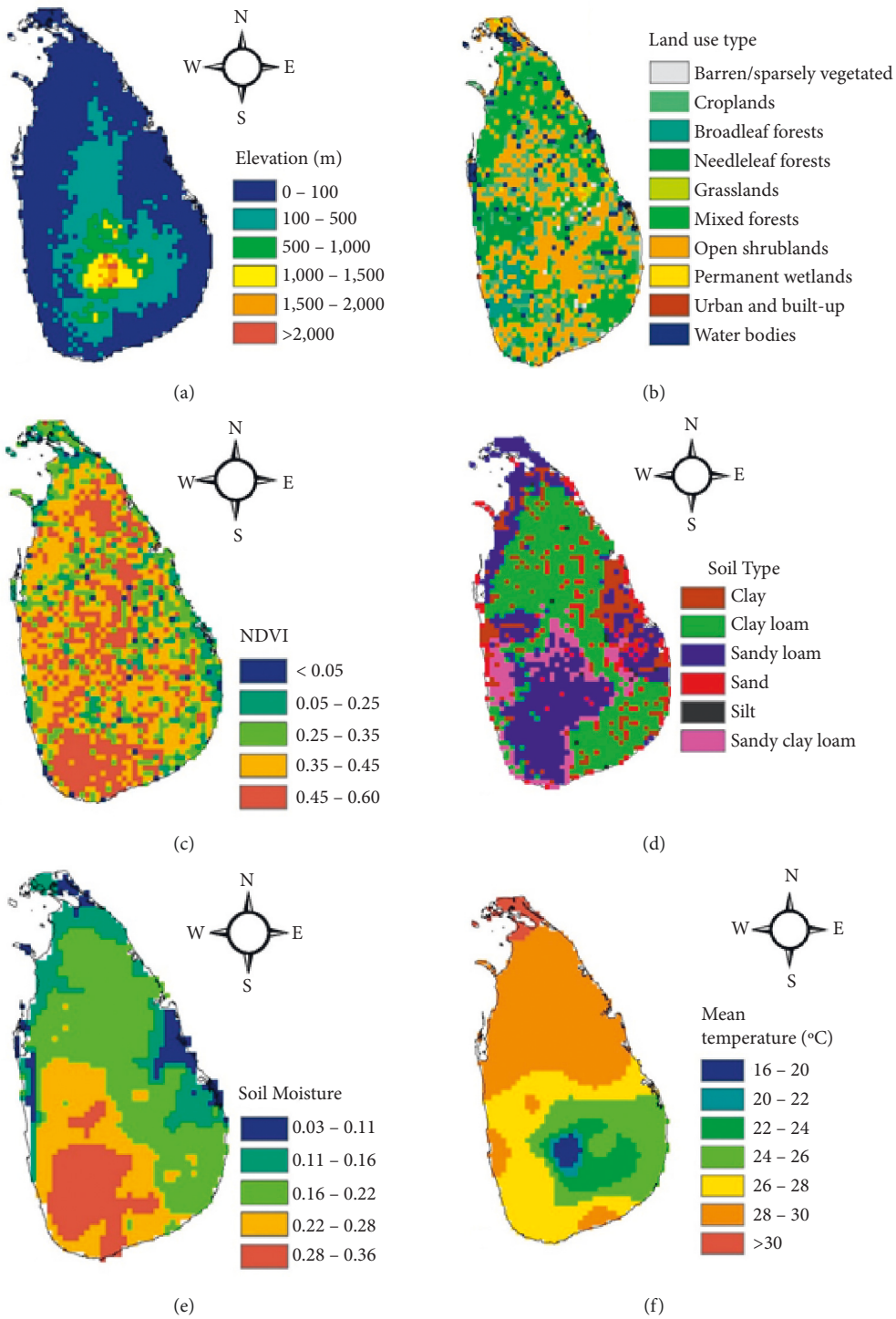


Figure 2: Continued.

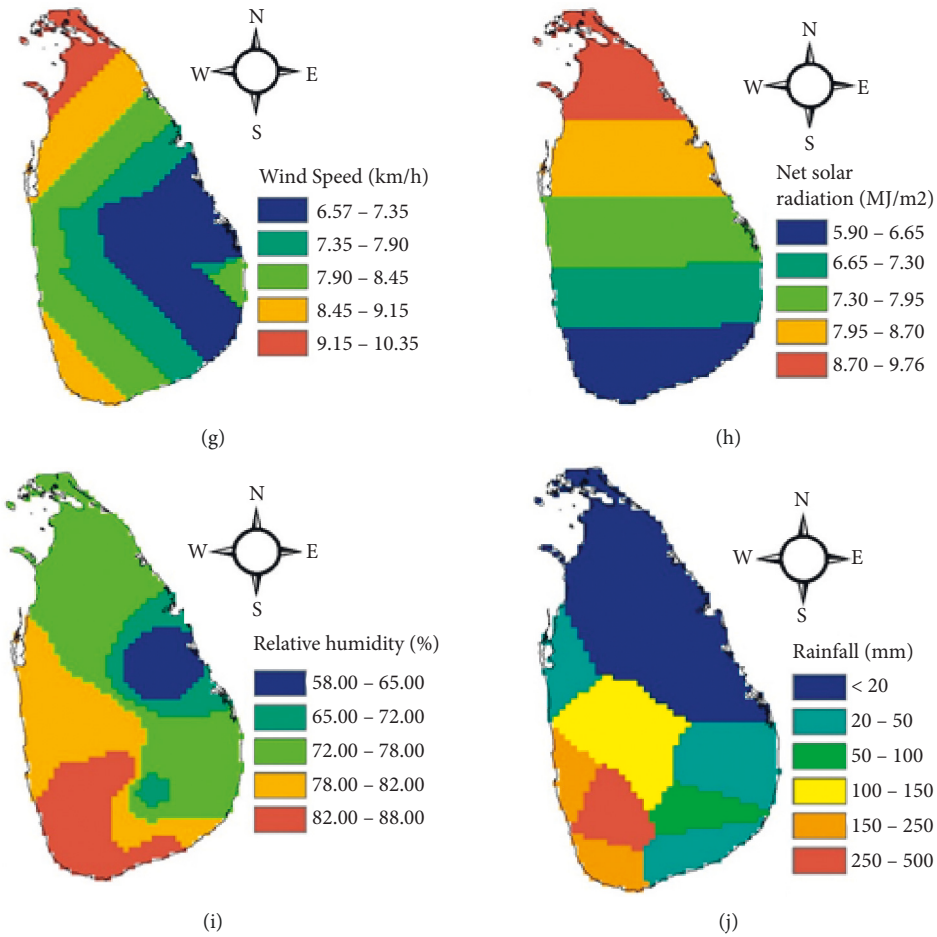


Figure 2: Model input parameters (created by the authors). (a) For elevation. (b) For land-use type. (c) For NDVI. (d) For soil type. (e) For soil moisture. (f) For mean temperature. (g) For wind speed. (h) For average daily net solar radiation. (i) For average relative humidity. (j) For monthly average rainfall.

potential limitation of the study. The NDVI values were computed as the difference between near-infrared (NIR) and red (RED) reflectance divided by their sum (equation (16)) in ArcGIS. Finally, the calculated NDVI layers' resolution was converted to $6km \times 6km$ (Figure 2(c)).

$$\mathrm{NDVI} = \frac{\mathrm{NIR} - \mathrm{Red}}{\mathrm{NIR} + \mathrm{Red}}. \tag{16}$$

3.2.4. Soil Type and Soil Moisture. Sri Lanka soil type map was obtained from the Department of Survey, Sri Lanka. The soil map of Sri Lanka was then recategorized into six soil textural classes based on the classifications of Moormann and Panabokke [27] (Figure 2(d)). Much of the dry zone consists of clay loam, and most of the wet zone contains sandy loam. Soil saturation and wilting point values were assessed based on the soil textural classes. The reclassified soil vector layer was converted to a raster layer with a $6\,\mathrm{km} \times 6\,\mathrm{km}$ grid.

Due to the limitation imposed by the unavailability of field measured soil moisture data in the country, monthly root zone soil moisture values were retrieved from NASA [28] from 2009 to 2019 (Figure 2(e)). When data for a month was missing, data from the previous or subsequent month (depending on the rainfall seasons) were utilized. The study may face limitations incurred by this alternative method. Retrieved data has a spatial resolution of $30m \times 30m$; thus, it was transformed to the model's spatial resolution. Incur.

3.2.5. Meteorological Data. Monthly time series data of maximum and minimum daily temperatures, relative humidity (day and night), daily precipitation, solar net radiation, wind speed, and evaporation from 2009 to 2019 were obtained from the Department of Meteorology of Sri Lanka. The model requires only the temperature, solar radiation, humidity, and wind speed data (Figures 2(f)–2(j)). Average values of maximum and minimum temperatures and day and night relative humidity data were fed into the model. Precipitation and evaporation data were used for trend analysis and model validation. The data was preprocessed in ArcGIS. The spatial distributions of monthly averaged temperature, relative humidity, and evaporation were

generated by Inverse Distance Weighted (IDW) interpolation. The Universal Kriging interpolation method was used to obtain the spatial distribution of solar net radiation and wind speed. The spatial distributions of monthly precipitation were generated by applying the Thiessen polygon method.

3.3. Overall Methodology. Processed data showcased in Figures 2(a)–2(j) were used to simulate the potential evapotranspiration for the whole of Sri Lanka. All the equations were modeled in a Microsoft Excel office package and then extracted to ArcGIS to develop the graphical presentation of potential evapotranspiration over the country. The results were validated by a comparison analysis using recorded pan evapotranspiration and the predicted potential evapotranspiration. RMSE was calculated between the predictions and recorded values.

4. Results and Discussion

4.1. Impact of Data Scarcity on Analysis. The model results of fourteen grid cells (refer to Table 3), in which the selected meteorological stations lie, compared with the recorded pan evaporation at the respective meteorological stations. They could lessen the error resulting from interpolation. The model was not calibrated in this study due to a lack of data and a shorter data period. Coupled with the limitations exerted by NDVI and soil moisture data, further challenges to model calibration were exerted by the uncertain canopy and soil parameters obtained from the literature. However, the root means square error (RMSE) of model prediction to observation, 0.75 mm/day, suggests that the model can predict PET and its component over Sri Lanka with moderate accuracy.

Historical solar radiation measurements are not commonly available at many meteorological stations in Sri Lanka. This reduced the number of meteorological stations we could use for this study. The density of selected meteorological stations is notably lesser in the dry zone. Since there is no station in the South-East (SE) part of the country, observed values at Badulla and Bandarawela seem to influence the interpolation outputs in the SE region, irrespective of the interpolation method. As per the actual situation, in the SE low land, some meteorological parameters such as rainfall, temperature, and humidity show a drastic difference from those at Badulla and Bandarawela. Therefore, the accuracy and precision of the model predictions in the SE region are questionable. However, the model can be applied to the rest of the dry zone and intermediate zone with moderate confidence and the wet zone with high confidence.

4.2. Comparison of Predicted PET, Pan Evaporation, and Precipitation. The model generated PET values are greater than the observed, except at the Colombo station, at which the simulated results are significantly smaller than recorded values by an average of 32%. The disparity between simulated and observed values at Colombo can be attributed to several factors. The model is highly sensitive to LAI, and there is a positive correlation between LAI and simulated results, which has been confirmed by Zhou et al. [20]. Colombo is the capital city of Sri Lanka and is densely populated; thus, the vegetation cover is significantly lower than the rest of the wet zone. Hence, the predicted transpiration from vegetation cover decreases. As a result of the large extent of built-up areas, the evaporation from soil is also limited. Consequently, it is reasonable to assume that pan evaporation values are higher than the actual evapotranspiration in Colombo and the simulated results closely represent the actual evapotranspiration than pan evaporation. These discussion points are showcased in Table 3, which explicitly averaged annual pan evaporation, annual simulated PET, and annual rainfall at the 14 stations which were analyzed.

At Puttalam, Polonnaruwa and Hambantota averaged annual simulated figures and observed results differ only by 4–6 mm. However, at all three locations, predicted values are higher than those observed during the rainy season and lower than those observed during dry months. The highest variations between the average annual figures of simulated and pan evaporation can be seen in the wet zone, which could be due to good vegetation coverage throughout the year. In Jaffna, Puttalam, and Hambantota, generally the hottest areas in Sri Lanka, annual evapotranspiration is higher than the annual precipitation. Deficit water must be coming from the groundwater, as the aquifers in these areas are being abstracted excessively (Imbulana et al., [21]). In the rest of the dry zone, although the evapotranspiration is less than the rainfall, 80–93% of precipitation loses to the atmosphere as evapotranspiration, as per the predictions, whereas observed values indicate 75–85% evaporation loss. The reason for this massive evaporation loss might be due to the extensive network of small to large scale man-made irrigation tanks, which provides a constant supply of water over a large land area. In addition, paddy, the major cultivation in these areas, relies on flood irrigation, keeping the soil saturated and aiding both the evaporation and transpiration processes. In the wet zone, the ratio of PET to rainfall ranges between 40 and 65%, and the higher percentages are in the highlands. The predicted annual PET over the entire country is about 2 280 674 mm, with an average of 1243 mm/year. Approximately 65% of PET is from canopy transpiration.

4.3. Seasonal Changes of PET. Five stations, i.e., Vavuniya, Kurunegala, Bandarawela, Katugastota, and Galle, which give the smallest RMSEs, were selected from the dry zone, intermediate zone lowlands, intermediate zone high lands, wet zone high lands, and wet zone lowlands, respectively, to graphically illustrate the temporal variations of average monthly (1) PET, (2) ETs, (3) ETc, (4) rainfall, and (5) average daily Net radiation (refer to Figure 3).

Figure 3 indicates that the total monthly evapotranspiration follows a pattern similar to the seasonal fluctuations of net radiation. Nevertheless, a few minor deviations correlated with rainfall and cultivation seasons can be observed.

Table 3: Predicted annual averaged PET, ETs, and ETc from different land-use types.

	PET (mm/year)	ETs (mm/year)	ETc (mm/year)
Paddy	1391.65	505.03	886.63
Rubber	1292.72	168.20	1124.52
Coconut	1276.18	206.75	1069.42
Chena (arid cultivation)	1255.75	331.32	924.43
Tea	1199.56	295.68	903.87
Grass lands	663.07	429.49	233.58
Shrub lands	1173.70	285.13	888.57
Home gardens	872.00	300.53	571.47
Urban and built-up areas	405.46	405.46	

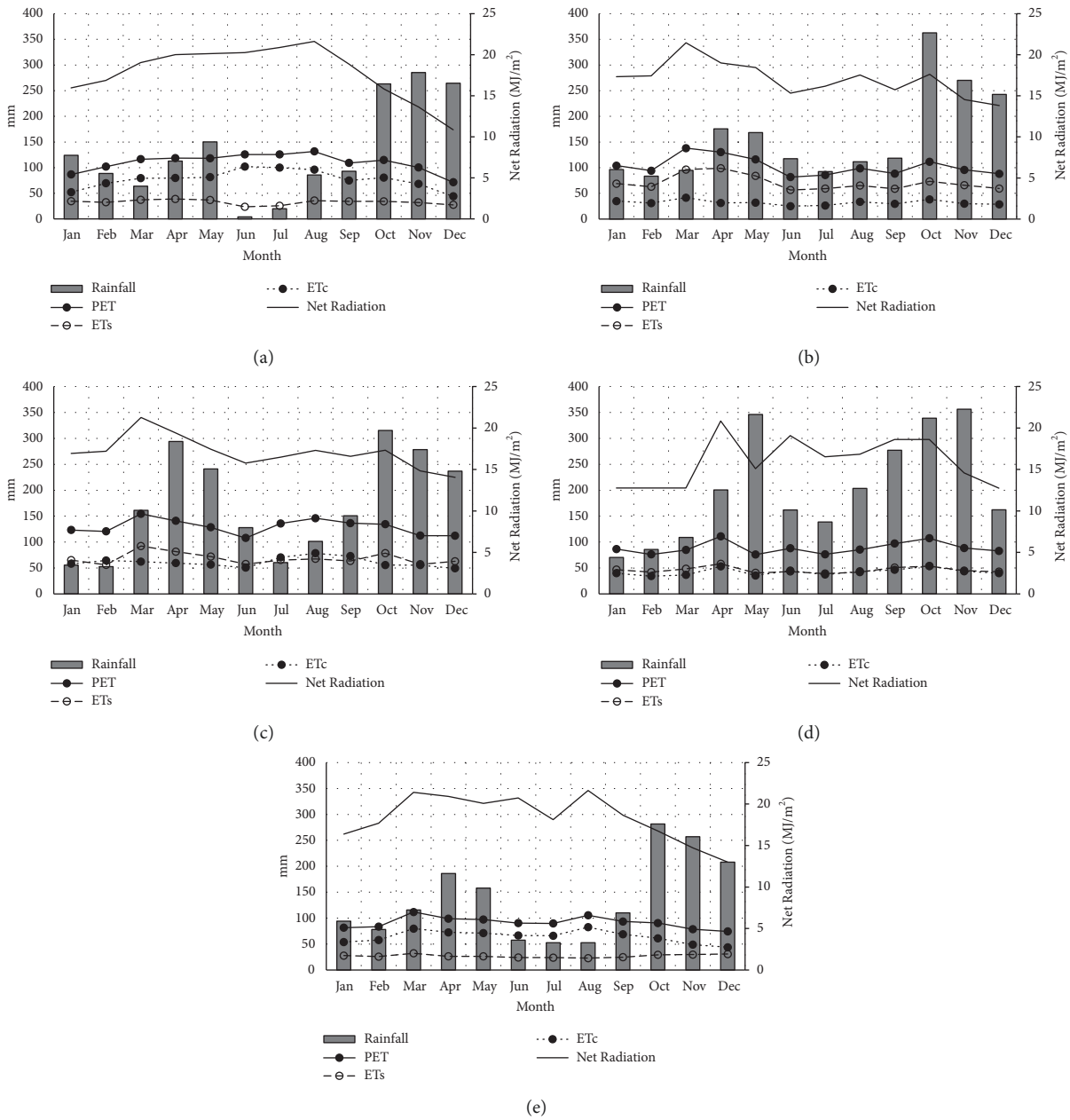


Figure 3: Seasonal patterns of monthly PET, ETs, ETc, rainfall, and averaged daily net solar radiation (all parameters are averaged for the period of 2009–2019). (a) For Vavuniya. (b) For Kurunegala. (c) For Bandarawela. (d) For Katugastota. (e) For Galle.

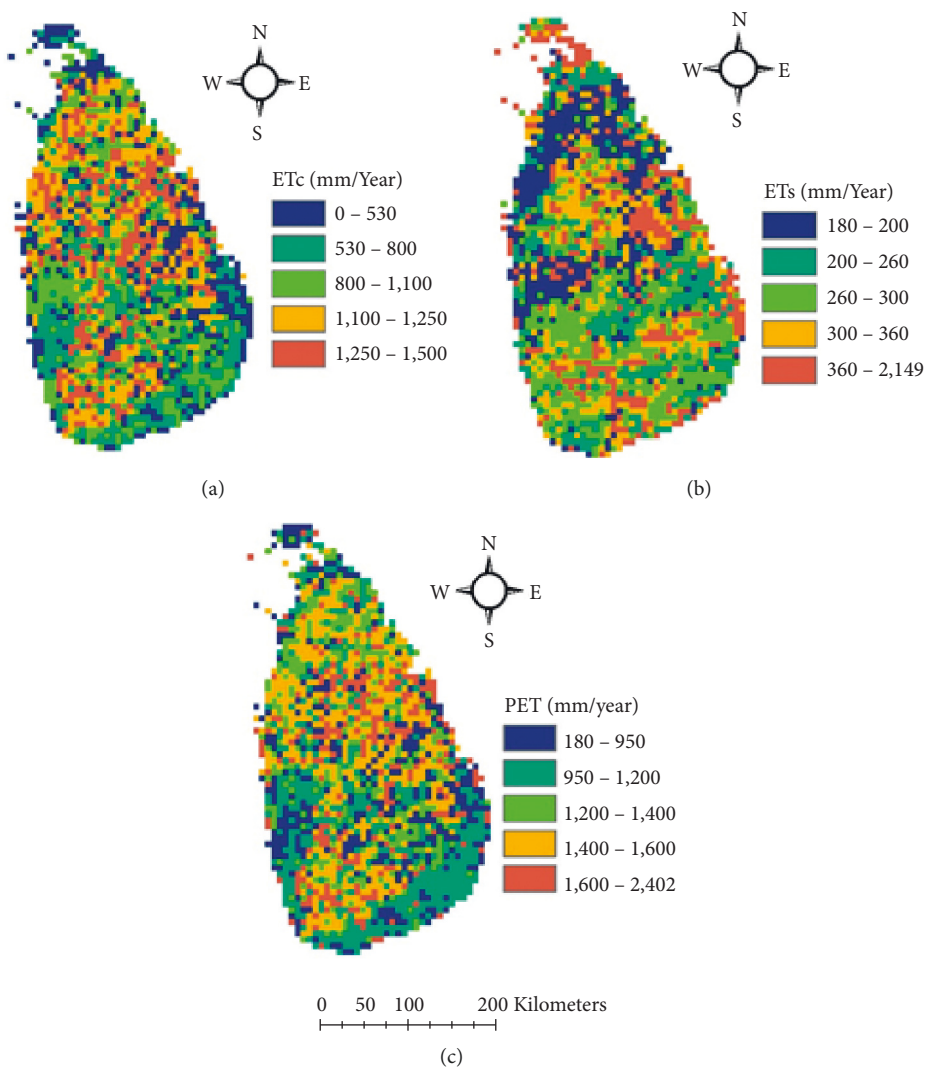


Figure 4: Spatial distribution of averaged annual ET_c, ET_s, and PET in $6km \times 6km$ grid (created by the authors). (a) For ET_c. (b) For ET_s. (c) For PET.

Within a year, monthly PET values fluctuate between 125 and 70 mm in the Vavuniya (84%), Katugastota (69%), and Bandarawela (50%) and Kurunegala, whilst Galle (46.3%) and Kurunegala (43%) are seeing much more minor variations around 125 mm and 100 mm, respectively. Vavuniya experiences the highest PET, ETc values in the period from June to August, while peaks arise either in July or August. The reason is the high incoming solar radiation and the irrigated cultivation, which facilitates the continuous supply of water despite the dry spell the dry zone experiences from June to August. The two stations from the intermediate zone show two distinct peaks of PET, one in March or April and the other one in July or August. Similarly, two peaks are seen in the wet zone, either in March or April and October. A decrease in incoming radiation fluxes has been observed all over the country from October to December. Relating to this observation, PET shows a decreasing trend from October to December in all three climatic zones, with a mean rate of 10 mm/month. ETs, as predicted, is associated with the rainy seasons. The highest amount of rainfall has been recorded from October to November, and this period is a rainy season for the entire island. Thus, a peak in ETs can be observed everywhere from October to December, and it decreases from January to mid-year until ETs record their lowest in June or August. When annual sums are considered in all the climate zones, ETc is superior to ETs, unless at bare lands, build-up lands, and water bodies. In Vavunia and Bandarawela, ETc is always greater than ETc, indicating that the dry substrate has limited evaporation from the soil.

4.4. Spatial Distribution of PET. The model outputs, PET, ETs, and ETc, were averaged from 2009 to 2019 and these were exported to the ArcGIS environment to visualize the spatial distribution of PER, ETs, and ET covering the entire country. These distributions are presented in Figure 4.

The spatial distribution of these parameters shows significant heterogeneity over the small landmass of Sri Lanka.

Table 4: Maximum LAI values for different vegetation types [20].

Code	Land use type	Classification as per the literature	LAI_{max}	h_c (m)	l_{max} (m)	F_{cl}	$r_{st\,min}$ (sm^{-1})	$NDVI_{98\%}$	z_{og} (m)
1	Coconut	Evergreen needle leaf forests	5.5	17	0.001	1	150	0.689	0.02
2	Rubber	Evergreen broadleaf forests	7	30	0.05	0	150	0.611	0.02
3	Forest, unclassified	Mixed forests	5.7	20	0.04	0.5	150	0.721	0.02
4	Homesteads/garden	Open shrub lands	3	1	0.01	1	100	0.674	0.02
5	Shrublands		3	1	0.01	1	100	0.674	0.02
6	Tea		3	1	0.01	1	100	0.674	0.02
7	Grasslands	Grasslands	1.8	0.8	0.01	0	115	0.674	0.01
8	Marshy lands	Permanent wetlands	6	1	0.01	0	65	0.674	0.01
9	Chena	Croplands	7	0.6	0.01	0	90	0.674	0.05
10	Other cultivations		7	0.6	0.01	0	90	0.674	0.05
11	Paddy		7	0.6	0.01	0	90	0.674	0.05
12	Urban and built-up	Urban and built-up	0	0	0	0	0	0.674	0.02
13	Barren land	Barren or sparsely vegetated	0.3	0.05	0.01	1	120	0.674	0.01
14	Water bodies	Water bodies	0	0	0	0	0	0.674	0.001

Low to medium PET occurs in the high lands. ETs also account for the evaporation from water bodies. It is shown that higher ETs occur in the North-central part, the Anuradhapura region, most evidently due to a large number of irrigation tanks in the region. The wet zone also has shown high ETs, as expected, due to high precipitation keeping the soil moist throughout the year.

Table 4 showcases the predicted values of annual PET, ETs ETc for different land uses. As discussed earlier, nine land uses, including paddy, rubber, coconut, Chena, tea, grasslands, shrublands, home gardens, and urban and built-up areas, were considered in generating the predicted results.

As shown in Table 4, PET from paddy lands, which has been predicted to have an annual average of 1392 mm and an annual maximum of 2035 mm, is the highest among the PET from cultivated lands. This is related to flood irrigation. It is also shown that PET from paddy fields is highest in the dry zone and lowest in the wet zone. This spatial distribution mainly correlates with the net solar radiation flux, which is higher in the dry zone. Rubber, which is planted only in the wet zone on gentle slopes, transpires water at an average rate of 1293 mm/year and seconds only to paddy. The evergreen broadleaf plant, growing on year-long moist soil, explains the second-highest transpiration rate of rubber. The lowest PET values occur in the urban built-up areas, grasslands, and home gardens with an average of 405 mm/year, 665 mm/year, and 872 mm/year, respectively.

5. Implications on Water Resources Management

As the S-W model accounts for leaf area index, with proper calibration, the model can be used in agricultural and irrigation water management, which was evident from the observations of this work: (1) peaks in plant growing seasons (Figure 3) and (2) significantly higher PETs in extensively cultivated areas (Figure 4). However, given the complexity of the model, the model may use for long-term irrigation planning instead of seasonal planning. Since the model shows an excellent correlation with pan evaporation, which has been used in water management in Sri Lanka, the S-W model could be utilized for future predictions of evapotranspiration for climate change resilience attempts.

6. Summary and Conclusions

This study aims to develop the S-W model and evaluate the model's efficiency in predicting PET over entire Sri Lanka. The model demands many meteorological data, soil, and vegetation parameters. Since some required meteorological data are not readily available, especially in data-scarce countries like Sri Lanka, the model holds many limitations. In addition, most of the soil and vegetation parameters are impossible or hard to measure. These facts restrict the extent to which the model can be calibrated and validated and may incur significant yet inevitable uncertainties in the model outputs.

However, even under these limitations, the model developed gives an RMSE of 0.75 mm/day when compared with historic pan evaporation values. It can be concluded that a moderate to high confidence can be placed on model predictions, except in the SE quadrant of the country. The lower density of weather stations in the SE region caused this variation. The annual ET_c, ET_s, and PET were presented in $6km \times 6km$ grid to showcase the spatial distribution of these important parameters. Therefore, the country now has its own spatial distribution maps for the PET.

The model shows that the PET over Sri Lanka is primarily governed by the incoming radiation flux, followed by rainfall and cultivation seasons. The model is highly sensitive to LAI, thus the vegetation cover. In the dry zone, 80–90% of rainwater escapes to the atmosphere through evaporation, while the percentage loss in the wet zone varies within 40–65%. Evaluation of annual figures indicates that approximately 65% of the total evapotranspiration amount is coming from vegetation canopy through transpiration. The model can be used as a guide to develop new agricultural policies in the changing climate of Sri Lanka. In addition, this study can be used as an example for tropical countries like Sri Lanka.

Appendix

A. Set of Equations

The evapotranspiration model is

$$PM_c = \frac{\Delta(R_n - G) + \left(\left\{24 \times 3600 \times \rho C_p D - \Delta r_{ac}(R_{ns} - G)\right\}/(r_a + r_{ac})\right)}{\Delta + \gamma[1 + (r_c/(r_a + r_{ac}))]}, \tag{A.1}$$

$$PM_s = \frac{\Delta(R_n - G) + \left\{24 \times 3600 \times \rho C_p D - \Delta r_{as}(R_n - R_{ns})\right\}/(r_a + r_{as})}{\Delta + +\gamma[1 + (r_s/(r_a + r_{ac}))]}, \tag{A.2}$$

$$C_c = \frac{1}{1 + (R_c R_a)/[R_s(R_c + R_a)]}, \tag{A.3}$$

$$C_s = \frac{1}{1 + (R_s R_a)/[R_c(R_s + R_a)]}, \tag{A.4}$$

$$R_a = (\Delta + \gamma) r_a, \tag{A.5}$$

$$R_c = (\Delta + \gamma) r_{ac} + \gamma r_c, \tag{A.6}$$

$$R_s = (\Delta + \gamma) r_{as} + \gamma r_s, \tag{A.7}$$

$$\lambda = 2.501 - 0.002361T. \tag{A.8}$$

$$\Delta = \frac{4098}{(237.3 + T)^2} \times 0.611 \times \exp\left(\frac{17.27T}{237.3 + T}\right), \tag{A.9}$$

$$\rho = \frac{P}{R_a T_k}, \tag{A.10}$$

$$\gamma = \frac{C_p P}{0.622\lambda}. \tag{A.11}$$

λ, Δ, ρ and γ are directly related to climatic factors [23].

$$P = 101.3\left(\frac{293 - 0.0065z}{293}\right)^{5.26}, \tag{A.12}$$

where T is the average temperature (°C). P is the atmospheric pressure (kPa). R_a is the specific gas constant (= 0.287 kJ·kg –1K–1). T_k is the mean temperature in Kelvin (= 273 + T) (K)

Radiation terms are

$$\mathrm{SR} = \frac{1 + \mathrm{NDVI}}{1 - \mathrm{NDVI}}, \tag{A.13}$$

$$\mathrm{FPAR} = \mathrm{FPAR}_{\min} + (\mathrm{FPAR}_{\max} - \mathrm{FPAR}_{\min}) \frac{(\mathrm{SR} - \mathrm{SR}_{\min})}{(\mathrm{SR}_{\max} - \mathrm{SR}_{\min})}, \tag{A.14}$$

$$\mathrm{LAI} = (1 - F_{cl}) \cdot \mathrm{LAI}_{\max} \frac{\ln(1 - \mathrm{FPAR})}{\ln(1 - \mathrm{FPAR}_{\max})} + (F_{cl}) \cdot \mathrm{LAI}_{\max} \frac{\mathrm{FPAR}}{\mathrm{FPAR}_{\max}}, \tag{A.15}$$

where SR is the simple ratio of hemispheric reflectance for the near-infrared light to that for the visible light. NDVI is the normalized difference vegetation index. PAR is the fraction of photo-synthetically active radiation. $\mathrm{FPAR}_{\min} = 0.001$. $\mathrm{FPAR}_{\max} = 0.95$. $\mathrm{SR}_{\min}$ is the SR estimated for NDVI at 5% vegetation population (NDVI at 5% = 0.039 globally). $\mathrm{SR}_{\max}\mathrm{SR}_{\max}$ is the SR estimated for NDVI at 95% vegetation population (NDVI at 95% refer to Table 2). $F_{cl}F_{cl}$ is the fraction of clumped vegetation (refer to Table 2).

Water vapor deficit at the reference height is

$$e_s = 0.611 \times \exp\left(\frac{17.27T}{237.3 + T}\right), \tag{A.16}$$

$$e_a = R_h e_s, \tag{A.17}$$

where e_s is the saturation vapor pressure (kPa). e_a is the ambient vapor pressure (kPa). R_h is the relative humidity.

Aerodynamic resistance between canopy source and reference level is

$$h_c = \begin{cases} 0, & \mathrm{LAI}_{\max} = 0, \\ h_{c\min} + (h_{c\max} - h_{c\min}) \dfrac{\mathrm{LAI}}{\mathrm{LAI}_{\max}}, & \mathrm{LAI}_{\max} \neq 0, \end{cases} \tag{A.18}$$

$$u_* = \frac{k u_a}{\ln((z_a - d_0)/z_o)}, \tag{A.19}$$

$$d_0 = \begin{cases} h_c - z_{oc}/0.3, & \mathrm{LAI} \geq 4, \\ 1.1 h_c \ln\left[1 + (C_d \mathrm{LAI})^{0.25}\right], & \mathrm{LAI} < 4, \end{cases} \tag{A.20}$$

$$\eta = \begin{cases} 2.5, & h_c \leq 1, \\ 2.036 + 0.194 h_c, & 1 < h_c < 10, \\ 4.25, & h_c \geq 10, \end{cases} \tag{A.21}$$

$$K_h = k u_* (h_c - d_0), \tag{A.22}$$

$$Z_0 = 0.13 h_c, \tag{A.23}$$

$$d_p = 0.63 h_c, \tag{A.24}$$

$$z_0 = \begin{cases} 0.3(h_c - d_0), & 0 < C_d \mathrm{LAI} < 0.2, \\ z_{og} + 0.3 h_c (C_d \mathrm{LAI})^{0.5}, & 0.2 < C_d \mathrm{LAI} < 1.5, \end{cases} \tag{A.25}$$

$$z_{oc} = \begin{cases} 0.13h_c, & h_c \le 1, \\ 0.139h_c - 0.009h_c^2, & 1 < h_c < 10, \\ 0.05h_c, & h_c \ge 10, \end{cases} \quad (A.26)$$

$$C_d = \begin{cases} 1.4 \times 10^{-3}, & h_c = 0, \\ \dfrac{\left[-1 + \exp\left(0.909 - 3.03z_{oc}/h_c\right]^4\right.}{4}, & h_c > 0, \end{cases} \quad (A.27)$$

where $\mathrm{LAI}_{\max}$ is the maximum LAI (refer to Table 2). u_a is the wind speed at the reference height (m s-1). z_o is the roughness length of the canopy (m). z_{oc} is the roughness length of the closed canopy (m). C_dC_d is the mean drag coefficient for individual leaves. z_{og} is the roughness length of ground (m) (refer to Table 2).

Canopy resistance [20] is

$$\mathrm{LAI}_e = \begin{cases} \mathrm{LAI}, & \mathrm{LAI} \le 2, \\ 2, & 2 < \mathrm{LAI} \le 4, \\ 0.5\mathrm{LAI}, & \mathrm{LAI} \ge 4, \end{cases} \quad (A.28)$$

$$f\left(R_n\right) = \frac{dR_n}{c + R_n}, \quad (A.29)$$

$$f\left(D\right) = \begin{cases} 1 - 0.409\ D, & \text{for short vegetation}, \\ 1 - 0.238\ D, & \text{for tall vegetation}, \end{cases} \quad (A.30)$$

$$f\left(T_k\right) = \begin{cases} 1 - 1.6 \times 10^{-3}\left(298 - Tm\right), & T_k \ge 298, \\ 1 - 1.6 \times 10^{-3}\left(298 - T_k\right), & 273 < T_k < 298, \end{cases} \quad (A.31)$$

$$f\left(\theta\right) = \begin{cases} 1 - 1, & \theta \ge \theta_c, \\ \dfrac{\theta - \theta_w}{\theta_c - \theta_w}, & \theta_w \le \theta < \theta_c, \\ 0 - 1, & \theta < \theta_w, \end{cases} \quad (A.32)$$

where c and d are constants, which are defined as $d = 1/(1 + c)$ [29]. $c = 100$ for forests and 400 for crops. R_n is the net radiation in W m −2. θ is the soil moisture. θ_w is the plant permanent wilting point. θ_c is the critical soil moisture at which transpiration is stressed. $\theta_c \times 0.75$ is the saturated soil moisture.

Data Availability

The climatic data and the analysis data are available for research purposes from the corresponding author upon request.

Disclosure

The research was carried out in the Sri Lanka Institute of Information Technology environment.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

The authors thank the support received from Sri Lanka Institute of Information Technology, Sri Lanka, to carry out this research work. This research was carried out under financial support of SLIIT Research under Grant FGSR/RG/FE/2021/11.

References

[1] D. Lang, J. Zheng, J. Shi et al., "A comparative study of potential evapotranspiration estimation by eight methods with FAO penman-monteith method in southwestern China," *Water*, vol. 9, no. 10, p. 734, 2017.

[2] H. L. Penman, "Natural evaporation from open water, bare soil and grass," *Proceedings of the Royal Society of London - Series A*, vol. 193, no. 1032, pp. 120–145, 1948.

[3] S. Dunn and R. Mackay, "Spatial variation in evapotranspiration and the influence of land use on catchment hydrology," *Journal of Hydrology*, vol. 171, pp. 49–73, 1995.

[4] K. J. Kristensen and S. E. Jensen, "A model for estimating actual evapotranspiration from potential evapotranspiration," *Hydrology Research*, vol. 6, no. 3, pp. 170–188, 1975.

[5] W. Verstraeten, F. Veroustraete, and J. Feyen, "Assessment of evapotranspiration and soil moisture content across different scales of observation," *Sensors*, vol. 8, no. 1, pp. 70–117, 2008.

[6] J. Lu, G. Sun, S. G. McNulty, and D. M. Amatya, "A comparison OF six potential evapotranspiration methods for regional use IN the southeastern United States," *Journal of the American Water Resources Association*, vol. 41, no. 3, pp. 621–633, 2005.

[7] C. W. Thornthwaite, "An approach toward a rational classification of climate," *Geographical Review*, vol. 38, no. 1, p. 55, 1948.

[8] C. H. B. Priestley and R. J. Taylor, "On the assessment of surface heat flux and evaporation using large-scale parameters," *Monthly Weather Review*, vol. 100, no. 2, pp. 81–92, 1972.

[9] J. L. Monteith, "Evaporation and environment," *Symposia of the Society for Experimental Biology*, vol. 19, pp. 205–234, 1965.

[10] W. J. Shuttleworth and J. S. Wallace, "Evaporation from sparse crops-an energy combination theory," *Quarterly Journal of the Royal Meteorological Society*, vol. 111, no. 469, pp. 839–855, 1985.

[11] B. Choudhury and J. Monteith, "A four-layer model for the heat budget of homogeneous land surfaces," *Quarterly Journal of the Royal Meteorological Society*, vol. 114, no. 480, pp. 373–398, 1988.

[12] X. Mo, S. Liu, Z. Lin, and W. Zhao, "Simulating temporal and spatial variation of evapotranspiration over the Lushi basin," *Journal of Hydrology*, vol. 285, pp. 125–142, 2004.

[13] X. Li and Q. Zhang, "Estimating the potential evapotranspiration of poyang lake basin using remote sense data and shuttleworth-wallace model," *Procedia Environmental Sciences*, vol. 10, pp. 1575–1582, 2011.

[14] F. Yuan, L.-L. Ren, Z.-B. Yu, and J. Xu, "Computation of potential evapotranspiration using a two-source method for the xin'anjiang hydrological model," *Journal of Hydrologic Engineering*, vol. 13, no. 5, pp. 305–316, 2008.

[15] J. M. Chen, X. Chen, W. Ju, and X. Geng, "Distributed hydrological model for mapping evapotranspiration using remote sensing inputs," *Journal of Hydrology*, vol. 305, pp. 15–39, 2005.
[16] M. R. Herman, A. P. Nejadhashemi, M. Abouali et al., "Evaluating the role of evapotranspiration remote sensing data in improving hydrological modeling predictability," *Journal of Hydrology*, vol. 556, pp. 39–49, 2018.
[17] W. W. Immerzeel and P. Droogers, "Calibration of a distributed hydrological model based on satellite evapotranspiration," *Journal of Hydrology*, vol. 349, pp. 411–424, 2008.
[18] G. Kite and P. Droogers, "Comparing evapotranspiration estimates from satellites, hydrological models and field data," *Journal of Hydrology*, vol. 229, pp. 3–18, 2000.
[19] E. T. Linacre, "A simple formula for estimating evaporation rates in various climates, using temperature data alone," *Agricultural Meteorology*, vol. 18, no. 6, pp. 409–424, 1977.
[20] M. C. Zhou, H. Ishidaira, H. P. Hapuarachchi, J. Magome, A. S. Kiem, and K. Takeuchi, "Estimating potential evapotranspiration using Shuttleworth-Wallace model and NOAA-AVHRR NDVI data to feed a distributed hydrological model over the Mekong River basin," *Journal of Hydrology*, vol. 327, pp. 151–173, 2006.
[21] K. A. U. S. Imbulana, N. T. S. Wijesekara, and B. R. Neupane, *Sri Lanka National Water Development Report*, University of Moratuwa, Moratuwa, Sri Lanka, 2006.
[22] N. A. Jackson and J. S. Wallace, "Soil evaporation measurements in an agroforestry system in Kenya," *Agricultural and Forest Meteorology*, vol. 94, no. 3–4, pp. 203–215, 1999.
[23] V. T. Chow, D. R. Maldment, and L. W. Mays, *Applied Hydrology*, McGraw-Hill, Singapore, 1988.
[24] W. J. Shuttleworth and R. J. Gurney, "The theoretical relationship between foliage temperature and canopy resistance in sparse crops," *Quarterly Journal of the Royal Meteorological Society*, vol. 116, no. 492, pp. 497–519, 1990.
[25] X. Li, S. Kang, F. Li et al., "Applying segmented Jarvis canopy resistance into Penman-Monteith model improves the accuracy of estimated evapotranspiration in maize for seed production with film-mulching in arid area," *Agricultural Water Management*, vol. 178, pp. 314–324, 2016.
[26] USGS, https://earthexplorer.usgs.gov/, 2021.
[27] F. R. Moormann and C. R. Panabokke, *Soils of Ceylon: A New Approach to the Identification and Classification of the Most Important Soil Groups of Ceylon*, Valaichchenai: Government Press, Kalkudah, Sri Lanka, 1961.
[28] NASA, https://earthdata.nasa.gov/, 2021.
[29] J. Noilhan and S. Planton, "A simple parameterization of land surface processes for meteorological models," *Monthly Weather Review*, vol. 117, no. 3, pp. 536–549, 1989.

Hindawi
Advances in Meteorology
Volume 2022, Article ID 7067951, 19 pages
https://doi.org/10.1155/2022/7067951

Research Article

Hydrological Drought Analysis using Streamflow Drought Index (SDI) in Ethiopia

Kassa Abera Tareke [1,2] **and Admasu Gebeyehu Awoke** [2]

[1] *Department of Hydraulic and Water Resources Engineering, Wollo University, Kombolcha Institute of Technology KioT, Dessie, Ethiopia*
[2] *School of Civil and Environmental Engineering, Addis Ababa University, Addis Ababa Institute of Technology, Addis Ababa, Ethiopia*

Correspondence should be addressed to Kassa Abera Tareke; kassa.abera@aait.edu.et

Received 5 February 2022; Revised 16 March 2022; Accepted 4 April 2022; Published 22 April 2022

Academic Editor: Upaka Rathnayake

Drought is a natural disaster that has impacts on society, the environment, and the ecosystem. Ethiopia faced many horrible severe drought events in the last few decades. Even though there are some drought-related studies in the country, most of the investigations were focused on meteorological drought analysis. This study was focused on hydrological drought analysis in Ethiopia using the streamflow drought index (SDI). The main objective was to identify drought-prone areas and severe drought events years. Streamflow data were collected from 34 stations to analyze SDI in seasonal (3-month) and annual (12-month) timescales. The analysis implies that seasonal time scale (3-month) hydrological drought has a high frequency of occurrence but short duration, whereas annual (12-month) analysis has a low frequency with a large magnitude. The overall result shows that 1984/85, 1986/87, 2002/03, and 2010/11 were the most severe and extreme drought years in all river basins. The 1980s were found severe and extreme drought years in which most hydrological drought events occurred in the country. The spatial analysis shows that Tekeze, Abbay, and Baro river basins have similar characters; Awash and Rift Valley River basins show relatively the same character, and Genale Dawa and Wabishebele river basins have a similar character. But Omo Gibe River basin has a unique character in which the severe drought occurred in a different year of other river basins.

1. Introduction

Drought is a worldwide natural hazard and has a severe impact on the social, environmental, and economic aspects [1]. Hydrological extreme events both flood (high flow) and drought (low flow) are the most concerning issues in the world [2]. However, drought is one of the most common natural disasters that have a great negative impact on agriculture and water resources projects in a wide range [3]. But researchers are highly focused on flood disasters than drought.

Drought is a complex phenomenon, and due to its notorious nature, there is no universal definition for it [4]. Albeit it has no universal definition, droughts can be (1) meteorological, which is a scarcity of precipitation [5–7], (2) hydrological drought, which is the scarcity of streamflow water, reduction of reservoir and Lake water level [8–10], (3) agricultural drought related to deficit of soil moisture [11, 12], and (4) socioeconomic drought which is the imbalance between supply and demand [13, 14] and each drought is a cause of the other [15]. Meteorological and agricultural droughts are mainly causing the failure of crops, while hydrological drought is a deficiency of water supply, a decrease in reservoir water level and groundwater level, and a decrease in irrigation and hydropower production [16]. The aggregated impact of meteorological, agricultural, and hydrological drought and the contradiction between water supply and demand results in socioeconomic drought [17, 18], in which the overall ecosystem of the drought region is disturbed, and there is a loss of life [19]. The overall

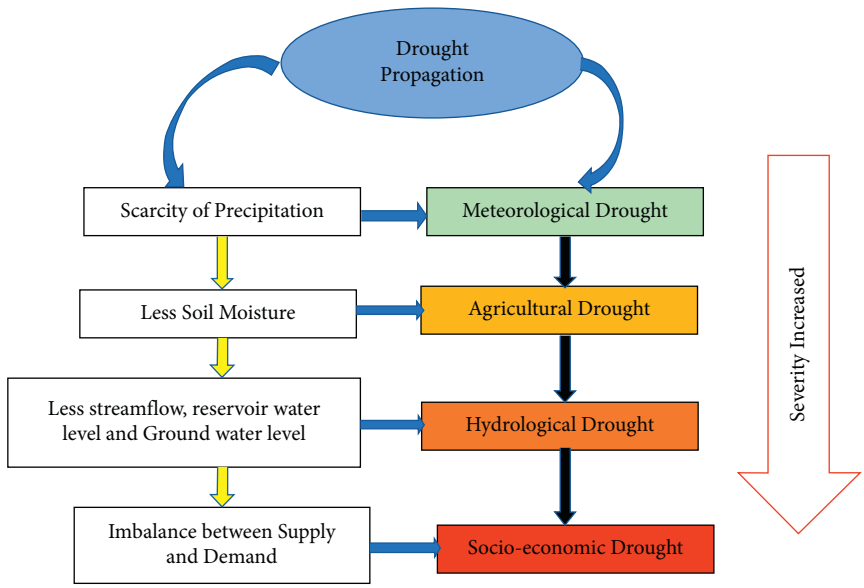


Figure 1: Relationship and propagation of different types of droughts.

propagation of each drought development is summarized in Figure 1.

The scientific analysis of the recurrence and persistence of drought seeks to establish an estimate of probabilities that will contribute to the planning of strategies for mobilization and management of water resources [19]. Due to climate change, the global vulnerability to floods and drought has been increasing [20]. The assessment of droughts affecting agricultural areas in Africa is highly relevant. Many African countries strongly rely on rain-fed agriculture. Since rainfall is commonly the limiting factor for rain-fed agricultural systems, droughts can lead to severe socioeconomic consequences like crop failure, food shortages, and even humanitarian crises [21]. For example, drought events have caused more than 800,000 deaths and affected about 262 million people in Africa from 1900 to 2013 [22].

Ethiopia has also experienced severe drought in the past centuries. Researchers have thought that severe drought has occurred once in three to ten years' recurrence interval [23]. Even though this recurrence time is quite short, no permanent drought mitigation measures have been established effectively. Measures are more localized and usually come in the form of food aid during the drought period.

Large-scale drought analysis studies have been recently conducted globally or continental level for present and future climate change [24]. But in the case of Ethiopia, drought studies were focused on catchment level and they do not give good information about the drought condition of the country. Few drought studies were conducted in the northern and eastern parts of Ethiopia, which are mainly focused on meteorological drought analysis [25]. Finally, the researchers try summarizing the effect of drought events throughout the country. But drought events are inconsistent across all of Ethiopia. The variability of the temporal, spatial, and magnitude characteristics of drought events requires in-depth analysis to determine different approaches for mitigation. The government of Ethiopia develops and planned different water resource infrastructures to reduce poverty. Yet, hydrological drought analysis is not well studied. Therefore, to develop permanent and effective drought mitigation measures, it is important to analyze the historical hydrological drought over the nation. So, the main objective of this study is to analyze and characterize historical hydrological drought in Ethiopia using the streamflow drought index (SDI) for a better understanding of its impact on water resource infrastructure development.

2. Methods

2.1. Description of the Study Area. Ethiopia is located in the eastern horn of Africa, which locates in 3°N and 15°N latitude and 33°E and 48°E longitude. The area of the country is about 1.13 million km^2 and this study covers 87.3% of the country. Of 12 major river basins, 4 are not included in this study due to a lack of data (Figure 2). The population of the country rapidly increases from decade to decade causing deforestation and expansion of urban areas. As a result, variability in climate conditions also increased. There are twelve major river basins, yet the development of water resources is still low. Of the existing 12 major river basins, three (3) are dry and one river basin has low flow and low recorded data as presented in Figure 2 northeast part. Due to this reason, this study was focused only on eight (8) major river basins. The drought phenomenon is a common natural hazard throughout the country, especially since the severity is high in the northern and eastern parts. The country receives high rainfall during the summer (June, July, and August). The mean annual rainfall and a temperature range from 510 to 1300 mm and 16°c to 27°c, respectively.

The topography of Ethiopia and the respective major river basins are described by the Digital Elevation Model (DEM), which was collected from the Ministry of Water, Irrigation and Electricity, GIS division with a spatial resolution of 30 m × 30 m. The river network and basin boundary

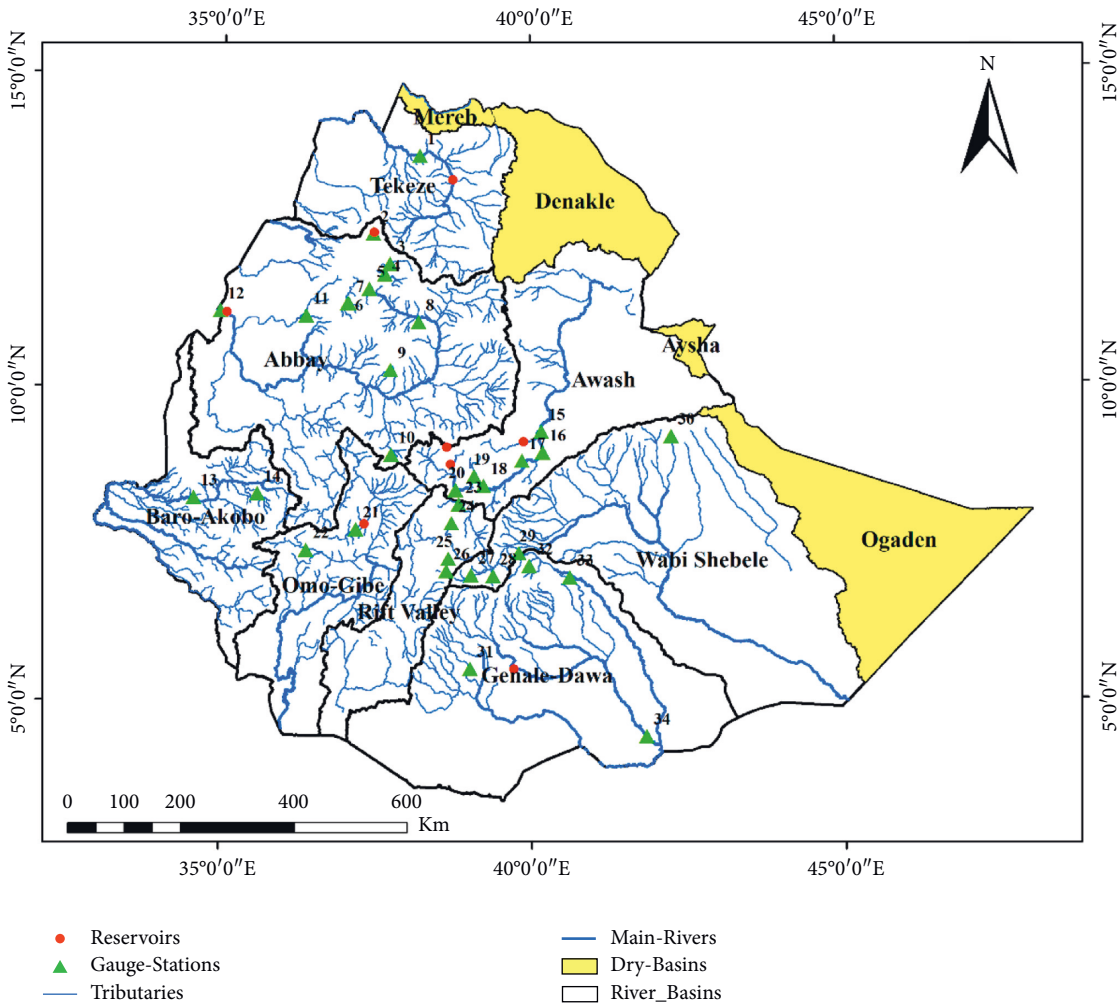


Figure 2: Location of streamflow stations.

are automatically extracted from the DEM itself by performing the ArcGIS software and the spatial location of 34 streamflow gauge stations and some major reservoirs are indicated in Figure 2.

2.2. Data Collection and Preparation. Streamflow data were collected from the Ministry of Water, Irrigation, and Electricity (MoWIE). The stations were selected based on data quality, recorded length, and area coverage. The data were collected from 34 streamflow stations with a length of 35–41 years, except for the Tekeze basin, which has only 21 years' record length (Table 1). The streamflow stations in the Tekeze basin have more missing data and it is difficult to analyze drought trends with minimum data availability. Therefore, for this basin, only one streamflow station is considered, which has 21 years of recorded data from 1994 to 2014. But, the other seven basins fulfill the minimum data length required (30 years) for drought analysis. Table 1 summarizes the spatial location of the gauge station and its corresponding data record length.

The annual flow in each basin varies from season to season. Abbay, Baro Akob, and Omo Gibe River basins (northwest, western, and southern parts) have a high flow than other basins, 52.6, 23.6, and 17.9 BM^3 annual runoff, respectively, whereas Tekeze Genale Dawa, Rift Valley, Awash, and Wabishebele (northern, central and eastern parts) have a relatively moderate flow, 7.6, 5.8, 5.6, 4.6, and 4.6 BM^3 annual runoff, respectively. Figure 3 illustrates the mean monthly flow rate (m^3/s) of all river basins. As shown in Figure 3, Abbay has a high flow rate followed by the Omo Gibe River basin.

2.3. Hydrological Drought Indicators Selection. There is no single drought indicator for all types of drought in a specific region [26] because all available drought indicators have their limitation during development and application [27]. Therefore, drought indicator selection requires a thorough investigation related to the type of drought and the respective drought indicator based on the availability of data, ease of communication, result implication, strength and limitations of the indices, and the objective of the investigation [26]. Of many drought indices, the most common are discussed in Table 2.

From the water resources development and management perspective, it is important to define the reference flow levels and indicators of drought severity (what drought duration

Table 1: Streamflow stations location and a corresponding record year in each river basin.

No	Station name	Basin	Latitude (°)	Longitude (°)	Area (km^2)	Recorded year
1	Embamadre	Tekeze	13.73	38.20	45694	1994–2014
2	Megech	Abbay	12.48	37.45	462	1980–2014
3	Rib	Abbay	12.00	37.72	1592	1972–2012
4	Gummera	Abbay	11.83	37.63	1394	1973–2014
5	Bahir Dar	Abbay	11.60	37.38	15319	1973–2014
6	Gilgel Abbay	Abbay	11.37	37.03	1664	1972–2012
7	Koga	Abbay	11.37	37.05	244	1972–2012
8	Kessie	Abbay	11.07	38.18	65784	1973–2014
9	Chemoga	Abbay	10.30	37.73	364	1973–2009
10	Gilgel Beles	Abbay	11.17	36.37	675	1973–2014
11	Guder	Abbay	8.95	37.75	524	1973–2009
12	Abbay border	Abbay	11.23	34.98	17254	1973–2014
13	Gambela	Baro	8.25	34.58	23461	1973–2014
14	Sorie	Baro	8.32	35.60	1622	1973–2014
15	Melkawerer	Awash	9.32	40.17	31183	1973–2010
16	Awash7 killo	Awash	8.98	40.18	19110.75	1973–2013
17	Metehara	Awash	8.85	39.85	16416.8	1982–2014
18	Wonj	Awash	8.45	39.23	11690	1973–2014
19	Mojo	Awash	8.60	39.08	1264.4	1973–2014
20	Hombel	Awash	8.38	38.78	7656	1973–2014
21	Assendabo	Omo Gibe	7.75	37.18	2966	1980–2018
22	Gojeb	Omo Gibe	7.42	36.38	3577	1980–2018
23	Meki	Rift valley	8.15	38.83	2433	1973–2014
24	Kekersitu	Rift valley	7.85	38.72	7488	1980–2014
25	Dedeba	Rift valley	7.28	38.67	156	1980–2014
26	Wosha	Rift valley	7.08	38.63	20	1980–2014
27	Wabi bridge	Wabishebele	7.02	39.03	1035	1976–2014
28	Leliso	Wabishebele	7.00	39.38	135	1976–2014
29	Weiyb	Wabishebele	7.37	39.80	7719	1983–2014
30	Erer	Wabishebele	9.23	42.25	469	1983–2014
31	Chenemasa	Genale Dawa	5.52	39.02	10574	1983–2014
32	Shaya	Genale Dawa	7.17	39.97	4338	1981–2014
33	Weib	Genale Dawa	6.98	40.62	3576.9	1984–2014
34	Halewe	Genale Dawa	4.43	41.83	54093	1984–2014

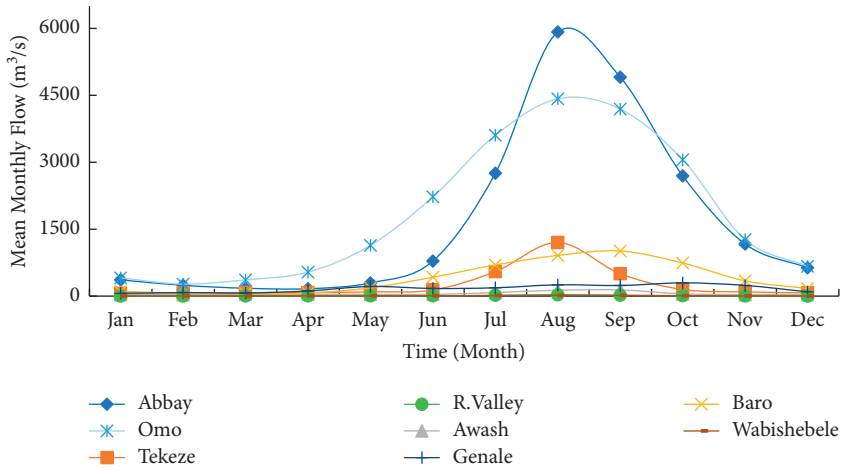


Figure 3: Seasonal variation of mean monthly streamflow over Ethiopian River Basins.

and/or flow deficit constitutes mild or severe drought). Several indices measure how much streamflow for a given period has deviated from historically established norms. Most water supply planners find it useful to consult one or more indices before making a decision [28, 29].

Hydrological drought can be defined in terms of the differences between the water supply and water demand time series. The supply time series is characterized by a river flow and the demand time series-by the demand of a particular user (e.g., irrigation) or by the total demand for all

Table 2: Summary of drought indicator's strengths and weakness.

No	Indicator	Strength	Weakness
1	Decile	Easy for analysis and used for all types of droughts at a different time scale	Uses only precipitation and does not consider temperature and other hydrological variables
2	Palmer drought severity index (PDSI)	Widely used, good for agricultural drought	The need for serially complete data cause problems
3	Palmer hydrology drought index (PHDI)	Its water balance approach allows the total water system to be considered	Frequencies will vary by region and time of year
4	Standardized precipitation index (SPI)	Easy to calculate, applicable in all climate regimes	The temperature effect is not considered
5	Standardized precipitation evapotranspiration index (SPEI)	Account for the impact of temperature on a drought situation, the output applies to all climate regimes	The requirement for a serially complete dataset for both temperature and precipitation limits its use due to insufficient data
6	Reconnaissance drought index (RDI)	Easy to use, precipitation and temperature as input, consider the water balance of the region	Comparing RDI for different climate regions may be subject to error
7	Streamflow drought index (SDI)	The program is widely available and easy to use. Missing data are allowed	A single input (streamflow) does not consider management decision
8	Surface water supply index (SWSI)	Considering the full water resources of a basin provides a good indication of the overall hydrological health of a particular basin or region	As data sources change or additional data are included, the entire index must undergo recalculation

Source: Handbook of Drought Indicators and Indices; Integrate Drought Management Program (IDMP); World Meteorological Organization WMO—No. 1173 (2016).

users. When demand exceeds supply, water shortages occur, which represents that the start of a such as storage is also a useful indicator of water shortages, due to data availability on a daily or weekly basis. Simultaneously, these data are strongly influenced by the reservoir operation rules [30].

The hydrological drought indices (HDIs) depend on the purpose of hydrologic monitoring and prediction, such as reservoir operation and water allocation for hydroelectric generation, irrigation, and domestic or industrial water supply. Most hydrological drought indices are largely based on stream flows. Streamflow deficit regarding normal conditions is not always the true representative, particularly in the river systems with diversion of virgin flows. Therefore, hydrological drought is characterized by more factors than just low flows. The preparation of streamflow data for drought management or any purpose is hampered by the lack of adequate data because shorter record lengths and artificial influences (such as abstraction) and even the longer recorded station have high missing values in developing countries such as Ethiopia.

Drought indices and definitions based solely on flow or reservoir storage are normally designed for reservoir operation and are seldom (if at all) used as triggers for drought relief, or drought monitoring over vast territories. Although none of the major indices is inherently superior to the rest in all circumstances, some indices are better suited than others for certain uses [31].

In Ethiopia, there is a high scarcity of data for streamflow analysis of both flood and drought. Most reservoirs in the country have no gauge staff and groundwater level is not well known in all basins, which makes it difficult to use more data-intensive drought indicators for hydrological drought analysis. Therefore, in such circumstances, SDI is the best alternative for hydrological drought analysis due to its less input requirement and simplicity for analysis and interpretation [32]. Streamflow is the most crucial variable in the quantity of water that expresses the availability of surface water resources. Therefore, in terms of normal conditions, a hydrological drought occurrence is linked to the streamflow deficit [20, 24]. Therefore, due to the following reasons, SDI is selected for this study: (i) the areal extent of a drought event is very useful for meteorological drought, but it is not of interest for hydrological droughts since water managers are interested in streamflow only as a small number of points in space (Basin outlets, Reservoir inlets, and outlets); (ii) streamflow at these points provides an integrated measure of spatially distributed runoff; (iii) furthermore, the river basin can be proposed as the unit for applying measures for water resource protection and management; (iv) in Ethiopia, there is a scarcity of data availability especially Lake level, reservoir level, soil moisture, and groundwater level. Relatively, precipitation and streamflow data are available from the 1970s to date.

2.3.1. Streamflow Drought Index. The streamflow drought index (SDI) was developed by [33], which is used to characterize the streamflow drought conditions. Its calculation is similar to SPI and therefore has the same characteristics of simplicity and efficiency. The SDI is based on monthly observed streamflow volumes at different time scales and thus offers the advantage of controlling streamflow drought or the supply of water in the short, medium, and long term. The calculation is given by

$$V_{i,j} = \sum_{j=1}^{k} Q_{i,j}\, i = 1, 2, \ldots,\ j = 1, 2, \ldots, 12,\ k = 1, 2, 3, 4, \quad (1)$$

where $V_{i,j}$ is the cumulative streamflow volume for the i-th streamflow year and the k-th reference period, and K is a seasonal value (four-season, Ethiopia case). $Q_{i,j}$ is monthly streamflow volume at i^{th} streamflow year and j^{th} month within that year.

Based on the cumulative streamflow volumes $V_{i,k}$, the streamflow drought index (SDI) is defined for each reference period k of the i-th streamflow year as follows:

$$SDI = \frac{Vi,k - Vkm}{Sk}, \quad (2)$$

where i = 1, 2; . . ., and k = 1, 2, 3, 4.

Vkm and *sk* are, respectively, the mean and the standard deviation of cumulative streamflow volumes of the reference period k as these are estimated over a long period. The range of wetness and dryness of SDI ranges between −2 and +2. The extremely dry and wet values are below −2 and above +2, respectively. According to Nalbantis and Tsakiris, 2009, hydrological drought classification using SDI is shown in Table 3.

Table 3: Drought classification according to the SDI values (Nalbantis and Tsakiris, 2009).

SDI value	Category
≥2	Extremely wet
1.5–1.99	Severely wet
1–1.49	Moderately wet
0.5–0.99	Slightly wet
−0.49–+0.49	Normal
−0.5–−0.99	Mild drought
−1–−1.49	Moderately drought
−1.5–−1.99	Severely drought
≤−2	Extremely drought

2.4. Hydrological Drought Characterization. The drought index plays a great role in order to evaluate the consequences of drought impact and to decide various drought characterization, such as duration (D), severity (S), magnitude (M), and relative frequency (RF) [34, 35]. Drought duration is the time taken between consecutive drought events (onset and end of drought). The duration is from starting with a negative SDI value and turns to a positive SDI value. Drought severity is the summation of negative SDI values from onset to end of a drought event as defined by Equation (3). The magnitude of drought is the ratio between drought severity and drought duration which is defined by equation (4):

$$Si = -\sum_{i=1}^{D} S\ DI\ i, \quad (3)$$

$$M = \frac{S}{D}. \quad (4)$$

The relative drought frequency is the ratio between the number of droughts (n) with negative SDI in drought duration and the total number of drought years in the analysis (N) [36] and RF is defined as

$$RF = \frac{n}{N} * 100. \quad (5)$$

3. Results and Discussion

3.1. Hydrological Drought Analysis using Drought Indices. Indices are important to give quantitative information about drought events in terms of duration, intensity, frequency, and recurrence interval for readers and scientists. Such information is extremely important for planning and water resource management [37]. There are many hydrological drought indices such as SDI, SWSI, PHDI, and ADI. But except for SDI, all indices are more data-intensive [32, 38]. Therefore, in this study, the historical hydrological drought trend in Ethiopia was analyzed using SDI. SDI is a point or site drought indicator that gives information about the river's temporal and special variation. The analysis was computed using the DrinC model (Drought Indicator Calculator), which is developed to determine SPI, RDI, and SDI using monthly input data [39, 40].

3.1.1. Hydrological Drought Analysis using SDI. The daily streamflow data were prepared monthly and normalized to zero mean and unit standard deviation to illuminate the flow variation of stations in time and space. Then, SDI was estimated using Equation (2) for short and long-time scales (SDI3 and SDI12), the seasonal base for the three-month duration (SDI3), and the annual base for the twelve-month duration (SDI 12). In Ethiopia, the summer covers from June to August (JJA); during this season nearly all river basins receive high rainfall and streamflow volume increased. Hydrological drought is developed gradually due to the scarcity of precipitation in the region when there is no rainfall for some consecutive months or a year. Therefore, this study focuses on the summer season (SDI3) and annually based (SDI12) hydrological drought analysis because the remaining three seasons, autumn (SON), winter (DJF), and spring (MAM) follow the summer season drought condition and it can be summarized by the annual drought conditions. The analysis result indicates that the 3-month time scale (SDI3) analysis frequency of drought occurrence was high compared to the long-term time scale, the 12-month (SDI12). This finding is agreed with previous studies such as [6, 19, 38, 41, 42]. But for the long-term time scale, drought duration is maximum and the SDI value is minimum as shown in Figure 4. In this study, the temporal and spatial variation of streamflow drought in Ethiopia was analyzed using SDI. Therefore, the most severe and extreme drought years over different river basins were identified. As shown in Table 4, the Rift Valley River basin was highly affected by 13 severe and 4 extreme droughts followed by the Abbay river basin for the last four decades.

In the Abbay river basin, 11 streamflow stations were considered for hydrological drought analysis using SDI. Since SDI is a point or site drought indicator, the discussion here considers all gauging stations within the basin and the result of each station is shown in Figure 4. The correlation of each station implies that the Abbay Border station correlates well with Kessie, Ribb, Megech, Gummera, Bahir Dar (0.85, 0.61, 0.61, 0.6, and 0.44), respectively. This 42-year drought analysis shows that the lower and central parts of the basin have similar drought occurrence seasons (Abbay border, Chemoga, Kessie, Guder, Gilgel Abbay, and Gilgel Beles),

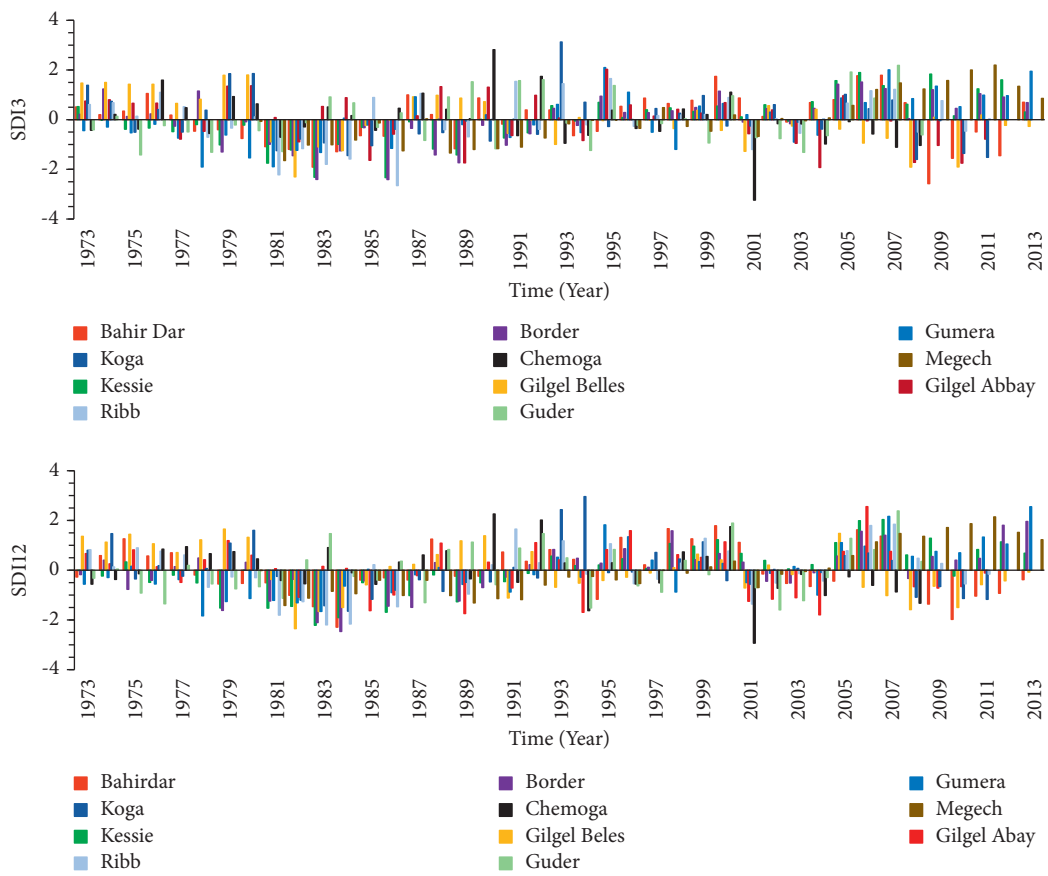


Figure 4: Temporal variation in seasonal (SDI3) and annual (SDI12) for Abbay river basin.

Table 4: Hydrological drought years in different river basins in Ethiopia from 1973 to 2014.

River basin	Severe drought years	Extreme drought year
Abbay	1979, 1981, 1983, 1984, 1986, 1994, 2010	1983/1984, 1984/1985
Awash	1979, 1996, 1997, 2001	1986/1987, 1987/1988, 2002/2003
Baro	1982, 1984, 1985, 2002	2002/2003, 2004/2005
Genale Dawa	1996, 2002, 2003, 2010	
Omo Gibe	1980, 1981, 1986	
Rift valley	1980, 1983, 1984, 1985, 1987, 1990, 1999, 2001, 2002, 2004, 2010, 2011, 2012	1984, 1985, 2003, 2012
Tekeze	1996	
Wabishebele	2001, 2002, 2004, 2005, 2007	2002

and the upper part of the basin is affected by drought differently (Megech, Bahir Dar, Ribb, and Koga). Commonly, the most severe streamflow drought year of all gauging stations in the Abbay basin was obtained in 1979, 1980, 1981, 1982, 1983, 1986, 1987, and 2010. This implies that for seven consecutive years (1980–1987), most parts of the basin were affected by severe drought (Figure 4). Besides these drought events, the period of 2001 to 2004 was dominated by moderate and severe drought conditions in most stations. But Chemoga station has been specifically affected by extreme drought in both seasonal (SDI3) and annual (SDI12) timescale during 2001 (Figure 4). In this basin, an extreme drought occurred in 1983/1984, and 1984/1985 in the upper, middle, and lower parts of the basin. The result is supported by [18, 43]; however, these studies were focused on the upper parts of the basin only but the lower and middle parts of the basin have not yet been studied before.

The Awash river basin is country-locked and it is extensively used for different purposes such as water supply, irrigation, and hydropower generation compared to other river basins in the country. However, according to [44], the basin is frequently affected by severe drought in the last few decades. In this study, six streamflow stations were considered to analyze hydrological drought over the whole basin, and the result indicated that the most severe drought years were found in 1979, 1996, 1997, and 2001, whereas extreme droughts occurred in 1986, 1987, and 2002 as shown in Figure 5. In addition to [44], the result is supported by other previous studies, such as [45]. According to this researcher, 2002, 1965, and 1984 were the worst drought years;

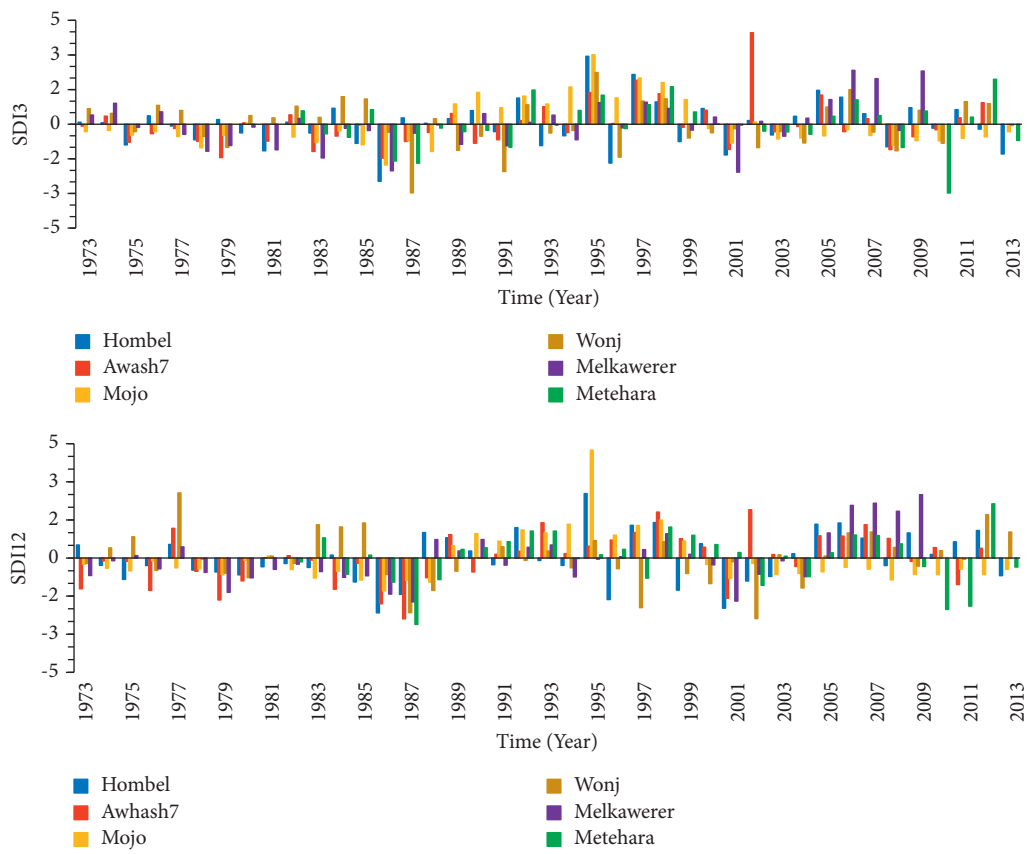


Figure 5: Temporal variation in seasonal (SDI3) and annual (SDI12) for Awash river basin.

99%, 97%, and 95% of the total basin area were affected by severe drought during the listed year, respectively. But during this analysis 1984/1985 is a moderate drought year and its magnitude is less relative to the Abbay river basin, but other drought years agreed with the previous studies.

Two streamflow stations (Gambela and Sore) were considered in the case of the Baro Akobo River basin drought analysis. This basin is received rainfall in two seasons, spring (MAM) and summer (JJA). So, the frequency of severe and extreme drought occurrence is less compared to other river basins in the country. In the Baro Akobo river basin, a drought analysis study is not conducted before this study. Results of this study show that the Baro Akobo basin was under moderate to severe drought conditions in the periods from 1980 to 1986 and 2001 to 2005 (seasonal and annual timescale analysis), and the most severe drought years of this basin occurred during the period of 1982/1983 and 1984/1985 and extreme drought years were 2002/2003 and 2004/2005, respectively, as shown in Figure 6.

Genale Dawa river basin is one of the drought-prone areas and receives less annual rainfall [46]. For the last few decades, drought analysis studies are not conducted in this basin, particularly hydrological drought analysis. Four streamflow stations were considered to identify the historical hydrological drought conditions of the basin during the period from 1981 to 2014 (34 years). As shown in Figure 7, SDI3 has high frequency and magnitude than the annual timescale (SDI12) in this basin. From 1981 to 1988, the basin was under mild-to-moderate drought conditions, whereas 1996 and 1999 were moderate drought years, and from 2001 to 2004, the basin was under moderate to severe drought. Previous meteorological studies concluded that the eastern and southeastern part of the country is frequently affected by the drought in which this basin is located [47–52]. However, meteorological drought results cannot be a good indicator of hydrological drought because the source of surface water for Genale and Wabishebele river basins is from the central part of the country, in which the flow travels a long distance in the arid area. Besides, the hydrological drought of this basin has been low if the central part of the country received good rainfall during the summer. The selected streamflow stations are from the upper part of the basin near the source. Therefore, in this study, the result generally shows that the basin highly experienced moderate to severe drought from the period of 1981to 2012.

Omo Gibe and Baro Akobo river basins are the wettest basins compared to other basins in the country. These basins received rainfall in two seasons (spring and summer). Assendabo and Gojeb streamflow stations were selected for Omo Gibe River basin hydrological drought analysis for the period from 1980 to 2017 (38 years). Results indicated that 1980–1981 and 2001 were severe drought years for Asseendabo station whereas 1986, 2002, and 2016 were for Gojeb station. Generally, this basin was under severe-moderate-mild drought conditions from 1980 to 1989 and 2001 to 2004, respectively, in the order of the years. Figure 8

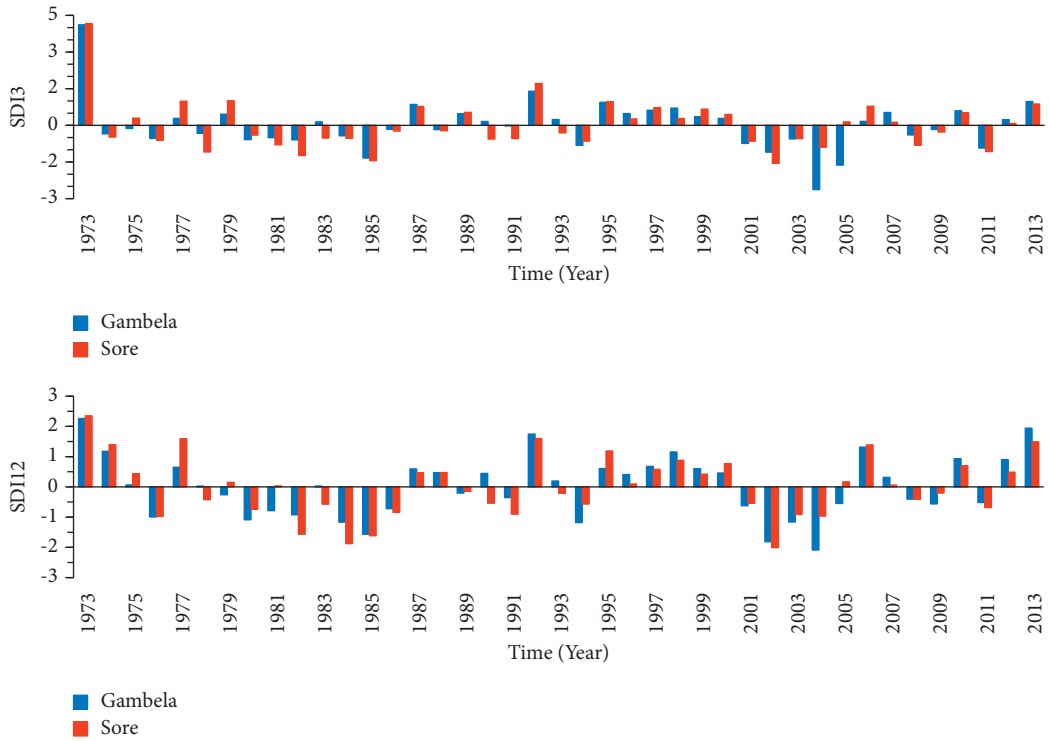


Figure 6: Temporal variation in seasonal (SDI3) and annual (SDI12) for Baro river basin.

Figure 7: Temporal variation in seasonal (SDI3) and annual (SDI12) for Genale Dawa river basin.

indicated that 1980–1987 and 2001 to 2004 were the most drought event year in the basin. But its severity is less than compared to Abbay river basins in the same drought event such as 1980 to 1987.

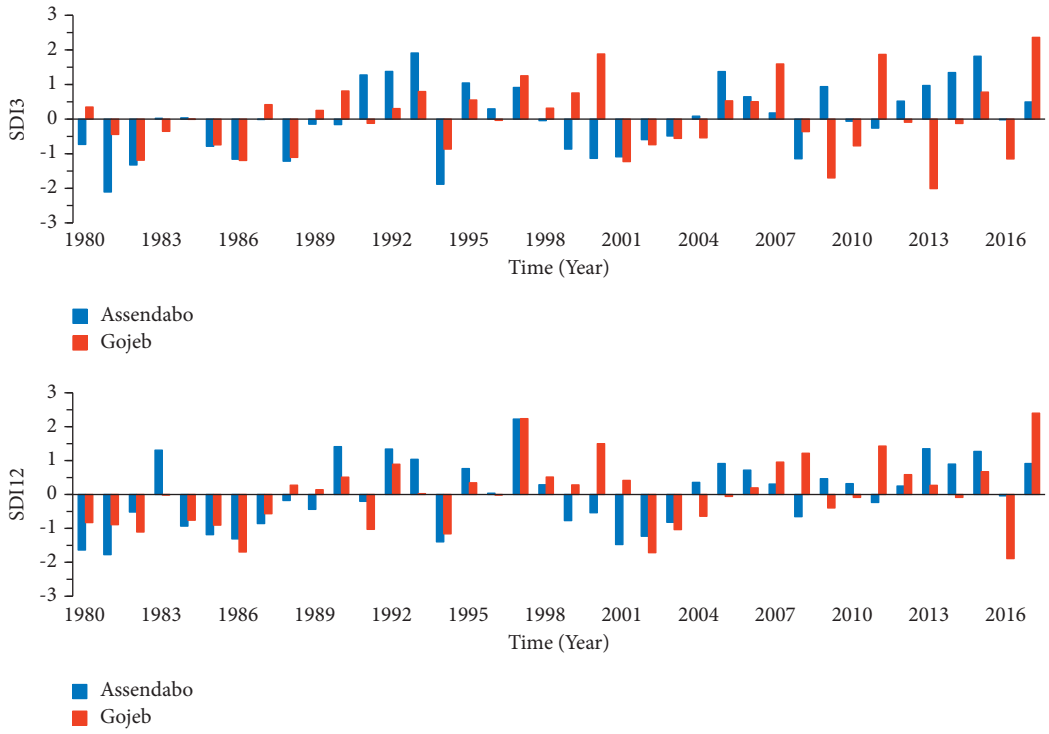


Figure 8: Temporal variation in seasonal (SDI3) and annual (SDI12) for Omo Gibe River basin.

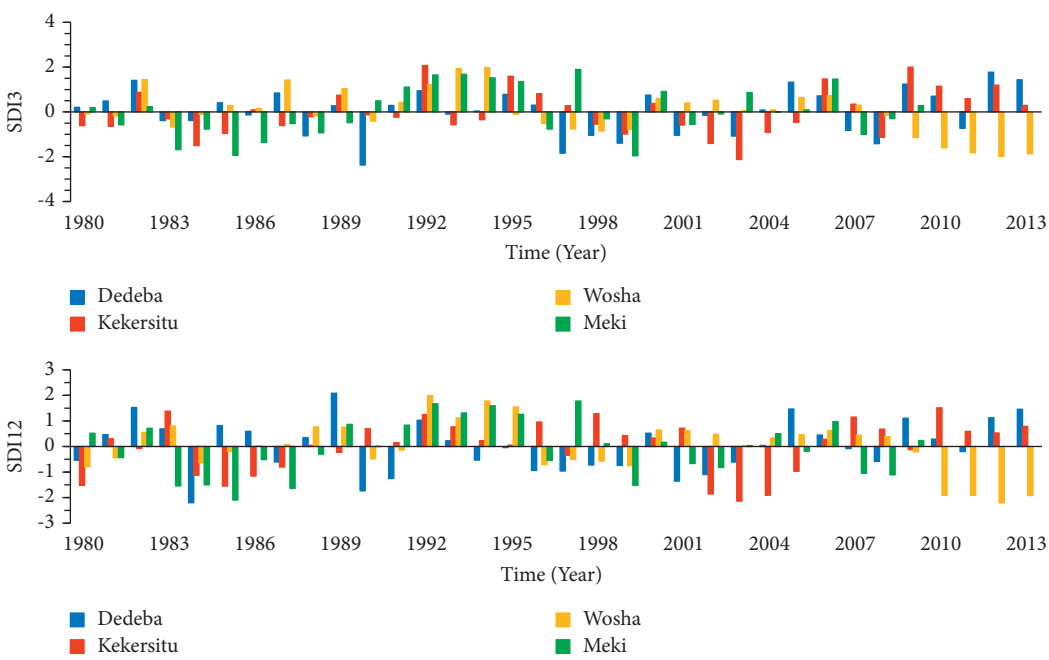


Figure 9: Temporal variation in seasonal (SDI3) and annual (SDI12) for Rift Valley River basin.

Relatively Rift Valley River basins have been frequently affected by severe drought compared to other basins in the last three decades (Figure 9). This result is agreed with [53] in which the basin is hit by moderate to severe droughts in more than 13 events in the last three decades. Particularly in this basin, the severity of drought increases from decade to decade. Figure 9 indicates that both seasonal and annual timescale analyses imply the occurrence of severe and extreme droughts over the three continuous decades (the 1980s, 1990s, and 2000s) increased. In particular, after 2010, the severity of drought at Wosha station increased because Wosha station is the most downstream station of the four considered stations in the basin for this study and which is nearer to arid and semiarid river basins such as Genale Dawa

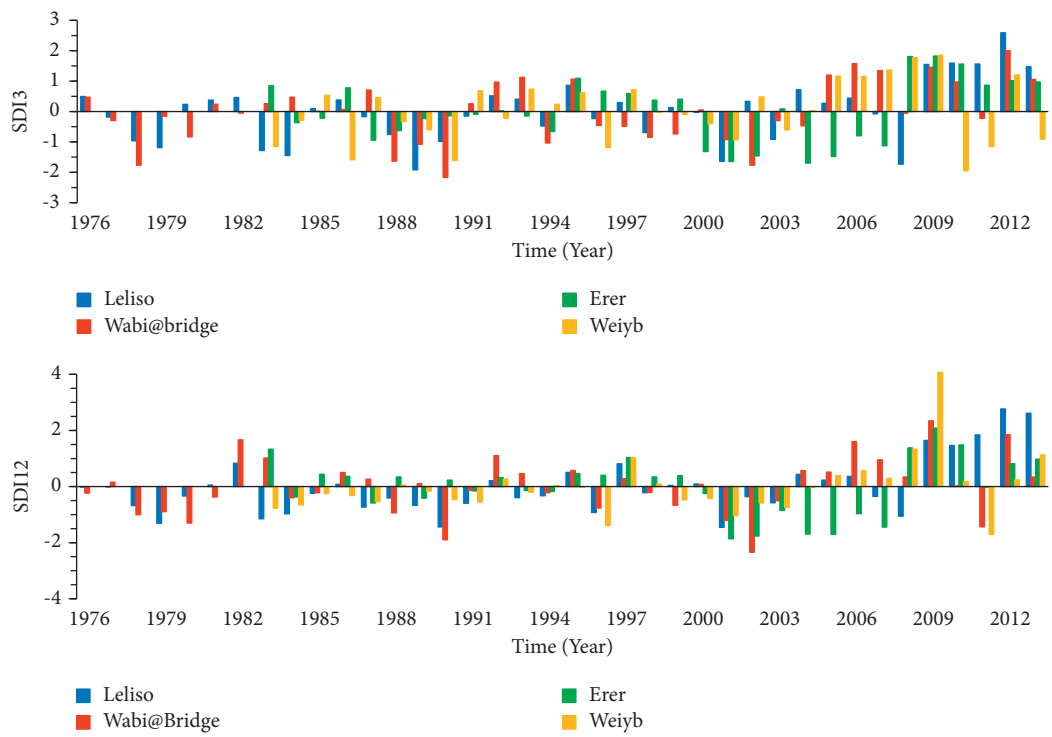


Figure 10: Temporal variation in seasonal (SDI3) and annual (SDI12) for the Wabishebele river basin.

and Wabishebele. Two stations from Genale Dawa river basin (Weib and Halewe) and two stations from Wabishebele river basin (Wabi@bridge and Weiyb) have also indicated the same drought condition after 2010 as Wosha (see Figures 7 and 10).

Drought analysis for the Tekeze basin was considered from 1994 to 2014 (21 years). Most flow stations located above Tekeze dam have no good data for drought analysis purposes. So, the Embamadre station, which is located below the dam, was considered for this study. After the completion of the Tekeze dam, the downstream flow is regulated starting from 2008 [54]. Therefore, for the last decade, the streamflow was increasing due to the conservation of flow by the Tekeze reservoir. As a result, the occurrence of severe as well as extreme drought is dramatically decreased. Figure 11 shows that the annual timescale has severe drought in 1996 whereas the seasonal timescale implies that a severe drought occurred in 2008 and 2009. But from 2001 to 2014, the result indicates wet years. Different studies reveal that the northern part of Ethiopia is frequently affected by drought [23, 47–49, 52, 55]. Therefore, this station is not a good representation for Tekeze basin hydrological drought analysis and it needs further investigation.

The Wabishebele river basin is one of the water-scarce basins in the country. While having the largest area coverage, its annual runoff and water availability are one of the lowest among the major river basins [56]. Before a decade, one drought study was conducted by Awass, 2009, in this river basin. His result shows that the most severe drought years of the basin were 1984/85, 1991/92, and 1998. However, in this study, the most severe and extreme drought events were 1986, 1988, 1989, 1990, 1991, 2001, 2002, 2004, 2005, and 2011. Figure 10 indicated that the magnitude of SDI3 was maximum compared to SDI12 for the above-listed drought years.

In arid and semiarid areas, the variation of streamflow is not significant because the low flow is the common characteristic of these areas. Therefore, moderate drought is more dominant than severe and extreme drought conditions for Genale Dawa and Wabishebele river basins, respectively (Figures 7 and 10).

In all river basins, the occurrence of extreme drought is decreased starting from 1990. For the case of Abbay, Awash, and Baro river basins, the analysis was considered for the long term (1974–2014) which results in different drought conditions, especially since the 1980s was the major drought event decade at the national level.

3.2. Hydrological Drought Characteristics. The hydrological drought characteristics of the 12-month (SDI12) timescale are shown in Tables 5–13. The result implies that the differences in duration (D), severity (S), magnitude (M), and relative frequency (RF) have high variation at different gauge stations within a basin (see Tables 6–13) but the average value of each drought characteristics has the relatively same value for all river basins (see Table 5). The maximum drought duration, severity, magnitude, and relative frequency are observed in the Abbay river basin at a period of 1980–1998 (19 years) at Megech station with 55.88% relative frequency. The hydrological drought condition of the Rift Valley River basin is agreed with Yisehak et.al. (2020); he was investigating drought characteristics at the Kulfo river and the result is almost the same as this finding [53].

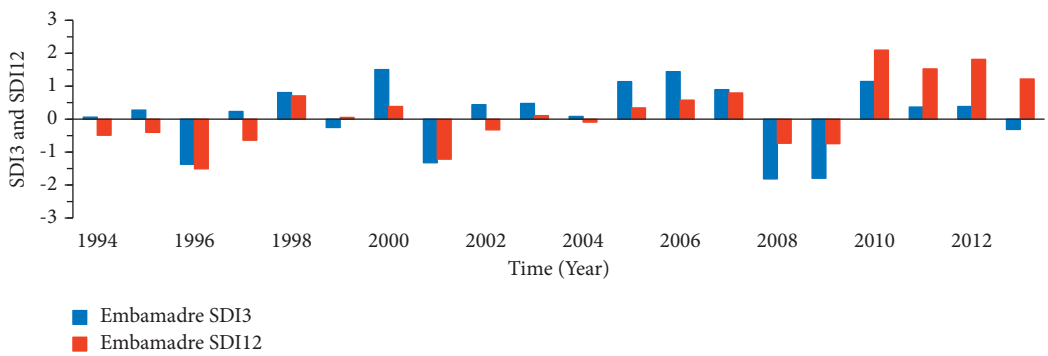


Figure 11: Temporal variation in seasonal (SDI3) and annual (SDI12) for the Tekeze river basin.

Table 5: Recapitulation of hydrological drought characteristics.

Basin	D (year)			S			M			RF (%)		
	Max	Min	Aver	Max	Min	Aver	Max	Min	Aver	Max	Min	Aver
Abbay	19	1	3.2	15.6	0.16	2.48	2.12	0.09	0.72	55.88	2.44	8.25
Awash	14	1	3.03	8.54	0.15	2.33	1.64	0.15	0.79	34.15	2.44	7.55
Baro	5	1	2.17	6.48	0.21	1.87	1.30	0.21	0.74	12.2	2.44	5.28
Genale Dawa	7	1	2.54	5.52	0.43	1.87	1.56	0.22	0.83	21.21	3.03	8.25
Omo Gibe	8	1	2.69	6.77	0.2	2.35	1.89	0.2	0.89	21.05	2.63	7.09
Rift valley	6	1	2.33	8.16	0.2	2.29	2.2	0.2	0.89	20	2.94	7.08
Tekeze	3	2	2.33	2.4	1.47	1.80	0.8	0.74	0.77	15.00	10.00	11.67
Wabishebele	8	1	2.43	10.52	0.22	1.82	1.7	0.16	0.70	25.81	2.63	7.03

Table 6: Hydrological drought characterization in Ethiopian major river basins, Tekeze river basin.

Station	Event	D (year)	S	M	RF (%)
Embamadre	1994–1996	3	2.4	0.80	15.00
	2001–2002	2	1.54	0.77	10.00
	2008–2009	2	1.47	0.74	10.00

3.3. Spatiotemporal Variability Analysis. The main important point in drought analysis and monitoring development is identifying the spatial extent of a drought over the river basin [36]. The drought-prone area identification was accomplished using ArcGIS software by applying a spatial analysis interpolation tool called inverse distance weighted (IDW). The result of SDI for the selected severe drought years was used as input for IDW in ArcGIS. The four severe and extreme drought years were selected to identify drought-prone areas over the region. Relatively, the common severe drought years for most river basins of the country were 1984/85, 1986/87, 2002/3, and 2010/11. The spatial variability of those identified drought years is shown in Figure 12. In 1984, the north and northwest parts of the country were affected by severe droughts (Tekeze, Abbay, and Baro) and some parts of the south and northeast parts were under moderate drought conditions (Rift Valley and Awash) (Figure 12(a)). In 1986, the central and south parts (Awash and Baro) of the country were affected by severe drought (Figure 12(b)). Severe drought was extended in the western part (Baro) and eastern part (Genale Dawa and Wabishebele) in 2002 (Figure 12(c)). But, in 2010, the drought was high in the Abbay basin and occurred in Genale Dawa and Rift Valley basins (Figure 12(d)).

Table 7: Hydrological drought characterization in Ethiopian major river basins, Omo Gibe River basin.

Station	Event	D (year)	S	M	RF (%)
Assendabo	1980–1982	3	3.93	1.31	7.89
	1984–1989	6	4.9	0.82	15.79
	1991	1	0.2	0.20	2.63
	1994	1	1.4	1.40	2.63
	1999–2003	5	4.48	0.90	13.16
	2008	1	0.66	0.66	2.63
	2011	1	0.23	0.23	2.63
Gojeb	1980–1987	8	6.77	0.85	21.05
	1991	1	1.03	1.03	2.63
	1994	1	1.16	1.16	2.63
	2002–2005	4	3.45	0.86	10.53
	2009–2010	2	0.49	0.25	5.26
	2016	1	1.89	1.89	2.63

Table 8: Hydrological drought characterization in Ethiopian major river basins, Abbay river basin.

Station	Event	D (year)	S	M	RF (%)
Bahir Dar	1973	1	0.28	0.28	2.44
	1978–1987	10	7.66	0.77	24.39
	1989–1990	2	0.48	0.24	4.88
	1995	1	1.17	1.17	2.44
	2002–2005	4	1.78	0.45	9.76
	2009	5	5.7	1.14	12.20
Border	1973	1	0.18	0.18	2.44
	1975–1976	2	1.18	0.59	4.88
	1979	1	1.62	1.62	2.44
	1981–1991	11	11.46	1.04	26.83
	2002–2004	3	1.05	0.35	7.32
	2008	1	0.33	0.33	2.44
G. Beles	1981–1985	5	6.47	1.29	12.20
	1991	1	2.12	2.12	2.44
	1993–1997	5	2.01	0.40	12.20
	2001	1	0.72	0.72	2.44
	2003–2004	2	0.2	0.10	4.88
	2006–2013	8	6.53	0.82	19.51
Gummera	1973–1984	12	10.7	0.89	29.27
	1986–1987	2	1.08	0.54	4.88
	1989–1992	4	1.63	0.41	9.76
	1994	1	0.28	0.28	2.44
	1998	1	0.88	0.88	2.44
	2001–2002	2	0.62	0.31	4.88
	2004	1	1	1.00	2.44
Guder	1973	1	0.32	0.32	2.78
	1975–1976	2	2.27	1.14	5.56
	1978–1981	4	3.1	0.78	11.11
	1984–1985	2	0.65	0.33	5.56
	1987	1	1.31	1.31	2.78
	1990	1	0.58	0.58	2.78
	1994	1	1.52	1.52	2.78
	1996–1997	2	1.51	0.76	5.56
	1999	1	0.18	0.18	2.78
	2001–2004	4	3.82	0.96	11.11
G. Abbay	1977	1	0.49	0.49	2.56
	1982	1	1.09	1.09	2.56
	1985–1987	3	2.85	0.95	7.69
	1989	1	1.75	1.75	2.56
	1991	1	0.72	0.72	2.56
	1994	1	1.69	1.69	2.56
	2001–2004	5	5.32	1.06	12.82
	2008–2011	4	2.15	0.54	10.26
Koga	1975	1	0.35	0.35	2.56
	1977	1	0.24	0.24	2.56
	1981–1990	10	8.69	0.87	25.64
	1992	1	0.31	0.31	2.56
	1995–1996	2	0.17	0.09	5.13
	2000–2001	2	0.98	0.49	5.13
	2008–2001	4	4.08	1.02	10.26
Ribb	1978–1984	7	8.7	1.24	17.95
	1986	1	1.47	1.47	2.56
	1989	1	0.96	0.96	2.56
	1994	1	0.16	0.16	2.56
	1996–1997	2	0.9	0.45	5.13
	2001–2004	4	2.46	0.62	10.26
	2010–2011	2	0.71	0.36	5.13

Table 8: Continued.

Station	Event	D (year)	S	M	RF (%)
Chemoga	1973–1975	3	1.7	0.57	8.33
	1981–1982	2	0.97	0.49	5.56
	1984–1985	2	0.66	0.33	5.56
	1989	1	0.34	0.34	2.78
	1991	1	0.49	0.49	2.78
	1994	1	1.63	1.63	2.78
	1996–1997	2	1.06	0.53	5.56
	2001–2008	8	7.92	0.99	22.22
Megech	1980–1998	19	13.01	0.68	55.88
	2001–2003	3	0.91	0.30	8.82
Kessie	1973–1974	2	0.25	0.13	4.88
	1976–1992	17	15.6	0.92	41.46
	2010	1	0.23	0.23	2.44

Table 9: Hydrological drought characterization in Ethiopian major river basins, Baro Akobo river basin.

Station	Event	D (year)	S	M	RF (%)
Gambela	1976	1	0.99	0.99	2.44
	1979–1982	4	3.06	0.77	9.76
	1984–1986	3	3.46	1.15	7.32
	1989	1	0.21	0.21	2.44
	1991	1	0.35	0.35	2.44
	1994	1	1.18	1.18	2.44
	2001–2005	5	6.23	1.25	12.20
	2008–2009	2	0.96	0.48	4.88
	2011	1	0.52	0.52	2.44
Sorie	1976	1	0.98	0.98	2.44
	1978	1	0.43	0.43	2.44
	1980	1	0.74	0.74	2.44
	1982–1986	5	6.48	1.30	12.20
	1989–1991	3	1.59	0.53	7.32
	1993–1994	2	0.78	0.39	4.88
	2001–2004	4	4.43	1.11	9.76
	2008–2009	2	0.62	0.31	4.88
	2011	1	0.69	0.69	2.44

Table 10: Hydrological drought characterization in Ethiopian major river basins, Awash river basin.

Station	Event	D (year)	S	M	RF (%)
Awash7	1973–1976	4	2.74	0.685	10
	1978–1981	4	3.1	0.775	10
	1984–1988	5	6.42	1.284	12.5
	1990	1	0.55	0.55	2.5
	2001	1	1.58	1.58	2.5
	2011	1	1.05	1.05	2.5
Hombel	1974–1976	3	1.43	0.48	7.32
	1978–1983	6	2.64	0.44	14.63
	1985–1987	3	4.53	1.51	7.32
	1993–1994	2	0.39	0.195	4.88
	1996	1	1.64	1.64	2.44
	1999	1	1.27	1.27	2.44
	2001–2003	3	3.62	1.21	7.32
Melka	1973–1974	2	0.8	0.40	5.41
	1976	1	0.43	0.43	2.70
	1978–1987	10	8.54	0.85	27.03
	1991	1	0.28	0.28	2.70
	1994–1995	2	0.8	0.40	5.41
	2000–2004	5	3.44	0.69	13.51
Wonj	1973	1	0.22	0.22	2.44
	1976	1	0.48	0.48	2.44
	1978–1980	3	1.43	0.48	7.32
	1986–1990	5	4.61	0.92	12.20
	1996–1997	2	2.37	1.19	4.88
	1999–2002	4	4.15	1.04	9.76
	2004	1	1.18	1.18	2.44
	2009	1	0.32	0.32	2.44
Metehara	1982	1	0.15	0.15	2.44
	1984	1	0.63	0.63	2.44
	1986–1988	3	4.47	1.49	7.32
	1997	1	0.79	0.79	2.44
	2002	1	1.08	1.08	2.44
	2004	1	0.73	0.73	2.44
	2009–2011	3	4.27	1.42	7.32
	2013	1	0.36	0.36	2.44
Mojo	1973–1980	8	2.83	0.35	19.51
	1982–1988	7	5.77	0.82	17.07
	2000–2013	14	7.59	0.54	34.15

4. Discussion

This study assessed the historical hydrological drought conditions in Ethiopia from 1973 to 2014. The analysis was computed using the streamflow drought index (SDI) on a 3-month and 12-month timescale for 34 streamflow stations in all basins. The results indicate that the occurrence of severe and extreme drought events declined from decade to decade, especially in the last two decades the 2000s and 2010s compared to earlier decades (1970s, 1980s, and 1990s). However, less severe and extreme drought events occurred from 2001 to 2004 in some river basins, particularly, Baro, Rift Valley, Genale Dawa, and Wabishebele river basins. Even though there is a limitation to hydrological drought analysis at a national level, some previous meteorological drought studies supported this finding such as [55].

This study also suggests that all river basins in Ethiopia experienced some degree of drought at both seasonal (3 months) and annual (12 months) timescales. The most severe and extreme drought events occurred in the period from 1981 to 1987 and 2002 to 2004. This result is also

Table 11: Hydrological drought characterization in Ethiopian major river basins, Genale Dawa river basin.

Station	Event	D (year)	S	M	RF (%)
Halawel	1984–1985	2	0.81	0.41	6.90
	1987	1	0.64	0.64	3.45
	1990–1991	2	1.68	0.84	6.90
	1993–1994	2	0.83	0.42	6.90
	1998–1999	2	1.56	0.78	6.90
	2001–2005	5	3.91	0.78	17.24
	2007–2008	2	1.46	0.73	6.90
	2012	1	0.8	0.80	3.45
Shaya	1981–1987	7	4.12	0.59	21.21
	1989–1991	3	1.43	0.48	9.09
	1993–1994	2	0.43	0.22	6.06
	1996	1	1.52	1.52	3.03
	1999–2005	7	5.52	0.79	21.21
Weib	1984–1988	5	1.36	0.27	16.67
	1990–1991	2	1.44	0.72	6.67
	1996	1	0.89	0.89	3.33
	1999	1	1.27	1.27	3.33
	2001–2003	3	3	1.00	10.00
	2008	1	1.38	1.38	3.33
	2010–2011	2	3.11	1.56	6.67
Chenemsa	1984–1988	5	2.31	0.46	16.13
	1990–1991	2	1.82	0.91	6.45
	1996	1	1.42	1.42	3.23
	1999	1	1.17	1.17	3.23
	2001–2003	3	3.85	1.28	9.68
	2007–2008	2	0.78	0.39	6.45

Table 12: Hydrological drought characterization in Ethiopian major river basins, Rift Valley River basin.

Station	Event	D (year)	S	M	RF (%)
Dedeba	1980	1	0.54	0.54	2.94
	1984	1	2.2	2.20	2.94
	1987	1	0.61	0.61	2.94
	1990–1991	2	3	1.50	5.88
	1994–1999	6	3.94	0.66	17.65
	2001–2003	3	3.07	1.02	8.82
	2007–2008	2	0.66	0.33	5.88
	2011	1	0.2	0.20	2.94
Kekersitu	1980	1	1.52	1.52	2.94
	1984–1987	4	4.66	1.17	11.76
	1989	1	0.23	0.23	2.94
	1997	1	0.35	0.35	2.94
	2002–2005	4	6.88	1.72	11.76
Meki	1981	1	0.44	0.44	3.33
	1983–1988	6	7.59	1.27	20.00
	1996	1	0.54	0.54	3.33
	1999	1	1.52	1.52	3.33
	2001–2002	2	1.49	0.75	6.67
	2007–2008	2	2.16	1.08	6.67
Wosha	1980–1981	2	1.25	0.63	5.88
	1984–1985	2	0.84	0.42	5.88
	1990–1991	2	0.62	0.31	5.88
	1996–1999	4	2.54	0.64	11.76
	2009–2013	5	8.16	1.63	14.71

assured by [18]. The proportional area affected by mild and moderate drought showed a statically increasing trend over the 42 years. The areas subjected to the droughts of high intensity (i.e., severe and extreme) also exhibited a statistically significant decrease with the peak of the decline from the 1990s and 2000s. These findings imply that the decrease in drought frequency in the recent decade is associated more with a decrease in the frequency of intense drought episodes. The overall analysis also indicates anon—synchronous occurrence of drought episodes over the country, whereas some severe drought episodes were witnessed in specific periods, while no droughts were corresponding to these episodes across other regions of the study area.

Table 13: Hydrological drought characterization in Ethiopian major river basins, Wabishebele river basin.

Station	Event	D (year)	S	M	RF (%)
Leliso	1978–1980	3	2.05	0.68	7.89
	1983–1985	3	2.36	0.79	7.89
	1987–1991	5	3.84	0.77	13.16
	1993–1994	2	0.73	0.37	5.26
	1996	1	0.92	0.92	2.63
	1998	1	0.22	0.22	2.63
	2001–2003	3	2.39	0.80	7.89
	2007–2008	2	1.41	0.71	5.26
Erer	1984	1	0.37	0.37	3.23
	1987	1	0.59	0.59	3.23
	1989	1	0.41	0.41	3.23
	1993–1994	2	0.32	0.16	6.45
	2000–2007	8	10.52	1.32	25.81
Wabi	1976	1	0.23	0.23	2.63
	1978–1981	4	3.49	0.87	10.53
	1984–1985	2	0.66	0.33	5.26
	1988	1	0.94	0.94	2.63
	1990–1991	2	2.04	1.02	5.26
	1994	1	0.22	0.22	2.63
	1996	1	0.77	0.77	2.63
	1998–199	2	0.88	0.44	5.26
	2001–2003	3	4.06	1.35	7.89
	2011	1	1.43	1.43	2.63
Weyb	1983–1987	5	2.54	0.51	16.13
	1989–1991	3	1.18	0.39	9.68
	1995–1996	2	1.41	0.71	6.45
	1999–2004	6	3.3	0.55	19.35
	2011	1	1.7	1.70	3.23

The spatial variability of hydrological drought occurrence has a wide range from basin to basin. These different drought occurrence patterns are related to diverse geographical areas and are associated with different temporal variations. It is observed that there is a strong relationship between streamflow and drought potential basins in the regions [20], where the most significant pattern of the drought was specifically found over the arid areas, while less significant patterns were observed over the most humid regions.

The 1980s was the driest year in all river basins. At a country level, four severe to extreme drought years were identified, 1984/1985, 1986/1987, 2002/2003, and 2010/2011 from 1973 to 2014. However, only 1984/1985 was nationally as well as globally get attention, but other severe drought years hindered the economic growth of the country in the last three decades.

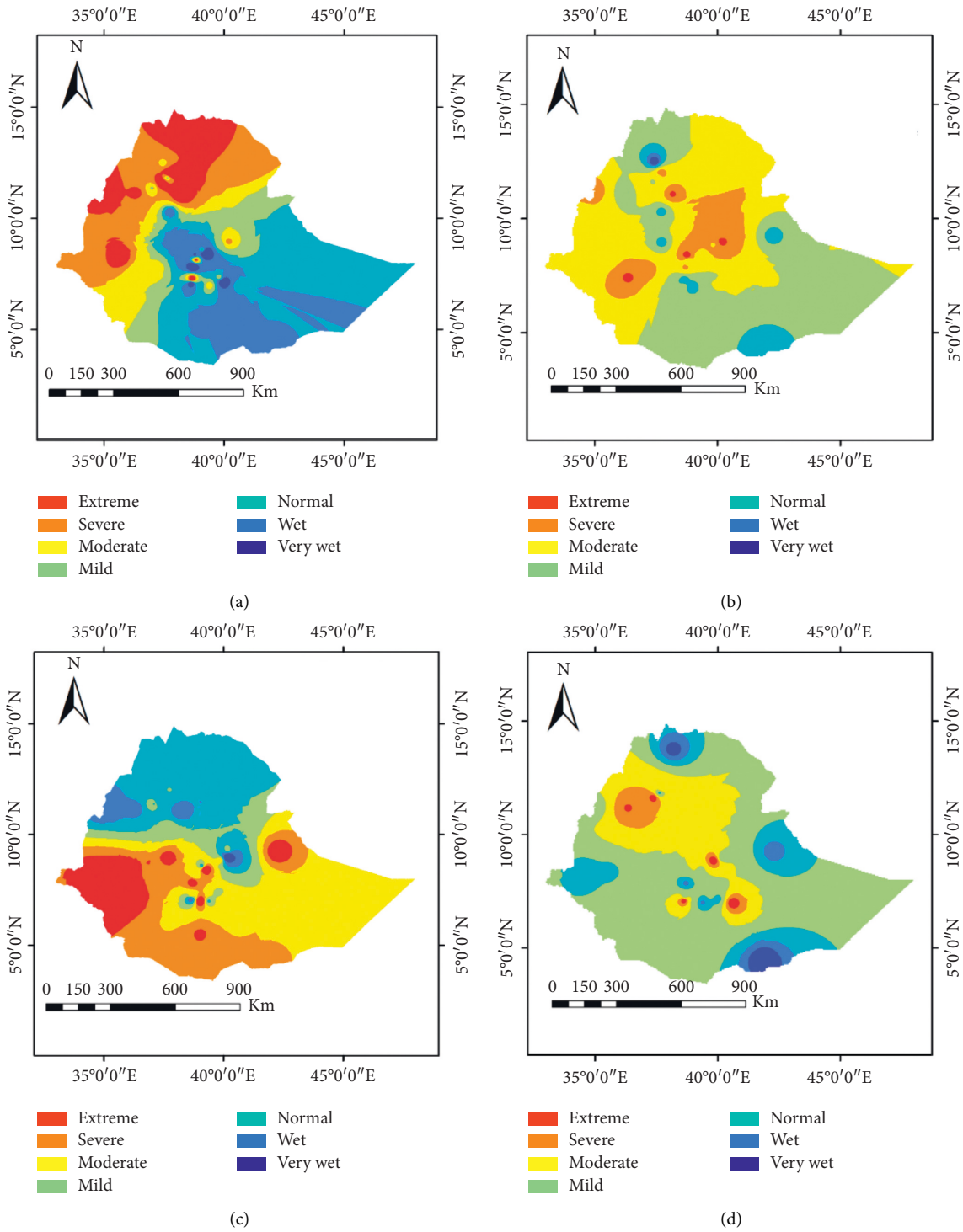


Figure 12: Spatial variation of streamflow drought in Ethiopia in four severe drought years. (a). SDI 1984. (b). SDI 1986. (c). SDI 2002. (d). SDI 2010.

5. Conclusion

In Ethiopia, many sectors such as agriculture, water supply, and hydropower are highly affected by hydrological drought in which most human activities depend on these sectors, but the hydrological drought is not well studied in the country. Particularly, Baro Akobo, Omo Gibe, Genale Dawa, and Rift Valley River basins are not been studied before. Abbay, Awash, Tekeze, and Wabishebele river basins are partially studied, more focused on meteorological drought than other droughts. Therefore, this study analyzed hydrological drought in Ethiopia using SDI to identify the most severe and extreme drought years and drought-prone areas. Accordingly, 1979, 1981, 1983, 1987, 1996, 1997, and 2001 were

the most severe drought years, whereas 1984, 1986, 2002, and 2010 were relatively extreme drought years in all river basins in Ethiopia from 1973 to 2014. However, globally, only the 1984 severe drought event was popularly recognized and other drought years were not focused on but affected food security, water supply, and hydropower production system over the country. Abbay, Awash, and Rift Valley River basins were frequently affected by severe streamflow drought, but for the case of Baro Akobo and Omo Gibe River basins, the frequency of severe drought was less. Genale Dawa and Wabishebele river basins were located in arid and semiarid areas in which drought conditions are dominated by their aridity nature and the occurrence of severe and extreme drought events was rare. But mild and moderate drought events frequently occur in these basins. Generally, severe and extreme drought frequency is relatively decreased for all river basins in the 1990s and 2000s compared to the 1970s and 1980s. The source of streamflow for all river basins is from highlands and mountainous areas. Therefore, most of the streamflow stations considered for this study were from the upstream and middle parts of the basins. Therefore, the drought conditions for these stations can represent the downstream parts of the basins. Generally, Abbay and Awash river basins are the most populated basins and cover large investment areas in which the investigation of hydrological drought and development of drought early warning systems is important. Therefore, this study gives good information about hydrological drought conditions for all river basins for decision-makers for good mitigation measurement development.

Data Availability

This research is not completed and the author is working on future forecasting of hydrological drought conditions in Ethiopia using the data used in this article. Therefore, the data will be provided after completing the forecasting analysis. But at this level, the data are not provided due to future extension of the research. So, the data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no known conflicts of financial interest or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The corresponding author acknowledges the Ethiopian Ministry of Water and Energy (MoWE) for their support of hydrological data (Streamflow)

References

[1] R. Shah, V. L. Manekar, R. A. Christian, and N. J. Mistry, "Estimation of reconnaissance drought index," *International Journal of. Environmental and Ecological Engineering*, vol. 7, pp. 1719–1722, 2013.

[2] D. C. Edossa, Y. E. Woyessa, and W. A. Welderufael, "Analysis of droughts in the central region of South Africa and their association with SST anomalies," *International Journal of Atmospheric Sciences*, vol. 2014, Article ID 508953, 8 pages, 2014.

[3] A. Asadi and S. F. Vahdat, "The efficiency of meteorological drought indices for drought Monitoring and evaluating in kohgilouye and boyerahmad Province, Iran," *International Journal of Modern Engineering Research*, vol. 3, pp. 2407–2411, 2013.

[4] C. Svensson, J. Hannaford, and I. Prosdocimi, "Statistical distributions for monthly aggregations of precipitation and streamflow in drought indicator applications," *Water Resources Research*, vol. 53, no. 2, pp. 999–1018, 2017.

[5] S. Giri, A. Mishra, Z. Zhang, R. G. Lathrop, and A. O. Alnahit, "Meteorological and hydrological drought analysis and its impact on water quality and stream integrity," *Sustainability*, vol. 13, pp. 1–24, 2021.

[6] P. Alegre, E. T. De Jesus, R. Junqueira, and M. R. Viola, "Meteorological and hydrological drought from 1987 to 2017 in Doce river basin, southeastern Brazil," *Revista Brasileira de Recursos Hidricos*, vol. 25, pp. 1–10, 2020.

[7] J. Nasir, E. Assefa, T. Zeleke, and E. Gidey, "Meteorological drought in northwestern escarpment of Ethiopian Rift valley: detection seasonal and spatial trends," *Environmental Systems Research*, vol. 10, no. 1, 2021.

[8] R. Kolachian and B. Saghafian, "Hydrological drought class early warning using support vector machines and rough sets," *Environmental Earth Sciences*, vol. 80, no. 11, pp. 1–15, 2021.

[9] Z. A. Alemu, E. C. Dioha, E. C. Dioha, and M. O. Dioha, "Hydro-meteorological drought in Addis Ababa: a characterization study," *AIMS Environmental Science*, vol. 8, no. 2, pp. 148–168, 2021.

[10] C. M. Botai, J. O. Botai, J. P. de Wit et al., "Hydrological drought assessment based on the standardized streamflow index: a case study of the three cape provinces of South Africa," *Water*, vol. 13, 2021.

[11] J. A. Keyantash and J. A. Dracup, "An aggregate drought index: assessing drought severity based on fluctuations in the hydrologic cycle and surface water storage," *Water Resources Research*, vol. 40, no. 9, pp. 1–14, 2004.

[12] D. T. Tsige, V. Uddameri, and F. Forghanparast, "Agriculture-Related Drought Indicators," *Water*, vol. 11, no. 2218, pp. 1–24, 2019.

[13] S. Kruse and I. Seidl, "Social capacities for drought risk management in Switzerland," *Natural Hazards and Earth System Sciences*, vol. 13, no. 12, pp. 3429–3441, 2014.

[14] J. Juana, "Socioeconomic impact of drought in Botswana the socio-economic impact of drought in Botswana," *International Journal of Environment and Sustainable Development*, vol. 11, 2015.

[15] B. Bąk and K. Kubiak-Wójcicka, "Assessment of meteorological and hydrological drought in Torun (central Poland town) in 1971–2010 based on standardized indicators," in *Proceedings of the 3rd International Conference Water Resources and Wetlands Conference Proceedings*, pp. 164–170, Tulcea Romania, 2014.

[16] A. F. Van Loon, "Hydrological drought explained," *Wiley Interdisciplinary Reviews: Water*, vol. 2, no. 4, pp. 359–392, 2015.

[17] S. Liu, J. Zhang, N. Wang, and N. Wei, "Large-scale linkages of socioeconomic drought with climate variability and its evolution characteristics in northwest China," *Advances in Meteorology*, vol. 2020, Article ID 2814539, 13 pages, 2020.

[18] Y. Bayissa, S. Maskey, T. Tadesse et al., "Comparison of the performance of six drought indices in characterizing historical drought for the upper blue nile basin, Ethiopia," *Geosciences*, vol. 8, no. 3, p. 81, 2018.

[19] B. Boudad, H. Sahbi, and I. Mansouri, "Analysis of meteorological and hydrological drought based in SPI and SDI index in the Inaouen basin (Northern Morocco)," *Journal of Materials and Environmental Science*, vol. 9, no. 1, pp. 219–227, 2018.

[20] H. H. Hasan, S. F. M. Razali, N. S. Muhammad, and A. Ahmad, "Research trends of hydrological drought: a systematic Review," *Water*, vol. 11, pp. 1–19, 2019.

[21] C. Rosenzweig and D. Hillel, "Climate variability and the global harvest," *Impacts of El Niño and other Oscillations on Agro-Ecosystems*, Oxford University Press, Oxford, UK, 2008.

[22] D. Guha-Sapir and P. W. Philippe Hoyois, "Below regina annual disaster statistical review 2016: the numbers and trends," *Review Literature and Arts of the Americas*, pp. 1–50, 2016, http://www.cred.be/sites/default/files/ADSR_2016.pdf.

[23] Y. Mohammed, F. Yimer, M. Tadesse, and K. Tesfaye, "Meteorological drought assessment in north east highlands of Ethiopia," *International Journal of Climate Change Strategies and Management*, vol. 10, no. 1, pp. 142–160, 2018.

[24] H. A. J. Van Lanen, N. Wanders, L. M. Tallaksen, and A. F. Van Loon, "Hydrological drought across the world: impact of climate and physical catchment structure," *Hydrology and Earth System Sciences*, vol. 17, no. 5, pp. 1715–1732, 2013.

[25] T. Gebrehiwot, A. van der Veen, and B. Maathuis, "Spatial and temporal assessment of drought in the northern highlands of Ethiopia," *International Journal of Applied Earth Observation and Geoinformation*, vol. 13, no. 3, pp. 309–321, 2011.

[26] S. Barua, "Drought assessment and forecasting using a nonlinear aggregated drought index," Victoria Unversity, Melbourne Australia, M.Sc, 2010.

[27] D. S. Nairizi, "Irrigated agriculture development under drought and water scarcity," *International Commission on Irrigation and Drainage.*, pp. 1–168, 2017.

[28] A. Cancelliere, G. Giuliano, V. Nicolosi, and G. Rossi, "Optimal short-term operation of a multipurpose reservoir system under limited water supply," in *Proceedings of the Water Resources Systems-Hydrological Risk. Management and Development*, pp. 200–207, Sapporo, Japan, 2003.

[29] J. Wu, Z. Liu, H. Yao et al., "Impacts of reservoir operations on multi-scale correlations between hydrological drought and meteorological drought," *Journal of Hydrology*, vol. 563, pp. 726–736, 2018.

[30] L. J. Barker, J. Hannaford, S. Parry, K. A. Smith, M. Tanguy, and C. Prudhomme, "Historic hydrological droughts 1891–2015: systematic characterisation for a diverse set of catchments across the UK," *Hydrology and Earth System Sciences*, vol. 23, no. 11, pp. 4583–4602, 2019.

[31] J. M. Davis and M. Stampone, *Analysis of Drought Conditions in New Hampshire*, University of New Hampshire, Durham, NH, USA, 2014.

[32] I. Nalbantis, "Evaluation of a hydrological drought index," *Environmental Science*, vol. 23, pp. 67–77, 2008.

[33] I. Nalbantis and G. Tsakiris, "Assessment of hydrological drought revisited," *Water Resources Management*, vol. 23, no. 5, pp. 881–897, 2009.

[34] M. Eskandaripour and S. Soltaninia, "Analyzing the duration frequency and severity of drought using copula function in the Yazd city," *Journal of Water and Climate Change*, vol. 13, no. 1, pp. 67–82, 2022.

[35] Y. Soo, J. So, and H. Soo, "Drought severity-duration-frequency analysis of hydrological drought based on copula theory," pp. 82–89, 2012.

[36] B. Yisehak, H. Shiferaw, H. Abrha et al., "Spatio-temporal characteristics of meteorological drought under changing climate in semi-arid region of northern Ethiopia," *Environmental Systems Research*, vol. 10, no. 1, 2021.

[37] G. Buttafuoco and T. Caloiero, "Drought events at different timescales in southern Italy (Calabria)," *Journal of Maps*, vol. 10, no. 4, pp. 529–537, 2014.

[38] J. I. Christian, J. B. Basara, A. Otkin, S. J. Sutanto, F. Wetterhall, and H. A. J. Van Lanen, "Hydrological drought forecasts outperform meteorological drought forecasts," *Environmental Research Letters*, vol. 15, 2020.

[39] D. Tigkas, H. Vangelis, and G. Tsakiris, "The drought indices calculator (DrinC)," in *Proceedings of the 8th International EVRA Conference*, pp. 1333–1342, Porto, Portugal, 2013.

[40] D. Tigkas, H. Vangelis, and G. Tsakiris, "DrinC: a software for drought analysis based on drought indices," *Earth Science India*, vol. 8, no. 3, pp. 697–709, 2015.

[41] A. Shahidi, "Assess the impact of climate change on the hydrological drought in southern Khorasan province, Iran," in *Proceedings of the 6th International Conference on Informatics, Environment, Energy and Applications*, pp. 38–42, New York, NY, USA, 2017.

[42] R. M. Wambua, B. M. Mutua, and J. M. Raude, "Hydrological drought frequency estimation using stream flow drought index and modified gumbel method in upper Tana river basin," *International Journal of Current Research and Review*, vol. 7, 2015.

[43] Y. A. Bayissa, S. A. Moges, Y. Xuan et al., "Spatio-temporal assessment of meteorological drought under the influence of varying record length: the case of upper Blue Nile basin, Ethiopia," *Hydrological Sciences Journal*, vol. 60, pp. 1–16, 2015.

[44] A. Belayneh, J. Adamowski, B. Khalil, and B. Ozga-Zielinski, "Long-term SPI drought forecasting in the Awash river basin in Ethiopia using wavelet neural network and wavelet support vector regression models," *Journal of Hydrology*, vol. 508, pp. 418–429, 2014.

[45] D. C. Edossa, M. S. Babel, and A. Das Gupta, "Drought analysis in the Awash river basin, Ethiopia," *Water Resources Management*, vol. 24, no. 7, pp. 1441–1460, 2010.

[46] N. Kassahun and M. Mohamed, "Groundwater potential assessment and characterization of Genale-Dawa river basin," *Open Journal of Modern Hydrology*, vol. 8, no. 04, pp. 126–144, 2018.

[47] T. T. Zeleke, F. Giorgi, G. T. Diro, and B. F. Zaitchik, "Trend and periodicity of drought over Ethiopia," *International Journal of Climatology*, vol. 37, no. 13, pp. 4733–4748, 2017.

[48] I. Masih, S. Maskey, F. E. F. Mussá, and P. Trambauer, "A review of droughts in the African continent: a geospatial and long-term perspective," *Hydrology and Earth System Sciences Discussions*, vol. 11, 2014.

[49] G. A. Mera, "Drought and its impacts in Ethiopia," *Weather and Climate Extremes*, vol. 22, pp. 24–35, 2018.

[50] G. Gebremeskel Haile, Q. Tang, S. Sun, Z. Huang, X. Zhang, and X. Liu, "Droughts in east Africa: causes, impacts and resilience," *Earth-Science Reviews*, vol. 193, pp. 146–161, 2019.

[51] D. Temam, V. Uddameri, G. Mohammadi, and E. A. Hernandez, "Long-term drought trends in Ethiopia with implications for Dryland agriculture," *Water*, vol. 11, pp. 1–22, 2019.

[52] A. Teshome, "Increase of extreme drought over Ethiopia under climate warming," *Advances in Meteorology*, vol. 2019, Article ID 5235429, 18 pages, 2019.

[53] B. Yisehak, K. Adhena, H. Shiferaw, H. Hagos, H. Abrha, and T. Bezabh, "Characteristics of hydrological extremes in Kulfo river of southern Ethiopian Rift valley basin," *SN Applied Sciences*, vol. 2, no. 7, pp. 1–12, 2020.

[54] F. Fentaw, D. Hailu, A. Nigussie, and A. M. Melesse, "Climate change impact on the hydrology of Tekeze basin, Ethiopia: projection of rainfall-runoff for future water resources planning," *Water Conservation Science and Engineering*, vol. 3, no. 4, pp. 267–278, 2018.

[55] A. M. El Kenawy, M. F. McCabe, S. M. Vicente-Serrano, J. I. López-Moreno, and S. M. Robaa, "Changes in the frequency and severity of hydrological droughts over Ethiopia from 1960 to 2013," *Cuadernos de Investigación Geográfica*, vol. 42, no. 1, pp. 145–166, 2016.

[56] A. A. Awass, "Hydrological drought analysis—occurrence, severity, risks: the case of Wabishebele river basin," Dissertation, University of Siegen, Siegen, Germany, 2009.

Hindawi
Advances in Meteorology
Volume 2022, Article ID 6953306, 13 pages
https://doi.org/10.1155/2022/6953306

Research Article

Impact of Hydroclimate Change on the Management for the Multipurpose Reservoir: A Case Study in Meishan (China)

Yang Liu [1] **and Buwei Wang** [2]

[1] *Key Laboratory of Land Surface Pattern and Simulation, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China*
[2] *China Electric Power Research Institute Co.Ltd, Beijing 100192, China*

Correspondence should be addressed to Yang Liu; liu_yang@igsnrr.ac.cn

Received 23 November 2021; Revised 10 March 2022; Accepted 15 March 2022; Published 30 March 2022

Academic Editor: Upaka Rathnayake

China holds the largest amount of reservoirs in the world, while more than 80% of them were constructed 50–70 years ago and are approaching a critical stage of their designed lifetime. Before deciding the future of a reservoir, it is essential to find out whether it could still satisfy its original purpose in the context of hydroclimate change under global warming. Here, we present a case study of the Meishan reservoir in east-central China, which was primarily built for irrigation and flood control in the 1950s. We evaluate the impacts of rainfall change on the hedging and releasing rules over the historical period (1969–2008) by instrumental data and future period (2061–2100) based on simulations in a regional rivalry-mitigated scenario from the Coupled Model Intercomparison Project Phase 6. The main conclusions are as follows: (1) the annual total rainfall has a remarkable increasing trend from 2015 to 2100 and the annual precipitation variability exceeds the envelope range during the past 50-year period. The increased precipitation amount mainly occurs in spring (March to May). (2) The optimal regulation cycle is from September to August and from July to June for both historical and future periods. The limiting level during the nonflooded season is lower than the operating water level for more than five months in the historical period, which limits the ability of reservoir regulation and utilization of water resources. However, the water supply is no longer affected by flood control in 2061–2100 because of the redistribution of annual precipitation. (3) The projected irrigation and residential water demands of the Meishan reservoir are stable; thus, the improvement of the total economic benefit will mainly depend on power generation. This case provides a practical guide for many reservoirs serving water supply for small cities in eastern China, where the size of the population and cultivated land area is stagnant and the climate is getting wetter.

1. Introduction

Reservoir and dam system stores additional water in the rainy season and releases it in drier months, playing a fundamental role in achieving sustainable uses of the water resource. In practice, the storage or empty space to be maintained in a reservoir during different times of the year is determined by the rule curve, which is predefined in the design stage based on long-term (30 years or more) water inflow and demand statistics. Generally, the rule curve is fixed during the lifespan (~ a hundred years) of reservoirs [1, 2]. Almost 50% of hydropower reservoirs worldwide were originally commissioned more than 40 years ago [3], meaning that their rule curves are calculated from climate data in the mid-twentieth century or even earlier times. However, many observed changes in the climate system during the past decades are unprecedented over centuries, including precipitation and temperature-induced evaporation [4], which have already affected the regional runoff [5]. For the rest of the twenty-first century, the precipitation is projected to increase over high latitudes in parts of the monsoon regions and in wet regions over the tropics, but decrease in dry regions including large parts of the subtropics under 1.5°C and 2°C global warming scenarios. The changing patterns of precipitation minus evaporation and runoff are similar to precipitation [4]. Consequently, the

predefined reservoir rule curve in the mid-twentieth century may not remain as effective operation guidance in current and future situations. Moreover, aged reservoirs are about to reach the designed service life [3]. Maintaining or demolishing these reservoirs in the future requires an overall assessment of the energy-related benefits and ecological environment impacts, in which the most important issue is whether the reservoirs could still satisfy their original purpose in the context of hydroclimate change under global warming.

Studies on evaluating the impacts of climate change on reservoir management are carried out all over the world, for example, in northwest America [6], southeast Canada [7], eastern India [8], and northeast Thailand [9, 10]. Among countries owing dam-reservoir systems, China counts more than 40% of the total in the world [11]. Thus, a large number of cases are conducted in China, located in the Yangtze River basin [12–15], Xin'an River basin [16], Hai River basin [17], Yuan River basin [18], and Zhu River basin [19]. Some studies [12–14, 16–19] quantify the rainfall or streamflow change in the reservoir watershed and project the water budget in the future based on multiple theoretical models, but few of them provide the updated rule curves that could be directly used to guide the gate opening schedule of a dam. Others [6–10, 15] develop advanced optimization techniques to calculate rule curves but their future projections are extracted from a single climate model or based on previous-generation scenarios applied in the Coupled Model Intercomparison Project (CMIP) Phases 3 and 5. Now, the outputs of CMIP6 use a set of new plausible scenarios [4] and the multimodel ensemble is proved to be more reliable than a single model [20], which provides us more possible future pathways to assess the alteration of the rule curve under the climate change.

Over 80% of reservoirs in China were constructed between the 1950s and 1970s, most of which are not operated reliably and effectively now [21]. So far, however, the research in China mainly focuses on the electricity generation change of large or recently constructed hydropower reservoirs [12–14, 16, 18] and mega reservoirs serving the urban agglomerations [15, 17, 19]. There has been a little quantitative analysis of the rule curve change for the aged reservoirs. Here, we pay particular attention to the Meishan dam and reservoir system built during 1956–1958. It is located in Anhui Province (east-central China), where the province mean temperature and precipitation are projected to increase 2.3°C (with reference to 1850–1900) and 4.0% (with reference to 1985–2005) under 2°C global warming scenario [22]. Moreover, annual runoff changes in Anhui are basically consistent with precipitation, but the increased rates are different among seasons, that is, higher in spring and winter than in summer and autumn [23]. Thus, there is an urgent need to update the original regulating rules of Meishan reservoir developed 65 years ago. In this study, we use instrumental data for the past decades and the latest CMIP6 scenario simulations till 2100 to evaluate the potential impacts of climate change on the hedging and releasing strategies. A description of the study area, climate datasets, and statistical methods is given in Section 2. Regulating rule curves for historical and future periods are presented in Section 3. A summary of the main conclusions and the practical implications of the study case is discussed in Section 4.

2. Materials and Methods

2.1. Study Area. Meishan reservoir and dam project spans the Shi River in Anhui Province, eastern China (Figure 1(a)). Shi River is born in the Ta-pieh Mountains and joins the Huai River from the south side between Guan River and the west branch of Pi River (Figure 1(b)). The length of Shi River is 216 km and the catchment area is 6880 km^2. There are 11 branches upstream of the dam site, which is 130 km away from the confluence of the rivers Shi and Huai. The upstream watershed controlled by the dam is calculated from the hole-filled Shuttle Radar Topography Mission (SRTM) 90 m digital elevation data [24] by using a Matlab TopoToolbox [25]. The watershed area is 1970 km^2, which is mainly covered by forest and cropland (Figure 1(c)). The subtropical humid monsoon climate dominates the catchment area, where the annual mean temperature, precipitation, evaporation, and aridity are 15.5°C, 1381.5 mm, 1173.6 mm, and 0.8, respectively [26]. Meishan reservoir is a multipurpose facility and its main characteristics [27–29] are summarized in Table 1. The primary purposes are flood control and irrigation. The flood level is designed to accommodate an estimated 100-year recurrence interval event and it will provide the irrigation water needed for the agricultural sector to 25.5×10^2 km^2 farmland in five downstream counties (Figure 1(b)), that is, Jinzhai, Huoqiu, Liuan, Gushi, and Shangcheng [30]. Moreover, it releases the ecological flow to maintain the stability of the downstream ecosystem (e.g., biological protection, sediment deposition, and pollution control). However, hydropower generation is not mandatory and the turbines only work during irrigation season. The original capacity of the Meishan hydropower station was 40 MW, and it was increased to 52 MW after a retrofit project in 2012 [28].

2.2. Hydroclimate Observation and the Model for Predicting Inflow. The monthly inflow data at the dam site from 1969 to 2008 (Figure 2(a)) are provided by the Meishan hydrologic station, which could be used to directly calculate the water supply in the historical period. However, the water inflow is not an output variable of the CMIP6 models and we need to derive water inflow from other variables to calculate the water supply in the future period. Prior studies demonstrated that precipitation is the dominant factor of runoff in the upper-middle reaches of Huai River and its branches, while evaporation makes a minor contribution [31, 32]. Therefore, we test their relationships with water inflow. The rainfall data (1961–2014) are from a grid precipitation product [33] with a spatial resolution of 0.5°, and we use the data in the grid centered on 115.75°E, 31.75°N (Figure 2(b)) since it overlaps most of the reservoir area. The evaporation data are from the dataset of daily climate data from Chinese surface stations [34], and records of the nearest station

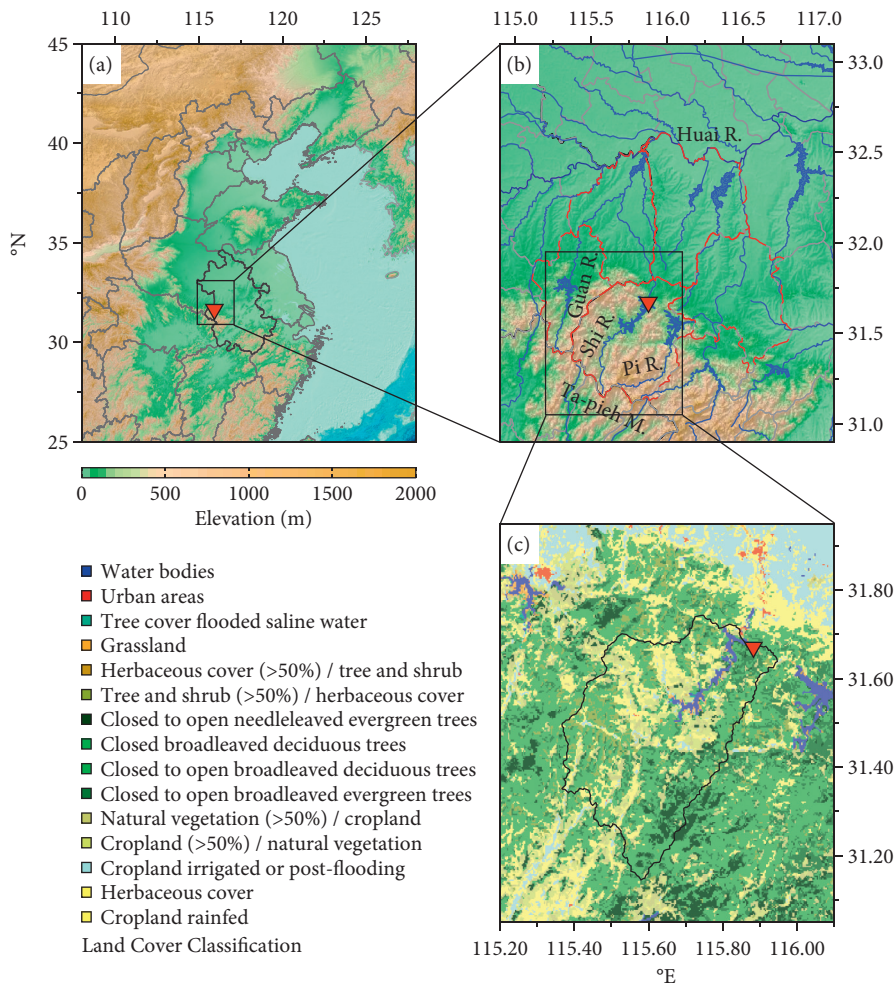


Figure 1: Basic geographic information of Meishan reservoir. (a) The location of Anhui Province (bold black line) in East Asia and the topography. (b) The location of the dam in Shi River and five counties (red dash line) irrigated by the reservoir. (c) The upstream catchment area of the dam and the land cover types in it. The red triangle indicates the dam position.

Table 1: Characteristics of the Meishan reservoir project [27–29].

Name	Value
Type of dam	Concrete multiple-arch dam
Controlled catchment area by the dam	1970 km^2
Total storage volume	22.63×10^8 m^3
Flood regulation capacity	10.65×10^8 m^3
Regulating storage	9.57×10^8 m^3
Dead storage	4.02×10^8 m^3
Annual mean inflow	14.05×10^8 m^3
Elevation of dam crest	140.17 m a.s.l.
Minimum operating elevation	107.07 m a.s.l.
Normal pool level	128.0 m a.s.l.
Dead water level	107.07 m a.s.l.
Design flood level	137.69 m a.s.l.
Check flood level	139.93 m a.s.l.
Limiting level during flood season	125.27 m a.s.l.
Limiting level during the nonflood season	126.0 m a.s.l.
Maximum dam height	88.24 m
Width crest	1.8 m
Top length of the dam	443.5 m

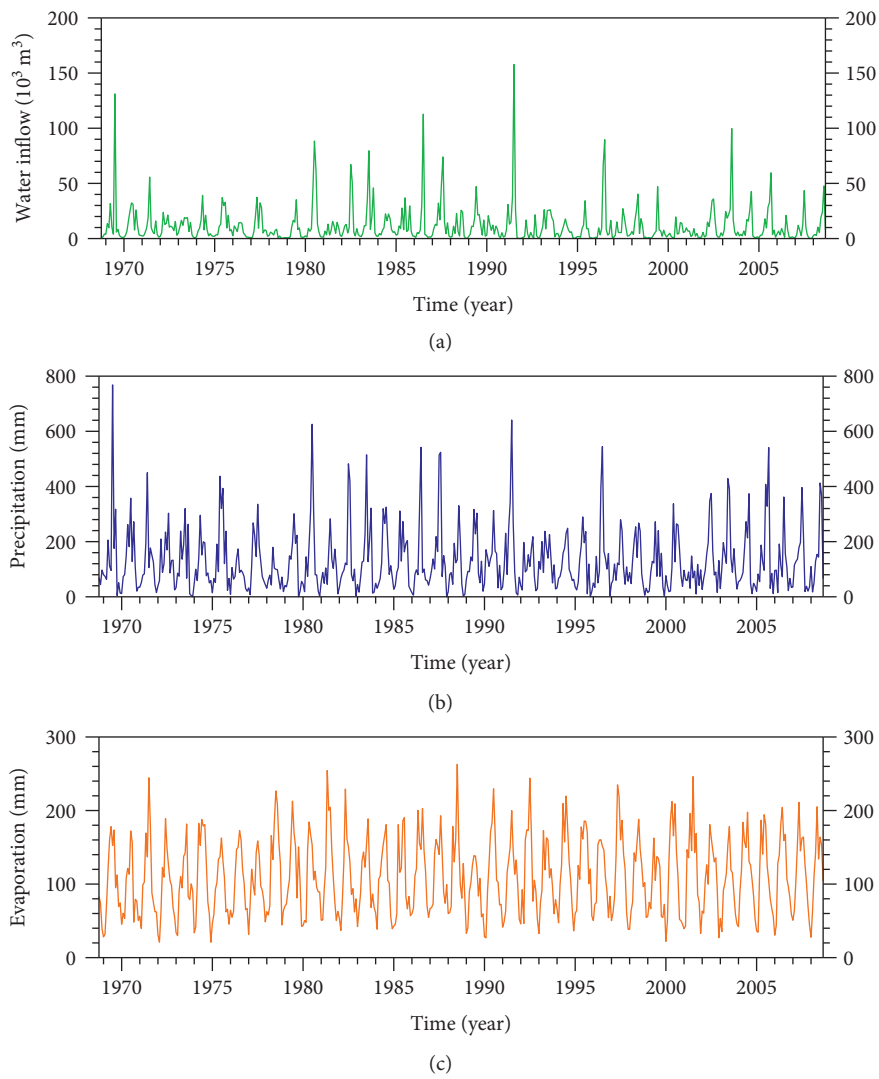


Figure 2: Monthly (a) water inflow, (b) precipitation, and (c) evaporation at Meishan reservoir.

(Jinzhai, 115.88°E, 31.68°N) are used to represent evaporation in the drainage basin (Figure 2(c)). Several regression models are calculated using different combinations of predictors and validated by the leave-one-out procedure (Table 2), and the result shows that the 2nd degree polynomial fitting by precipitation has the highest predicted variance of 85.99%. This nonlinear relationship is also observed in other reaches of the Huai River as that rains could induce more surface runoff after soil saturation [35]. However, introducing the evaporation parameter would reduce the model skill, such that only precipitation is used to predict future inflow change.

2.3. CMIP6 Data and Ensemble Method. Projected precipitation change is needed to generate the reservoir rule curve for the future. Here, we firstly describe the future scenario selection and then introduce the method to calculate weighted mean precipitation from multiple model outputs. CMIP aims to better understand climate changes in a multimodel context. The model performance is assessed by a set of experiments with idealized and historical radiative forcings. Future climate change simulations are based on designed scenarios [36]. CMIP has developed in phases with expanded number and complexity of climate models. In CMIP5, four Representative Concentration Pathway (RCP) scenarios (RCP2.6, RCP4.5, RCP6, and RCP8.5) represent different assumptions about atmospheric greenhouse gas concentrations based on population, economic growth, energy, and land use over the twentieth century has been developed [37]. They are labeled after a possible radiative forcing anomaly value (relative to the preindustrial level) in the year 2100. The latest CMIP6 has a new set of paths so-called SSP (shared socioeconomic pathways) scenarios, indicating different future socio-economic projections and political environments [38]. There are five high-priority SSP scenarios defined in terms of challenges to adaptation and mitigation, from an increasingly sustainable world (SSP1) to a high fossil fuel dependent world (SSP5). Combining the five SSP pathways with the different radiative forcings yields

Table 2: Summary of the regression models to predict water inflow by precipitation (P) and evaporation (E).

Model no.	1	2	3	4
Predictors	P	P, P^2	P, E	P, P^2, E
Equation	105.60 P	$44.52\,P + 0.1829\,P^2$	131.37 P − 46.00 E	$43.11\,P + 1.35\,E + 0.1849\,P^2$
R-square (adjusted)	79.30%	86.66%	81.94%	86.63%
R-square (predicted)	78.75%	85.99%	81.35%	85.90%

a scenario matrix and four of them are selected as standard scenarios (SSP585, SSP370, SSP245, and SSP126). It is worth noting that SSP585, SSP245, and SSP126 can be understood as an update or remake of the RCP8.5, RCP4.5, and RCP2.6, while SSP370 is a newly introduced scenario closing the gap between RCP6.0 and RCP8.5. Specifically, SSP370 represents a regional rivalry world that degrades the international priority for addressing environmental concerns and the corresponding forcing by 2100 is 3.7 W/m^2 [38]. On a global scale, the mean Earth surface temperature will increase 3.6°C over 2081–2100 in SSP370 compared to 1850–1900 and this scenario represents the upper limit of plausible future [4]. Therefore, hydroclimate change under the SSP370 pathway in the Meishan reservoir and potential operation change are investigated in this study. A total of 34 models containing precipitation simulation for SSP370 are used and their information is summarized in Table 3 [39]. All CMIP6 model data are downloaded from the Climate Explorer developed by the Royal Netherlands Meteorological Institute [40], including the SSP370 experiment from 2015 to 2100 and the historical run (forced with changing conditions that are consistent with observations) from 1850 to 2014, the latter is used to evaluate the model performance in our study area by comparing with instrumental data.

It is fundamentally impossible to describe all true climate processes in a single model and the uncertainties introduced by choices in the model design cannot be addressed by varying its parametrizations. Multimodel ensemble strategy is a way to exploit the advantage and to avoid the weakness of individual models due to substantial reduction of the systematic errors in individual models, and multimodel means often agree better with observations than individual model results [20]. In this study, we compute the multimodel mean by both equally and unequally weighted methods and select the best one to predict future precipitation. Note that some of the models have multiple ensemble members with different initial conditions. Therefore, there are three equally weighted multimodel ensemble methods: (1) calculating the mean of all members; (2) calculating the mean of all models with only one member per model; (3) first calculating the means of individual models by corresponding members and then calculating the mean of all models. Moreover, the unequal weights are determined through the best subset regression constrained by maximizing adjusted R-square. All model outputs are resized to 192 × 144 horizontal resolution to calculate the ensemble mean.

2.4. Reservoir Rule Curve Calculation. The reservoir rule curve guides the releasing schedule to impound excess water in wet periods and ensure water demand in dry periods because the water demand change is not consistent with the inflow change. Thus, except for water inflow data described in Section 2.2, water demand data are also required to calculate the rule curve. Meishan reservoir project is the main water source of the Shi River irrigation area. The irrigation release flow is 190 m^3/s and the average agricultural water consumption is 10.63×10^8 m^3 per year under the designed assurance level of 80%. In addition, the annual urban water demand is 400×10^4 m^3 and the downstream ecosystem flow is 5.47 m^3/s [28]. The total annual water consumption of the aforementioned water-using sectors is 12.36×10^8 m^3. According to the long-term statistics, the month-by-month water demands of the reservoir are shown in Figure 3(a).

To calculate the rule curve, the first step is identifying the regulation cycle. A complete operation cycle ranges from one day to multiple years, determined by the coefficient of reservoir storage capacity β (proportion of regulating storage to multiyear average inflow). The range of β for daily, seasonally, yearly, and multiyear (or carryover) reservoirs is $\beta \leq 2\%$, $2\% \beta \leq 8\%$, $8\% \beta \leq 20\%$, and $\beta > 20\%$, respectively [41]. In this study, β is equal to 68% (9.57×10^8 m^3/ 14.05×10^8 m^3, Table 1) such that the reservoir needs to be carryover regulated. Generally, the horizontal axis of the operating rule curve represents time (month or 10-day period unit), and the vertical axis indicates the reservoir water level or storage capacity. There are three time-varying curves between normal pool level and dead water level. The first is the flood limited curve, which constrains the reservoir water level in the flood season. The other two are the lower water supply curve (or upper rule curve) and limited water supply curve (or lower rule curve), guiding the water refill and release strategies. We select 3–6 representative hydrological years, in which the water inflow is close to the water demand, to calculate operating rule curves, as well as associated annual and carryover storage, and the step-by-step computation follows the engineering standard guidance from reference [42]. The elevation-volume relationship [43] is plotted in Figure 3(b).

3. Results

Generally, the regulating year is different from the calendar year, and the former usually starts from the beginning of the flood season and lasts for 12 months. Under the predefined operation schedule of Meishan reservoir, the regulating year is from June to May [43]. However, the reservoir was commissioned more than 60 years ago and the catchment area experienced significant climate change since then. In this section, we first explore how the optimal regulating rules change in the past decades (1969–2008, historical period). Then, we present the future

Table 3: List of CMIP6 models used in this study.

#	Model name	Institution (country)	Resolution (lon × lat)
1	ACCESS-CM2	Commonwealth Scientific and Industrial Research Organisation/ Australian Research Council Centre of Excellence for Climate System Science (Australia)	192 × 144
2	ACCESS-ESM1-5	Commonwealth Scientific and Industrial Research Organisation (Australia)	192 × 145
3	AWI-CM-1-1-MR	Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research (Germany)	384 × 192
4	BCC-CSM2-MR	Beijing Climate Centre (China)	320 × 160
5	CAMS-CSM1-0	Chinese Academy of Meteorological Sciences (China)	320 × 160
6	CanESM5-CanOE-p2 CanESM5-p2 CanESM5	Canadian Centre for Climate Modelling and Analysis, Environment and Climate Change Canada (Canada)	128 × 64
7	CESM2-WACCM CESM2	National Centre for Atmospheric Research, Climate and Global Dynamics Laboratory (USA)	288 × 192
8	CMCC-CM2-SR5	Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici (Italy)	288 × 192
9	CNRM-CM6-1-HR-f2	Centre National de Recherches Meteorologiques/Centre Europeen de Recherche et de Formation Avancee en Calcul Scientifique (France)	720 × 360
10	CNRM-CM6-1-f2		
11	CNRM-ESM2-1-f2		256 × 128
12	EC-Earth3-veg	EC-Earth-Consortium[1]	512 × 256
13	EC-Earth3		
14	FGOALS-f3-L	Chinese Academy of Sciences (China)	360 × 180
15	FGOALS-g3		180 × 80
16	GFDL-ESM4	National Oceanic and Atmospheric Administration, Geophysical Fluid Dynamics Laboratory (USA)	360 × 180
17	GISS-E2-1-G-f2	Goddard Institute for Space Studies (USA)	144 × 90
18	GISS-E2-1-G-p3		
19	INM-CM4-8	Institute for Numerical Mathematics, Russian Academy of Science (Russia)	180 × 120
20	INM-CM5-0		
21	IPSL-CM6A-LR	Institut Pierre Simon Laplace (France)	144 × 143
22	KACE-1-0-G	National Institute of Meteorological Sciences/Korea Meteorological Administration, Climate Research Division (Korea)	192 × 144
23	MIROC-ES2L-f2	Japan Agency for Marine-Earth Science and Technology/ Atmosphere and Ocean Research Institute, the University of Tokyo/National Institute for Environmental Studies/RIKEN Centre for Computational Science (Japan)	128 × 64
24	MIROC6		256 × 128
25	MPI-ESM1-2-HR	Max Planck Institute for Meteorology/Deutscher Wetterdienst/ Deutsches Klimarechenzentrum (Germany)	384 × 192
26	MPI-ESM1-2-LR	Max Planck Institute for Meteorology/Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research/Deutscher Wetterdienst/Deutsches Klimarechenzentrum (Germany)	192 × 96
27	MRI-ESM2-0-i2	Meteorological Research Institute (Japan)	320 × 160
28	MRI-ESM2-0		
29	NorESM2-LM	NorESM Climate Modeling Consortium[2]	144 × 96
30	NorESM2-MM		288 × 192
31	UKESM1-0-LL-f2	Met Office Hadley Centre/Natural Environment Research Council (UK)/National Institute of Meteorological Sciences/Korea Meteorological Administration, Climate Research Division (Korea)/National Institute of Water and Atmospheric Research (New Zealand)	192 × 144

[1,2] The full consortium list is available at https://github.com/WCRP-CMIP/CMIP6_CVs/blob/master/CMIP6_institution_id.json.

precipitation projection over 2061–2100 in the SSP370 scenario and associated rule curve change.

3.1. Reservoir Rule Curve of the Historical Time. The annual cycle of water inflow and demand is shown in Figure 3(a), and the flood season is from May to September. The irrigation water demand peak (in June) is a little earlier than the water inflow peak (in July) during 1969–2008; consequently, the predefined operation schedule (June to May) frequently suffers from water deficit in the early months. Moreover, the water demand cycle is largely overlapped with the water inflow cycle such that it is difficult to identify the appropriate regulating cycle. Therefore, we plot all possible

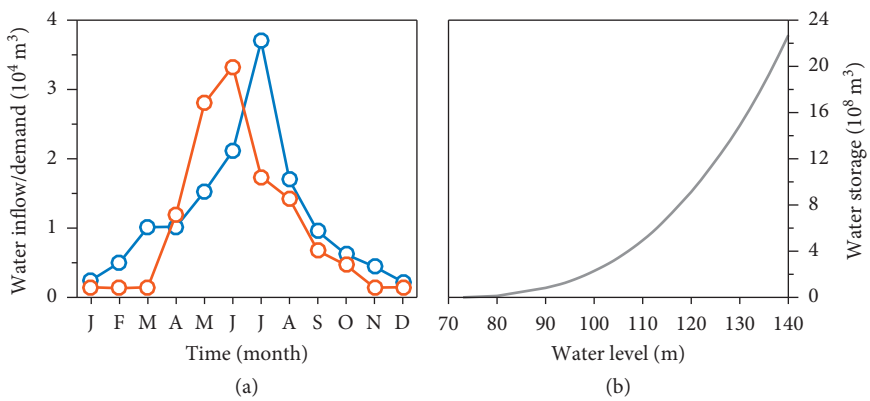


Figure 3: (a) Monthly water inflow (blue) and demand (orange) and (b) the elevation-volume curve.

cases together to make a comparison (Figure 4). For each case, six representative hydrological years (Table 4) are selected to calculate rule curves. Although the total inflow water amount is adjusted to match the water demand, there exist years that the water supply is damaged (negative storage anomaly) due to the uneven seasonal distribution of water inflow. In these cases, the carryover storage needs to be used although it is not a dry year. Therefore, the percentage of damage years is used to screen the appropriate regulating year cycle. On this basis, the best choice is from September to August, which only has one damage year (16.7%). The second optional choice is from July to June (two months later than the beginning of the flood season), which has three damage years (50%). The remaining annual cycles have at least four damage years (>66.7%) while all years are damaged from April to March and May to April. Figure 5 shows the reservoir rule curves starting from July and September, respectively. However, the limiting level during the non-flooded season is 126.0 m, which is lower than the operating water level for seven and five months in the above two cases. Such issues, to some extent, limit the ability of reservoir regulation and utilization of water resources.

3.2. Local Precipitation Change in SSP370 Scenario. Firstly, we evaluate the ability of a set of CMIP6 models to capture the observed changes by Taylor diagram [44] (Figure 6). For a good simulation, the correlation coefficient with the observation and the ratio of their variance should be close to one, while the centered root mean square error (RMSE) should be close to zero. In this study, we test the performance of multiple models in only one grid; thus, the statistics are calculated based on the historical time series instead of the average spatial pattern used in other studies. The statistics for the majority of models are clustered in the intervals in which the correlation coefficients are between 0.35 and 0.55, the ratios of variance are between 45% and 90%, and the centered RMSEs are between 0.8 and 1. The result indicates that most models have a reasonable performance in simulating temporal change of precipitation in the study site, while none of them shows outstanding skills. The three equally weighted multimodel ensemble means also give results that are similar to the best individual model. Moreover, we build the best subset regression ensemble by seven CMIP6 models whose correlation coefficients are larger than 0.5 and it finally maintains five input variables. The regression equation is as follows:

$$Y = 0.13X_2 + 0.17X_{10} + 0.21X_{15} + 0.31X_{22} + 0.13X_{23} + 0.13X_{24} + 0.15X_{26} - 13.64, \quad (1)$$

where Y is the ensemble predicted precipitation, X is the model simulated precipitation, and the subscript number represents the model in Table 3. This unequally weighted multimodel ensemble has the adjusted R-square of 0.43 (equal to the correlation coefficient of 0.66), which has 11% more explained variance than the best individual model (IPSL-CM6A-LR, correlation coefficient = 0.57, RMSE = 0.82) and three equally weighted ensemble methods. Based on the best subset regression, the future precipitation is projected to 2100 (Figure 7). Like the temperature, the annual total rainfall also has a remarkable increasing trend. The precipitation amount of the wettest year reaches 2760 mm in the 2090s, while it is only 831 mm in the driest year in the mid-twenty-first century, indicating that the annual precipitation variability exceeds the envelope range during the past 50-year period (Figure 7(a)). The mean annual precipitation is 2022 mm over 2061–2100, which is 500 mm more than that of 1961–2010. The increased precipitation amount mainly occurs in spring (March to May), while July to September experiences a reduction (Figure 7(b)).

3.3. Reservoir Rule Curve of the 2061–2100 Period. Since the monthly distribution of rainfall changes in the future, the reservoir regulating strategy also needs to be adjusted. We assume that the water demand is the same as the current time and plot all possible regulating cycles for 2061–2100 in Figure 8. It is worth noting that the standard deviation of the weighted ensemble mean is reduced by a factor of $\sqrt{1/N}$ since the ensemble members are sequentially independent and possess an identical standard deviation. Therefore, the range of representative hydrological years of the multimodel ensemble is also reduced. To eliminate this effect, we pick out three representative hydrological years for seven selected

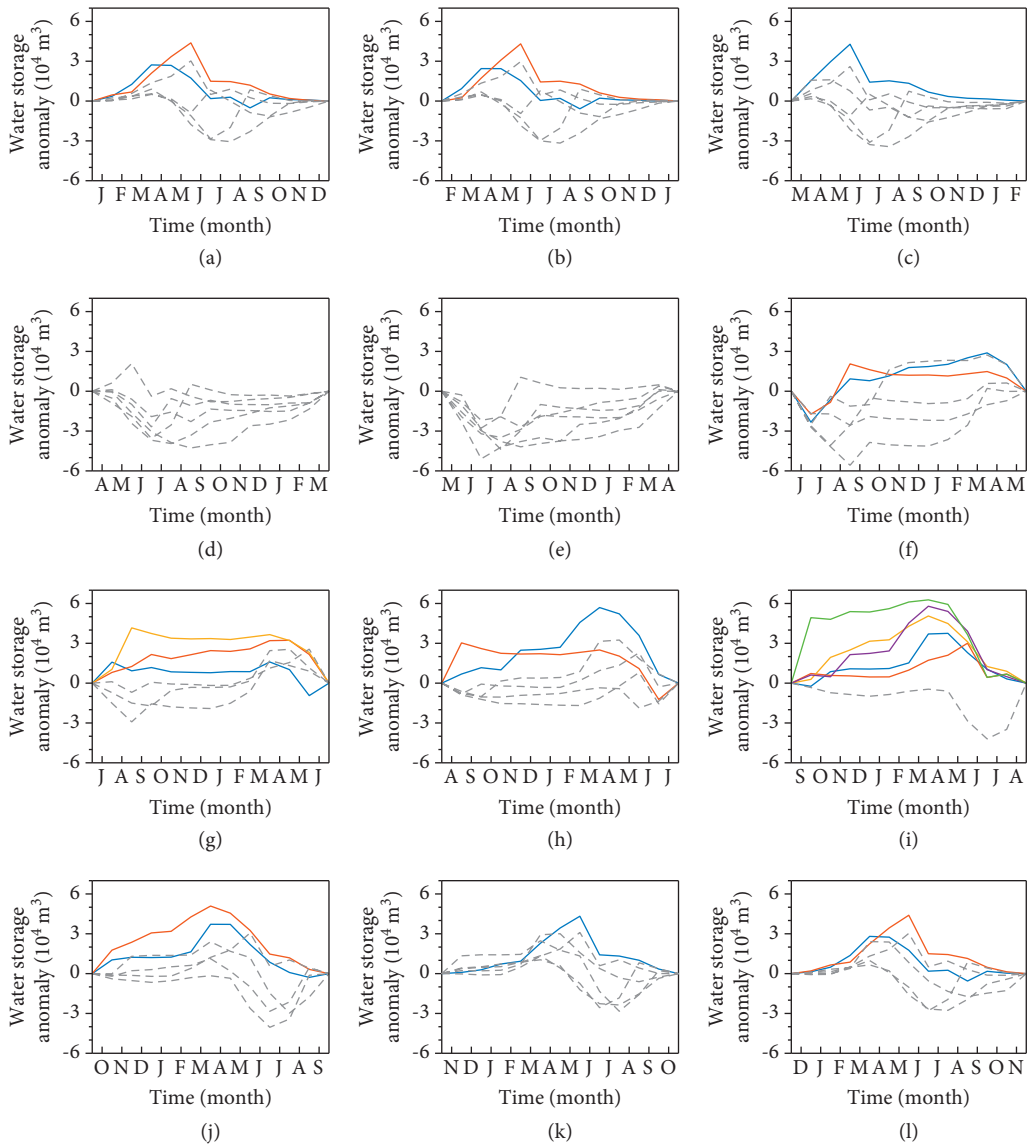


Figure 4: Water storage anomaly for different 12-month regulating cycles in historical time (1969–2008). The curves for damage years (negative storage anomaly for more than two months) are shown by gray dash lines while undamaged years by color solid lines (colors are only used to distinguish multiple lines). (a–l) Regulating cycles start from January to December.

Table 4: Six representative hydrological years under different regulation cycles.

Regulation cycle	Jan-Dec	Feb-Jan	Mar-Feb	Apr-Mar	May-Apr	Jun-May	Jul-Jun	Aug-Jul	Sep-Aug	Oct-Sep	Nov-Oct
Representative	1971	1971	1971	1972	1972	1973	1971	1971	1971*	1971*	1971
	1973*	1973*	1973	1973	1973	1977*	1973	1973	1973*	1973	1973
	1974	1974	1974	1974	1984	1985	1979*	1974	1975*	1975*	1983
hydrological years	1984	1984	1984	1979	1988	1988	1992	1976	1989*	1987	1987
	1998*	1998*	1998*	1984	1997	1992	1993*	1989*	2005*	2003	1997*
	2004	2004	2004	2004	2004	2004*	2004*	2004*	2007	2007	2003

*Undamaged water supply years.

models (Table 5) to generate the rule curves (Figure 8). Similar to the historical period, the number of damage years is least (48%) from September to August. The second optimal cycle is from July to June and November to October, which both have 70% damage years. The cycles with 100% damage years are increased from two to four. The selected rule curves are provided in Figure 9. Compared with the historical period, the annual storage is increased for the July-June scheme while decreased for the September-August scheme. Due to the redistribution of annual precipitation,

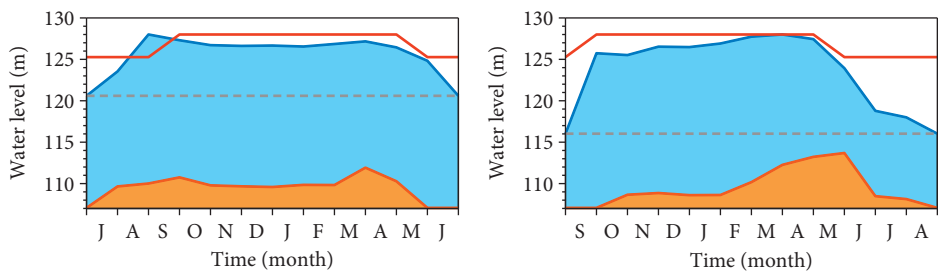


Figure 5: Reservoir rule curves for the July-June scheme (a) and September-August scheme in historical time (1969–2008). The red line is the flood limited curve. The blue/orange line is the lower/limited water supply curve. The gray dash line indicates the threshold of annual and carryover storage.

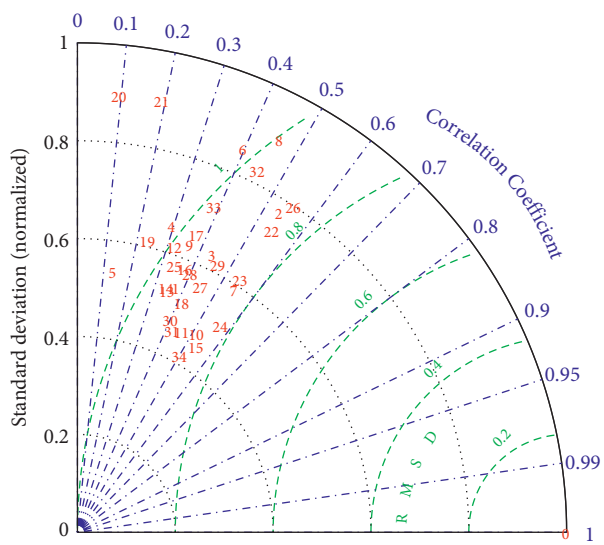


Figure 6: Taylor diagram for precipitation of 34 CMIP6 models. Historical model simulations and observations from 1961 to 2014 are used to calculate the statistics. The number 0 represents observation and models are indicated from 1 to 34 as that in Table 3.

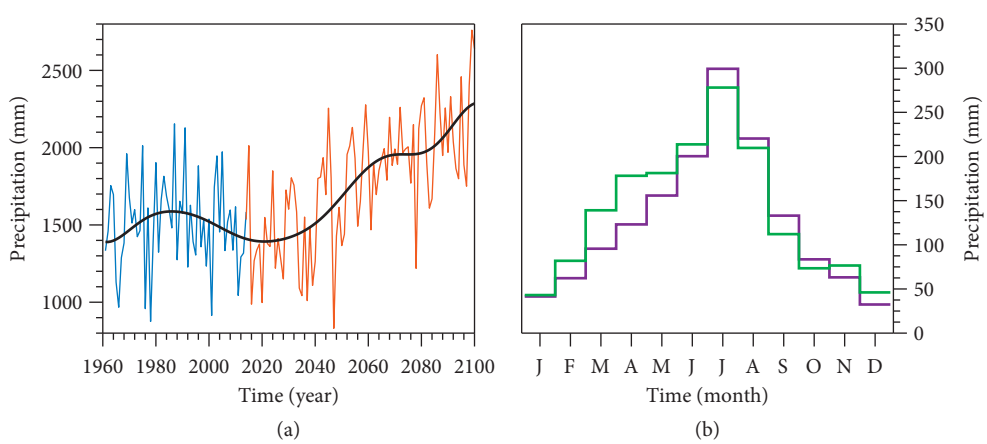


Figure 7: Precipitation change in the Meishan reservoir area in SSP370 scenario. (a) Multimodel ensemble precipitation projection from 2015 to 2100 (orange) and the comparison with the historical observation from 1961 to 2014 (blue). The black line is the 30-year low-pass curve by the Lanczos filter. (b) The monthly precipitation climatology from 1961 to 2014 historical observation (purple) and 2061–2100 multimodel ensemble (green).

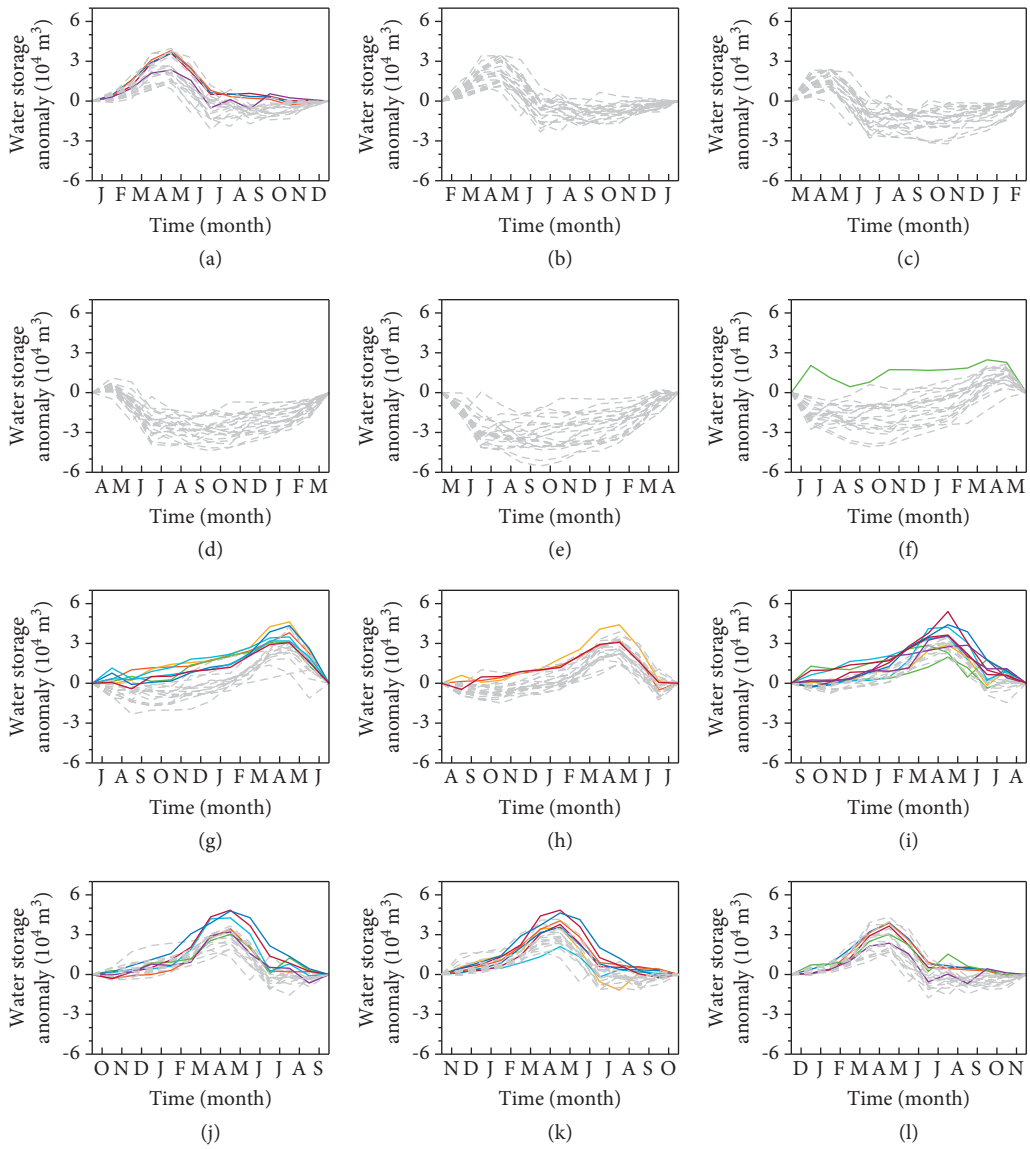


Figure 8: Water storage anomaly for different 12-month regulating cycles in the SSP370 scenario from 2061 to 2100. The curves for damage years (negative storage anomaly for more than two months) are shown by gray dash lines while undamaged years by color solid lines (colors are only used to distinguish multiple lines). (a–l) Regulating cycles start from January to December.

Table 5: Three representative hydrological years under different regulation cycles for five models.

Regulation cycle	Jan-Dec	Feb-Jan	Mar-Feb	Apr-Mar	May-Apr	Jun-May	Jul-Jun	Aug-Jul	Sep-Aug	Oct-Sep	Nov-Oct
ACCESS-ESM1-5	2063	2063	2063	2087	2083	2083	2082	2071*	2062	2062	2062
	2080	2080	2080	2089	2089	2087	2083*	2082	2079	2079	2079*
	2083	2083	2089	2094	2094	2094	2094*	2092*	2082	2082	2082
CESM2	2073	2073	2073	2073	2073	2086	2086*	2072	2072*	2072	2072*
	2086	2086	2086	2086	2090	2089	2090*	2089	2085	2078	2078
	2090	2090	2090	2090	2093	2090	2093	2090	2090*	2089	2089
EC-Earth3-veg	2078*	2066	2066	2066	2077	2084	2066	2077	2065	2065	2077*
	2088*	2078	2078	2077	2094	2087	2087	2087	2077	2077*	2079*
	2096*	2088	2088	2096	2096	2094	2096	2096	2087	2087	2087*
INM-CM4-8	2087	2087	2066	2066	2087	2066	2066	2067	2067*	2067*	2066
	2089	2089	2087	2087	2089	2076	2078	2086	2078*	2086	2086
	2092	2092	2092	2089	2091	2087*	2086	2091	2086*	2091	2091

Table 5: Continued.

Regulation cycle	Jan-Dec	Feb-Jan	Mar-Feb	Apr-Mar	May-Apr	Jun-May	Jul-Jun	Aug-Jul	Sep-Aug	Oct-Sep	Nov-Oct
INM-CM5-0	2068	2068	2068	2068	2068	2068	2068*	2067	2067*	2067*	2067
	2096	2096	2096	2096	2080	2087	2087	2095	2080*	2095*	2095*
	2097	2097	2097	2097	2096	2096	2096	2096	2096*	2096*	2096*
IPSL-CM6A-LR	2069	2071	2069	2066	2069	2069	2068	2074	2068	2068	2074
	2075*	2075	2075	2069	2074	2074	2074	2087	2074	2074	2084
	2096	2096	2096	2074	2085	2083	2095	2095	2095*	2095*	2095
MIROC-ES2L-f2	2077	2079	2079	2079	2079	2079	2076	2094*	2094*	2091	2076
	2098	2095	2092	2094	2092	2094	2087*	2097	2097	2097	2097
	2100	2098	2098	2098	2094	2095	2094*	2099	2099	2099	2099

*Undamaged water supply years.

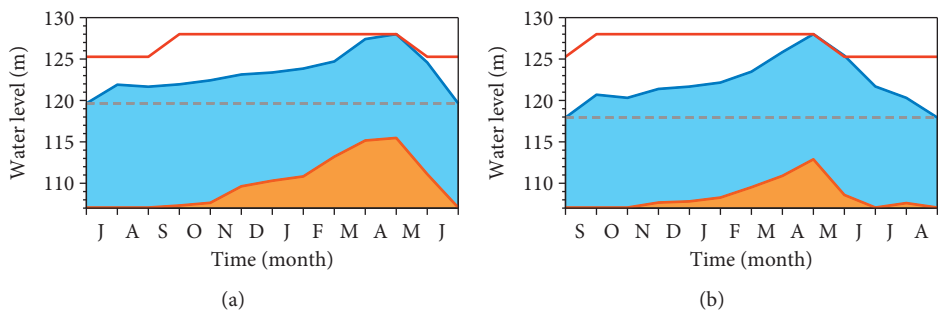


Figure 9: Reservoir rule curves for the July-June scheme (a) and September-August scheme in SSP370 scenario from 2061 to 2100. The red line is the flood limited curve. The blue/orange line is the lower/limited water supply curve. The gray dash line indicates the threshold of annual and carryover storage.

the operating water level is reduced during the nonflooded season and it is lower than the flood control level, indicating that the water supply is no longer affected by flood control in the future. As a result, the efficiency of water utilization is enhanced.

4. Discussion and Conclusions

In this study, we evaluate the impacts of hydroclimate change on the regulating rule curves of the Meishan reservoir constructed in the 1950s, for the historical period (1969–2008) by instrumental data and future period (2061–2100) by CMIP6 simulations. We find that the optimal regulation cycle is from September to August and from July to June for both historical and future periods. The limiting level during the nonflooded season is lower than the operating water level for more than five months in the historical period, which limits the ability of reservoir regulation and utilization of water resources. However, the water supply is no longer affected by flood control in 2061–2100 because of the redistribution of annual precipitation.

Therefore, the predefined operating strategy from June to May is not an efficient scheme for either historical or future scenarios. Operation log of the Meishan reservoir records that the water level is lower than the dead water level in nearly 25% of months from 1965 to 2015, and sometimes even lower than the elevation of irrigation water intake of 95.27 m [27]. Because the operation dead water level is lower than design, the utilization rate of water resources is reduced and the conflicts among water demand sectors are inevitable within the hydrological year. In addition, the associated facilities, such as dam, dike, gate, and plant, will deteriorate more quickly under an abnormal operating condition.

Many large reservoirs in China are designed to provide residential water to metropolises. The population in these cities will continue to grow under the urbanization process in future China [45]. Therefore, the water supply requirement of the large reservoirs is expected to increase year by year. On the contrary, the Meishan reservoir affords water supply for some small counties, which experience a net outflow of population. For example, the population statistics for Jinzhai County, which is the first residential area downstream of the Meishan reservoir, has dropped by 3.49% while the proportion of people aged 65 or above has increased by 6.33% from 2010 to 2020 [46]. Therefore, the demand for urban water is not expected to increase. Furthermore, the irrigation water demand in the future is likely to stay the same as the current level because the cultivated land area in Anhui Province is stable [47]. Without additional water pressure, the improvement of the total economic benefit of the Meishan reservoir will mainly depend on the increase in the benefit of power generation, which will also help China to reach the carbon peak and neutrality targets. Studies have shown that precipitation in many regions of China will increase under global warming scenarios [48]. The case of Meishan reservoir provides a possible reference for the planning and operation of medium-size

reservoirs in these regions. This study also presents a framework to calculate rule curves from multisource datasets (i.e., geographic raster, construction planning document, and climate data) by using a set of software packages. Rule curves for other similar reservoirs could be calculated following this framework, and the individual dataset or software could be substituted by state-of-the-art products.

Data Availability

The shaded relief map in Figures 1(a) and 1(b) is from GEBCO https://download.gebco.net/. The land cover data for 2015 in Figure 1(c) is from http://maps.elie.ucl.ac.be/CCI/viewer/download.php. The grid monthly precipitation data for China are from http://poles.tpdc.ac.cn/zh-hans/data/653191c3-df99-479b-95bd-a18bee30f2bd/. Requests for access to the inflow data of Meishan reservoir should be made to the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors acknowledge World Climate Research Programme's Working Group on Coupled Modelling for coordinating and promoting the CMIP6. The authors thank the climate modeling groups for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies for supporting CMIP6 and ESGF. This work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (grant no. XDA23100401) and the National Key Research and Development Program of China (grant no. 2017YFA0605303).

References

[1] F.-J. Chang, L. Chen, and L.-C. Chang, "Optimizing the reservoir operating rule curves by genetic algorithms," *Hydrological Processes*, vol. 19, no. 11, pp. 2277–2289, 2005.

[2] W. H. Wan, J. S. Zhao, and J. B. Wang, "Revisiting water supply rule curves with hedging theory for climate change adaptation," *Sustainability*, vol. 11, no. 7, 2019.

[3] E. Quaranta, G. Aggidis, and R. M. Boes, *Assessing the energy potential of modernizing the European hydropower fleet*, vol. 246, Energy Conversion and Management, Amsterderm, Netherlands, 2021.

[4] IPCC, *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change,* Cambridge University Press, Cambridge, England, 2021.

[5] S. Yan, J. Y. Liu, X. H. Gu, and D. D. Kong, "Global runoff signatures changes and their response to atmospheric environment, GRACE water storage, and dams," *Remote Sensing*, vol. 13, p. 20, 2021.

[6] N. Fayaz, L. E. Condon, and D. G. Chandler, "Evaluating the sensitivity of projected reservoir reliability to the choice of climate projection: a case study of bull run watershed, portland, Oregon," *Water Resources Management*, vol. 34, no. 6, pp. 1991–2009, 2020.

[7] H.-I. Eum, A. Vasan, and S. P. Simonovic, "Integrated reservoir management system for flood risk assessment under climate change," *Water Resources Management*, vol. 26, no. 13, pp. 3785–3802, 2012.

[8] D. Raje and P. P. Mujumdar, "Reservoir performance under uncertainty in hydrologic impacts of climate change," *Advances in Water Resources*, vol. 33, no. 3, pp. 312–326, 2010.

[9] A. Kangrang, H. Prasanchum, and R. Hormwichian, "Development of future rule curves for multipurpose reservoir operation using conditional genetic and tabu search algorithms," *Advances in Civil Engineering*, vol. 2018, 2018.

[10] T. Thongwan, A. Kangrang, and H. Prasanchum, "Multi-objective future rule curves using conditional tabu search algorithm and conditional genetic algorithm for reservoir operation," *Heliyon*, vol. 5, no. 9, Article ID e02401, 2019.

[11] M. Mulligan, A. van Soesbergen, and L. Sáenz, "GOODD, a global dataset of more than 38,000 georeferenced dams," *Scientific Data*, vol. 7, no. 1, p. 31, 2020.

[12] P. C. Qin, H. M. Xu, and M. Liu, "Assessing concurrent effects of climate change on hydropower supply, electricity demand, and greenhouse gas emissions in the Upper Yangtze River Basin of China," *Applied Energy*, vol. 279, 2020.

[13] W. J. Zhong, J. Guo, and L. Chen, "Future hydropower generation prediction of large-scale reservoirs in the upper Yangtze River basin under climate change," *Journal of Hydrology*, vol. 588, 2020.

[14] Y. Feng, J. J. Xu, and Y. Hong, "Reservoir scheduling using a multi-objective cuckoo search algorithm under climate change in jinsha river, China," *Water*, vol. 13, no. 13, 2021.

[15] Y. Zhou and S. Guo, "Incorporating ecological requirement into multipurpose reservoir operating rule curves for adaptation to climate change," *Journal of Hydrology*, vol. 498, pp. 153–164, 2013.

[16] S. Wu, J. Zhao, X. Lei, Z. Wang, and H. Wang, "Impacts of climate change on operation of Xin'an River reservoir and adaption strategies," *Journal of Hydroelectric Engineering*, vol. 36, pp. 50–58, 2017.

[17] T. Yan, J. Bai, T. Arsenio, J. Liu, and Z. Shen, "Future climate change impacts on streamflow and nitrogen exports based on CMIP5 projection in the Miyun Reservoir Basin, China," *Ecohydrology and Hydrobiology*, vol. 19, no. 2, pp. 266–278, 2019.

[18] Y. Guo, G. Fang, X. Wen, X. Lei, Y. Yuan, and X. Fu, "Hydrological responses and adaptive potential of cascaded reservoirs under climate change in Yuan River Basin," *Hydrology Research*, vol. 50, no. 1, pp. 358–378, 2019.

[19] C. Wu, G. Huang, H. Yu, Z. Chen, and J. Ma, "Impact of climate change on reservoir flood control in the upstream area of the beijiang River basin, south China," *Journal of Hydrometeorology*, vol. 15, no. 6, pp. 2203–2218, 2014.

[20] C. Tebaldi and R. Knutti, "The use of the multi-model ensemble in probabilistic climate projections," *Philosophical Transactions of the Royal Society A: Mathematical, Physical & Engineering Sciences*, vol. 365, no. 1857, pp. 2053–2075, 2007.

[21] E. Zhao, *Dam Safety Monitoring Data Analysis Theory and Assessment Methods*, Hohai University Press, Nanjing, China, 2018, (in Chinese).

[22] S. Wang, H. Xu, W. Yang, and X. Zhang, "Estimated projections in climate change and meteorological drought in Anhui Province under 1.5 and 2°C global warming based on RCP," *Journal of China Agricultural University*, vol. 23, no. 6, pp. 100–107, 2018, (in Chinese).

[23] Y. Yao, W. Qu, and J. X. Lu, "Responses of hydrological processes under different shared socioeconomic pathway scenarios in the huaihe River basin," *China, Water*, vol. 13, no. 8, 2021.

[24] H. I. Reuter, A. Nelson, and A. Jarvis, "An evaluation of void filling interpolation methods for SRTM data," *International Journal of Geographic Information Science*, vol. 21, no. 9, pp. 983–1008.

[25] W. Schwanghart and D. Scherler, "Short Communication: TopoToolbox 2 - MATLAB-based software for topographic analysis and modeling in Earth surface sciences," *Earth Surface Dynamics*, vol. 2, no. 1, pp. 1–7, 2014.

[26] Q. Xu, L. Ren, J. Liu, B. Yang, and X. Liu, "Forecast model of reservoir inflow based on DEM and its system development," *Chinese Journal of Geotechnical Engineering*, vol. 30, no. 11, pp. 1748–1751, 2008.

[27] L. Zhang, L. Wu, and J. Jin, "Study of staged drought-limited water level of the main reservoir in the large-scale irrigation district," *Journal of Hydraulic Engineering*, vol. 49, no. 6, pp. 757–766, 2018.

[28] Z. Liu, "Studies on regulation modes of meishan reservoir," *China Water Resources*, vol. 9, pp. 30-31, 2018.

[29] Z. Yin, "Discussion on water utilization and protection of meishan reservoir," *Jianghuai Water Resources Science and Technology*, vol. 3, pp. 5–25, 2012.

[30] C. Wang, "Analysis of comprehensive operation mode of Meishan Reservoir," *Journal of Anhui Technical College of Water Resources and Hydroelectric Power*, vol. 3, no. 1, pp. 15–17, 2003.

[31] Y. Xue, C. Gao, X. Zhang, Y. Xu, and P. Li, "Response of runoff in huaihe river upstream to change of different climate elements," *Journal of China Hydrology*, vol. 37, pp. 22–28, 2017.

[32] S. Wang, H. Xu, L. Liu, Y. Wang, and A. Song, "Projection of the impacts of global warming of 1.5 °C and 2.0 °C on runoff in the upper-middle reaches of huaihe River basin," *Journal of Natural Resources*, vol. 33, pp. 1966–1978, 2018.

[33] Y. Zhao, J. Zhu, and Y. Xu, "Establishment and assessment of the grid precipitation datasets in China for recent 50 years," *Journal of the Meteorological Sciences*, vol. 34, pp. 414–420, 2014.

[34] National Meteorological Information Center, *Daily Meteorological Dataset of Basic Meteorological Elements of China*, National Surface Weather Station (V3.0).

[35] Y. Ou, "Based on surface water dynamics research and application of grass slope," *Master's Thesis*, Nanchang University, 2012, (in Chinese).

[36] L. Touze-Peiffer, A. Barberousse, and H. Le Treut, "The Coupled Model Intercomparison Project: history, uses, and structural effects on climate research," *Wiley Interdisciplinary Reviews-Climate Change*, vol. 11, no. 4, 2020.

[37] D. P. van Vuuren, J. Edmonds, and M. Kainuma, "The representative concentration pathways: an overview," *Climatic Change*, vol. 109, no. 1-2, pp. 5–31, 2011.

[38] B. C. O'Neill, E. Kriegler, and K. L. Ebi, "The roads ahead: narratives for shared socioeconomic pathways describing world futures in the 21st century," *Global Environmental Change-Human and Policy Dimensions*, vol. 42, pp. 169–180, 2017.

[39] V. Eyring, S. Bony, G. A. Meehl et al., "Overview of the coupled model Intercomparison project phase 6 (CMIP6) experimental design and organization," *Geoscientific Model Development*, vol. 9, no. 5, pp. 1937–1958, 2016.

[40] Royal Netherlands Meteorological Institute, "Climate explorer".

[41] Y. Shang, S. Lu, Y. Ye et al., "China' energy-water nexus: hydropower generation potential of joint operation of the Three Gorges and Qingjiang cascade reservoirs," *Energy*, vol. 142, pp. 14–32, 2018.

[42] Z. Cui, *Operation Scheduling of Small and Medium Reservoirs*, Hohai University Press, Nanjing, China, 2017, (in Chinese).

[43] A. Ding, "Study on operations optimization of meishan reservoir," Master's Thesis, Hefei University of Technology, 2010.

[44] K. E. Taylor, "Summarizing multiple aspects of model performance in a single diagram," *Journal of Geophysical Research: Atmospheres*, vol. 106, no. D7, pp. 7183–7192, 2001.

[45] Y. Chen, F. Guo, J. Wang, W. Cai, C. Wang, and K. Wang, "Provincial and gridded population projection for China under shared socioeconomic pathways from 2010 to 2100," *Scientific Data*, vol. 7, no. 1, p. 83, 2020.

[46] Jinzhai County People's Government, *Report for Jinzhai County of the Seventh National Population Census*.

[47] S. Hu, L. Chen, L. Li et al., "Simulation of land use change and ecosystem service value dynamics under ecological constraints in Anhui province, China," *International Journal of Environmental Research and Public Health*, vol. 17, no. 12, 2020.

[48] T. Li, Z. H. Jiang, and H. L. Treut, "Machine learning to optimize climate projection over China with multi-model ensemble simulations," *Environmental Research Letters*, vol. 16, no. 9, 2021.

Hindawi
Advances in Meteorology
Volume 2022, Article ID 9717192, 9 pages
https://doi.org/10.1155/2022/9717192

Research Article

Air Pollutants Sources in Winter in Chang-Zhu-Tan Region of China

Yingfang Zhu,[1,2] **Juyang Liao**,[2,3,4] **Wei Gong**,[1] **Huili Wu**,[1] **Yaqi Huang**,[2,3] **Yan Liu**,[2,3] **and Meifang Zhao**[1]

[1]*College of Science, Central South University of Forestry and Technology, Changsha 410004, China*
[2]*Chang-Zhu-Tan Station for Scientific Observation and Research of Urban Agglomeration Ecosystems, Changsha 410116, Hunan, China*
[3]*Hunan Botanical Garden, Changsha 410116, China*
[4]*Beijing Forestry University, Beijing 10083, China*

Correspondence should be addressed to Juyang Liao; liaojuyang@163.com

Received 12 January 2022; Revised 15 February 2022; Accepted 23 February 2022; Published 25 March 2022

Academic Editor: Upaka Rathnayake

In order to analyze the primary sources of air pollutants in Chang-Zhu-Tan region, this article selected the environmental monitoring data and meteorological data in the winter of 2019 to calculate the backward airflow trajectories with the Chang-Zhu-Tan region as the starting point by using the backward trajectory model. Combined with the ground concentration monitoring data, cluster analysis, potential source contribution factor (PSCF) analysis, and concentration weighted trajectory (CWT) analysis were carried out to determine the pollutant transportation paths and sources of the potential source area. The results show that air mass transportation mainly comes from three directions: northwest, northeast, and southwest China. The airflow in northwest China moves faster and cleaner, while the airflow from the northeast and southwest moves slowly and carries a high concentration of pollutants. PSCF and CWT analyses show that the critical potential sources are mainly located in this area and some cities next to the study area. This study has important practical significance for the environmental research of Chang-Zhu-Tan region and can provide theoretical reference for regional joint prevention and control of air pollution.

1. Introduction

Regional air pollution is caused by human activities and local special geographical location, which has been the most severe environmental problem in recent years [1]. The spatial and temporal distribution characteristics of urban air pollution in China are prominent [2, 3]. In winter, significant disastrous weather is frequent, so air pollution is the most serious, followed by spring and autumn, and summer is the best [4]. Air pollution sources can be divided into two categories, natural and human-made sources. The anthropogenic source is the main aspect of air pollution. It is mainly produced by people's production activities and daily activities (industry, transportation, various combustion, garbage treatment, etc.). The primary pollutants include particulate matter (PM2.5, PM10), carbon monoxide (CO), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), and ozone (O_3) [5]. In most studies, scholars mainly analyzed the spatial-temporal pattern, pollution characteristics, and transportation path of PM2.5. However, there are few studies on the overall analysis of multiple pollutants [6].

The research contents of source analysis of air pollutants mainly include quantitative source analysis, cause and process analysis, and local source inventory analysis [7–18]. In the study of source analysis of air pollutants, some scholars used the constrained positive matrix decomposition method, Kriging interpolation method, positive matrix factorization method, and so forth [7, 8, 19]. HYSPLIT (Hybrid Single-Particle Lagrangian Integrated Trajectory) backward trajectory model is often used to deal with input fields of various meteorological elements and emission sources of different types of pollutants. The source,

transportation, air mass trajectory, diffusion, sedimentation, and other processes of air pollutants are also calculated and analyzed. The possible transmission path from the pollution source to the corresponding recipient area at a specific time is estimated to show the concentration contribution and spatial and temporal distribution characteristics of pollutants [9]. The research methods include trajectory cluster analysis, potential source contribution function (PSCF), and concentration weighted trajectory (CWT). Cluster analysis allocates trajectories into representative spatial groups. PSCF is a conditional probability function, which uses a backward trajectory to describe the geosynchronous orientation of potential sources and identify pollution sources. Because it only calculates the number of pollution trajectories, not the concentration of pollutants, it is difficult to distinguish pollution degrees of different pollution sources. CWT model can produce the difference of pollution trajectory in pollution degrees. Trajectory clustering analysis, PSCF, and CWT have been widely applied in published studies, and their combination can provide better information on the location of pollution sources [8, 9, 13–15, 20].

According to the information released on the website of the Ecology and Environment Department of Hunan, in recent years, Changsha, Zhuzhou, and Xiangtan (Chang-Zhu-Tan) have been at the bottom of Hunan province in the air quality index evaluation ranking. Chang-Zhu-Tan region is one of the national "two-type society" comprehensive reform experimental areas. In recent years, the apparent economic growth makes the task of environmental protection increasingly arduous. In the literature of air pollution source analysis, most of the domestic research areas are concentrated in Beijing-Tianjin-Hebei, Yangtze River Delta, north China, west China, and northwest China [8, 10–12, 15, 16, 20]. There are few related types of research in central China, especially in Chang-Zhu-Tan region [8, 10–12, 15, 16, 20].

In this study, the primary and potential sources of air pollutants in Chang-Zhu-Tan region in winter were studied based on hourly pollutant data by using trajectory cluster analysis, PSCF, and CWT. Our objectives were (i) to analyze the air mass transportation in Chang-Zhu-Tan region and (ii) to determine the pollutant transport path and pollution source of potential source area.

2. Materials and Methods

2.1. Site Description. Chang-Zhu-Tan Urban Agglomeration, including Changsha (28.23°N, 113°E), Zhuzhou (27.83°N, 113.16°E), and Xiangtan (27.87°N, 112.91°E) cities, is located in the central and eastern part of Hunan province in China. The three cities are distributed in the lower reaches of Xiangjiang River in an inclined "*T*" shape. Chang-Zhu-Tan belongs to the subtropical monsoon climate zone, with apparent seasonal variation, cold winter and hot summer, abundant precipitation, suitable temperature, less severe winter, more northwest wind, and vulnerability to cold wave and strong wind. The topography is mainly hills and plains, in which mountains, hills, plains, and rivers are distributed alternately. The terrain is high in the south and low in the north. The north is flat and open, and the south is hilly [21]. The area with an altitude below 100 m accounts for 88.72% of the total area of the study area.

2.2. Data Sources. In this study, the air quality data (mainly the concentration of pollutants) were from the China National Environmental Monitoring Centre. There are 24 ground air quality monitoring points in the study area, as shown in Table 1. The data collection time range was winter of 2019 (December to February of the next year), and the data's resolution was 1 hour. The meteorological data (including air pressure, dew point, wind direction, wind speed, temperature, cloud cover, precipitation, etc.) used in the backward trajectory model were recorded by the global data assimilation system of the United States every six hours, which are 0, 6, 12, and 18 o'clock, respectively [22].

2.3. Data Analysis

2.3.1. HYSPLIT Model. The backward trajectory model represents the latest state of the air mass, and the atmospheric mass trajectory can provide the information of the area affected by the point source. Taking Changsha, Zhuzhou, and Xiangtan as the simulation center, 500 m is selected as the average flow field of the atmospheric boundary layer near the surface of the study area. Based on the HYSPLIT 4 model developed by NOAA (National Oceanic and Atmospheric Administration) and BOM (Australia's Bureau of Meteorology), the 72 h backward airflow trajectories arriving at the research area every 6 hours from December 2019 to February 2020 are calculated. Driven by the global atmospheric reanalysis data of the National Centre for Environmental Prediction (NCEP), the model covers the whole process of transport, diffusion, and deposition and is widely used in pollutant transport path and source area analysis [23].

2.3.2. Trajectory Cluster Analysis. Cluster analysis is a multivariate statistical analysis method that classifies the original data according to their similarity and affinity [20]. The backward trajectory clustering in the HYSPLIT model is to cluster and group all air mass trajectories from the source to the receiving point according to the spatial similarity of air mass movement, including the velocity and direction of movement, to obtain different types of airflow trajectories. Based on the moving direction and velocity, the air mass trajectories in winter in Chang-Zhu-Tan are clustered. The backward trajectories of airflow movement are then classified according to the change rate of total space variance (TSV).

2.3.3. Potential Source Contribution Function (PSCF) Analysis. PSCF is a method based on conditional probability function to calculate and judge the geographical orientation of potential source area and further identify the pollution source area using airflow trajectory [13–15]. The

Table 1: Longitude and latitude of the air quality monitoring points in the study.

Monitoring point	Longitude	Latitude	Monitoring point	Longitude	Latitude
Jingkai district (Changsha)	113.08°E	28.23°N	Jianglu (Xiangtan)	112.89°E	27.87°N
Gaokai district (Changsha)	112.89°E	28.22°N	Yuetang (Xiangtan)	112.92°E	27.82°N
Hunan University of Chinese Medicine (Changsha)	112.89°E	28.13°N	Hunan University of Science and Technology (Xiangtan)	112.91°E	27.91°N
Hunan Normal University (Changsha)	112.94°E	28.19°N	Zhaoshan (Xiangtan)	113.00°E	27.92°N
Yuhua district (Changsha)	113.00°E	28.14°N	Shaoshan (Xiangtan)	112.49°E	27.92°N
Wujialing (Changsha)	112.98°E	28.26°N	Tiantao Hills (Zhuzhou)	113.14°E	27.82°N
New railway station (Changsha)	113.00°E	28.19°N	Zhuye hospital (Zhuzhou)	113.10°E	27.89°N
Tianxin district (Changsha)	112.98°E	28.12°N	No.4 Middle School (Zhuzhou)	113.17°E	27.87°N
Mapoling (Changsha)	113.08°E	28.21°N	Railway station (Zhuzhou)	113.14°E	27.84°N
Shaping (Changsha)	113.00°E	28.36°N	Zhuzhou (Zhuzhou)	113.13°E	27.85°N
Bantang (Xiangtan)	112.94°E	27.86°N	Dajing Scenic Spot (Zhuzhou)	113.25°E	27.83°N
Xiangtan (Xiangtan)	112.91°E	27.84°N	Yuntian Middle School (Zhuzhou)	113.18°E	28.00°N

study area is divided into i grids and calculated according to the formula.

$$\mathrm{PSCF}_i = \frac{M_i}{N_i}, \tag{1}$$

where M_i is the residence time in the i grid of all contaminated air mass tracks whose concentration exceeds the set threshold; N_i is the total residence time of all air mass trajectories in the i grid. The larger the result value, the higher the proportion of pollution tracks passing through the area. The corresponding region is more likely to be the potential source area of pollutants, and the trajectory of high value is the main path of pollutant transport.

2.3.4. Concentration Weighted Trajectory (CWT) Method. PSCF represents the proportion of pollution trajectories in all trajectories in each grid, only reflects the proportion of pollution trajectories, and cannot show the difference of pollution degree of pollution trajectories; that is, it cannot reflect the difference of pollution contribution of different source regions. Furthermore, the CWT of the trajectory can be calculated by using the concentration weighted trajectory analysis method, and the difference of the pollution trajectory in the pollution degree can be obtained to make up for the deficiency of the PSCF method. CWT [20] is used to calculate the weight concentration of trajectories, and the average weight concentration of grids is quantitatively given. The calculation formula is shown as follows:

$$\mathrm{CWT}_i = \frac{1}{\sum_{n=1}^{m} N_{in}} \sum_{n=1}^{m} C_{in}, \tag{2}$$

where CWT_i is the average weight concentration of grid i; n is the trajectory; N_{in} is the residence time of trajectory n in the i grid; m is the number of trajectories; C_{in} is the pollutant mass concentration corresponding to the trajectory n passing through the i grid.

3. Results

3.1. Analysis of Air Mass Transportation in Chang-Zhu-Tan Region. The distribution map of backward airflow trajectories in the winter of 2019 in Chang-Zhu-Tan is drawn, as shown in Figure 1. Obviously, the northwest, northeast, and southwest directions are the main transport directions of airflow in the study area, some of which are similar to the clockwise clover. The movement speed of the airflow is expressed by the length of the trajectory. The air mass moving faster leads to the trajectory longer. It can be seen that the airflow in the northwest direction is relatively rapid, mainly due to the influence of monsoon climate and strong wind.

3.2. Cluster Analysis of the Backward Trajectory. Based on the moving direction and velocity, the air mass trajectories of each month are clustered. The backward trajectories of airflow movement are divided into 3 or 4 categories according to the change rate of TSV value, as shown in Figure 2. The longest track reached about 40°N, and the air mass moved fastest. Compared with the trend of mountains in China, it can be seen that the track was along the Qilian Mountains, passing through the channel between Qinling Mountains and Daba Mountains into central China, passing through Inner Mongolia, Gansu, Shaanxi, Henan, and Hubei provinces and finally arriving at the research area. The trajectory of airflow transport from Hubei, the northern neighbouring province, was the shortest and the moving speed of air mass was slow. Simultaneously, combined with the air pollutant concentration data for statistical analysis, the occurrence probability of each type of trajectory and the corresponding average value of air pollutant concentration (see Table 2) are calculated to analyze the impact of different types of airflow on the concentration of pollutants. In order to facilitate analysis and comparison, the hourly mass concentration data of air pollutants in the winter of 2019 (December 2019 to February 2020) are used to calculate the average concentration of pollutants (see Table 3).

In December 2019, the fastest moving air mass was No. 1 from the north of the study area, passing through Shanxi, Henan, and Hubei, and the probabilities of its occurrence in all tracks in Changsha, Zhuzhou, and Xiangtan were 23%, 34%, and 25%, respectively. In the corresponding pollutant concentration table of the three places, the daily average level of O_3 was lower than the average value of daily average concentration in winter of 2019, and the concentrations of PM2.5, SO_2, NO_2,

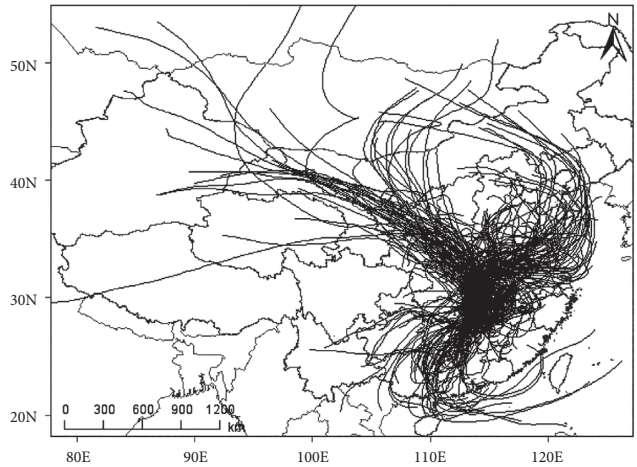


Figure 1: Distribution of backward airflow trajectories from December 2019 to February 2020 in Chang-Zhu-Tan region.

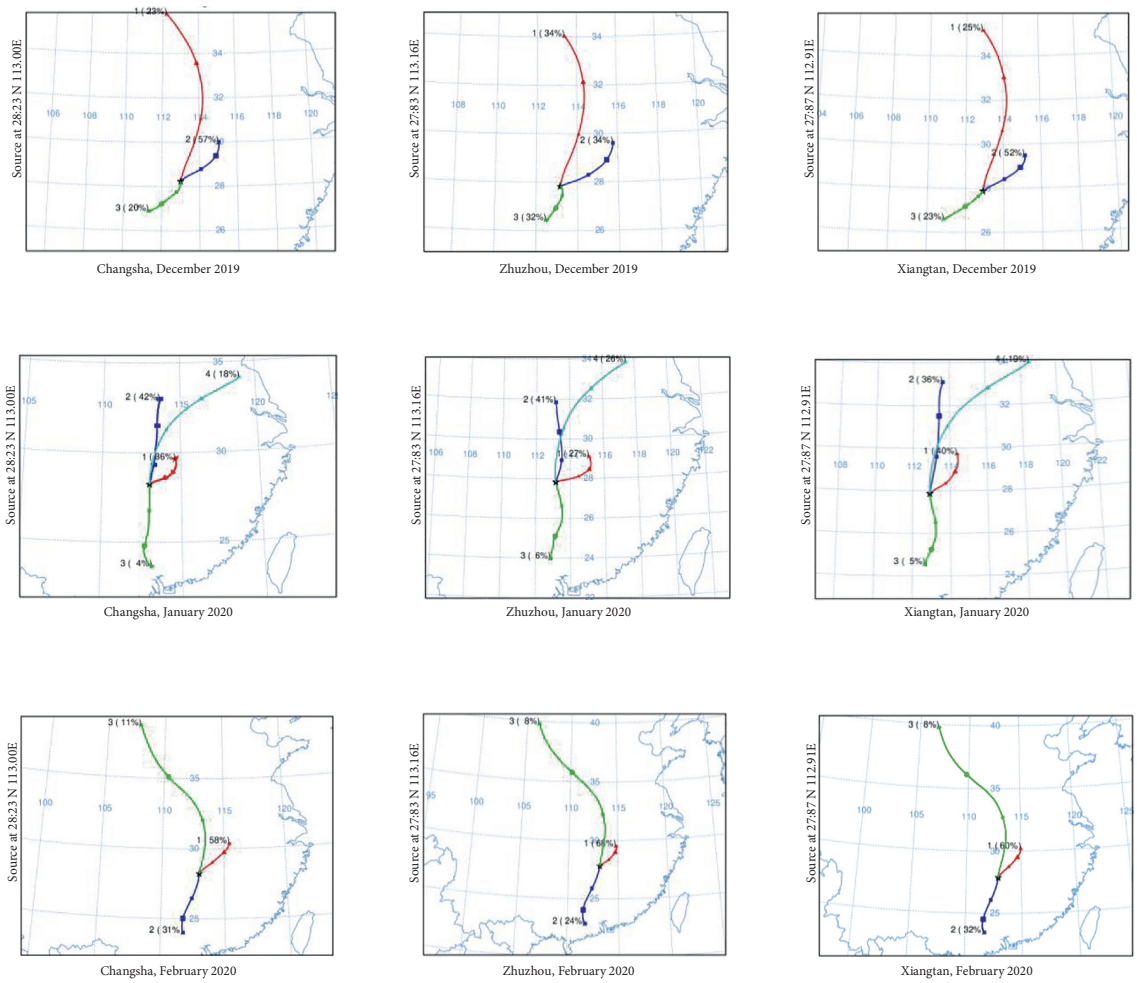


Figure 2: Clustering distribution of backward trajectories from December 2019 to February 2020 in Chang-Zhu-Tan region.

and CO in Changsha and Xiangtan were lower than the average values, and the carrying pollution was light. The concentration of PM10 in Changsha was 54.2 $\mu g/m^3$, close to the average value of 60.26 $\mu g/m^3$, but PM10 in Xiangtan was 70.7 $\mu g/m^3$ and that in Zhuzhou was 82.2 $\mu g/m^3$. It can be seen that Zhuzhou was the highest pollutant concentration affected by the No. 1 airflow trajectory, and the concentrations of SO_2, NO_2, and PM10 were higher than those of the other two cities. The No. 2 airflow trajectory passing through Jiangxi and Hubei provinces was the main transport channel, accounting for 57%, 34%, and 52% of

Table 2: Analysis of various trajectory paths and pollutant concentrations for December 2019 to February 2020 in Chang-Zhu-Tan.

Month	City	Track	Probability (%)	Airflow trajectories area	Average concentration of pollution ($\mu g/m^3$)					
					PM 2.5	PM 10	SO_2	NO_2	O_3	CO
December 2019	Changsha	1	23	Yueyang, Xianning, Western of Wuhan, Xiaogan, Zhumadian, Jincheng	63.5	54.2	7.2	39.7	26.5	1.1
		2	57	Jiujiang, Huangshi	87.4	84.4	8.7	56.6	26.4	1.1
		3	20	Xiangtan, Shaoyang	66.5	61.0	8.7	49.3	24.4	1.1
	Zhuzhou	1	34	Yueyang, Xianning, Western of Wuhan, Xiaogan, Zhumadian, Jincheng	65.2	82.2	10.8	46.9	25.3	0.9
		2	34	Jiujiang, Huangshi	82.0	98.9	12.2	58.2	27.0	1.1
		3	32	Xiangtan, Shaoyang	99.2	121.2	17.3	59.5	23.1	1.2
	Xiangtan	1	25	Yueyang, Xianning, Western of Wuhan, Xiaogan, Zhumadian, Jincheng	66.0	70.7	7.2	42.7	23.8	0.9
		2	52	Jiujiang, Huangshi	90.8	104.1	9.8	56.8	26.9	1.0
		3	23	Xiangtan, Shaoyang	60.7	67.5	11.6	47.6	23.0	1.1
January 2020	Changsha	1	42	Central Hubei, Central Henan	45.3	34.1	4.4	26.3	33.9	1.1
		2	36	Border of northwest Hunan and Hubei, Wuhan	73.4	61.8	5.4	37.8	30.1	1.2
		3	4	Central Hunan, Border of Hunan and Guangdong, Central Guangdong	62.6	53.0	5.6	36.6	20.8	1.3
		4	18	Central Hubei, Border of Hubei and Southern Henan, Border of Henan and Anhui, Border of Shandong and Anhui, Border of Shandong and Jiangsu	49.0	35.2	3.5	14.3	46.6	1.0
	Zhuzhou	1	27	Border of Western Hunan and Jiangxi, Central and Northern of Jiangxi	66.5	76.4	8.3	42.7	20.9	1.0
		2	41	Central Hubei, Central Henan	55.1	60.9	7.5	33.2	33.2	0.9
		3	6	Central Hunan, Border of Southeast Hunan and Guangdong, Central Guangdong	93.6	122.6	10.0	47.4	21.1	1.3
		4	26	Central Hubei, Border of Hubei and Southern Henan, Border of Henan and Anhui, Border of Shandong and Anhui, Border of Shandong and Jiangsu	36.9	34.7	5.3	22.7	33.6	0.8
	Xiangtan	1	40	Border of northwest Hunan and Hubei, Eastern Hubei	72.8	75.4	8.0	41.5	26.4	1.0
		2	36	Central Hubei, Central Henan, Border of Henan and Shansi	42.3	42.5	5.6	27.9	31.6	0.8
		3	5	Central Hunan, Border of Southeast Hunan and Guangdong, Central Guangdong	83.2	92.2	7.8	47.0	17.3	1.1
		4	19	Central Hubei, Border of Hubei and Southern Henan, Border of Henan and Anhui, Border of Shandong and Anhui, Border of Shandong and Jiangsu	46.8	42.7	4.0	19.3	40.1	0.8
February 2020	Changsha	1	58	Border of northeast Hunan and Hubei, Eastern of Hubei	45.6	36.6	4.7	15.0	50.7	0.8
		2	31	Central Hunan, Border of Southeast Hunan and Guangxi, Border of Eastern Guangxi and Guangdong	36.6	39.2	6.3	23.3	41.9	1.0
		3	11	Central Hubei, Border of Western Henan and Shanxi, Border of Shanxi and Shaanxi	15.7	19.9	4.7	6.5	66.3	0.6
	Zhuzhou	1	68	Border of northeast Hunan and Hubei, Southeast of Hubei	40.8	45.8	6.2	19.7	45.4	0.6
		2	24	Central Hunan, Border of Southeast Hunan and Guangxi, Border of Eastern Guangxi and Guangdong	30.6	33.9	8.0	19.9	46.8	0.8
		3	8	Central Hubei, Border of Western Henan and Shanxi, Border of Shanxi and Shaanxi, Border of Northern Shaanxi and Inner Mongolia	11.7	19.4	5.7	9.3	60.6	0.4
	Xiangtan	1	60	Border of northeast Hunan and Hubei, Southeast of Hubei	41.7	43.9	6.1	18.3	45.1	0.7
		2	32	Central Hunan, Border of Southeast Hunan and Guangxi, Border of Eastern Guangxi and Guangdong	37.2	44.2	8.9	25.3	35.3	1.0
		3	8	Central Hubei, Border of Western Henan and Shanxi, Border of Shanxi and Shaanxi, Border of Northern Shaanxi and Inner Mongolia	11.3	17.7	4.8	8.2	58.6	0.5

Table 3: Daily average concentration of pollutants in winter of 2019 in Chang-Zhu-Tan.

Pollution	Daily average concentration ($\mu g/m^3$)	National secondary standard limit
CO	0.93	4
NO_2	34.77	40
SO_2	7.78	150
PM2.5	59.43	75
PM10	60.26	150
O_3	55.70	160

the track probability of Changsha, Zhuzhou, and Xiangtan, respectively. The average concentrations of PM2.5, PM10, SO_2, and NO_2 carried by the airflow were the highest in Changsha and Xiangtan. The impact of the No. 3 airflow trajectory from the south on the concentration of atmospheric particulates corresponding to Changsha and Xiangtan was similar to that of No. 1, and the pollution carried by No. 3 had a more significant impact on Zhuzhou. The corresponding concentrations of PM2.5, PM10, SO_2, and NO_2 reached 99.2, 121.2, 17.3, and 59.5 $\mu g/m^3$, respectively, which were much higher than the average values in winter. No. 1, No. 2, and No. 3 air trajectories had little effect on O_3 and CO, and O_3 concentration was lower than the average level in winter. Among all the trajectories, the No. 2 airflow trajectory had a significant influence on the particulate matter concentration in Changsha and Xiangtan, while the concentrations of other pollutants except O_3 and CO in Zhuzhou were greatly affected by the No. 2 and No. 3 trajectories.

In January 2020, the three cities' backward trajectories were divided into four categories according to the change rate of TSV value. Among them, the first three types of tracks were similar to those of the previous month. The airflow of trajectory No. 4 accounted for 18%, 26%, and 19% of the total tracks of Changsha, Zhuzhou, and Xiangtan, respectively, passing through Henan, Anhui, the border areas of Shandong and Anhui, and the border areas of Shandong and Jiangsu. The air mass moved rapidly with long transmission distance and long trajectory. However, the average concentrations of PM2.5, PM10, SO_2, and NO_2 carried by the airflow were far lower than the seasonal average values; in particular, SO_2 was only between 3.5 and 5.3 $\mu g/m^3$, and the pollution is relatively light. However, compared with other airflow trajectories, the O_3 concentration carried by No. 4 was higher, about 40 $\mu g/m^3$. The concentration of air pollutants carried by the No. 2 airflow track from the north through central Hubei and central Henan was not high. Trajectory No. 3 accounted for the least proportion, accounting for 4%, 6%, and 5% of the total trajectories of Changsha, Zhuzhou, and Xiangtan, respectively. It carried about 1.3 $\mu g/m^3$ of CO through Guangdong to the central part of Hunan, which was higher than other airflow trajectories. The length of the No. 1 airflow trajectory imported from Hubei was relatively short, and the air mass movement speed was not fast. The PM2.5 concentration was about 10 $\mu g/m^3$ higher than the average seasonal value of 59.43 $\mu g/m^3$ and hovered around the national standard daily average concentration limit of 75 $\mu g/m^3$. It may be due to the impact of anthropogenic pollution, and the fine particulate matter pollution was more serious.

In February 2020, the main transport channel was the No. 1 airflow trajectory, accounting for 58%–68% of the total number of tracks. It passed through the eastern part of Hubei and the northeast of Hunan, with a short track and slow-moving air mass speed. The concentration of air pollutants carried except for O_3 by it was lower than the average seasonal value. Perhaps due to the impact of the SARS-CoV-2, during this period, the factory was shut down, and the traffic flow was reduced, so the air pollutants in the area along the route were also significantly reduced. Airflow trajectory No. 2 came from the border of eastern Guangxi and Guangdong and the central and southeast of Hunan. In addition to O_3, the concentrations of input air pollutants were also at a low level. The No. 3 airflow trajectory with the least probability of total trajectories, ranging from 8% to 11%, came from the boundary of northern Shaanxi and Inner Mongolia and passed through the border of Shanxi and Shaanxi, the border of western Henan and Shanxi, western Henan, and central Hubei. The trajectory was very long, and the air mass moved rapidly, but the concentrations of air pollutants carried by it were at a very low level except for O_3. Among the input pollutants, the concentrations of PM2.5 and PM10 were only 15.7 and 19.9 $\mu g/m^3$ in Changsha, 11.7 and 19.4 $\mu g/m^3$ in Zhuzhou, and 11.3 and 17.7 $\mu g/m^3$ in Xiangtan, while the concentrations of SO_2, NO_2, and O_3 were only 4.7–5.7, 6.5–9.3, and 0.4–0.6 $\mu g/m^3$, respectively. However, in February, the O_3 concentration brought by each airflow trajectory was not low; in particular, the O_3 concentration of No. 3 was above 60 $\mu g/m^3$, which was much higher than the average seasonal value.

On the whole, in winter, the airflow from northwest China carried fewer air pollutants and was relatively clean, while the airflow from the north by east and south passing through more developed areas had a higher average concentration of pollutants, and the pollution situation was relatively severe.

3.3. Analysis of Potential Source Area and Pollution Degree. In order to further analyze the air pollution transport source and confirm the potential source area of winter pollutants, the PSCF analysis of PM2.5 concentration in the study area was carried out by using MeteoInfo software. The contribution value of potential sources represents the proportion of pollution trajectories in the area. The greater the value, the more significant the contribution to PM2.5 concentration in the target study area and the greater the possibility of the region as the source of air pollution in the target area. The calculation results are shown in Figure 3.

In Figure 3, the darker the color, the greater the PSCF value. It can be seen that the PSCF value of PM2.5 in the study area has a wide span and a large coverage area, mainly concentrated in Hunan, Hubei, Jiangxi, and Henan provinces, which means that the main potential source areas of PM2.5 in the study area are concentrated in the adjacent areas of Chang-Zhu-Tan region and the surrounding developed areas.

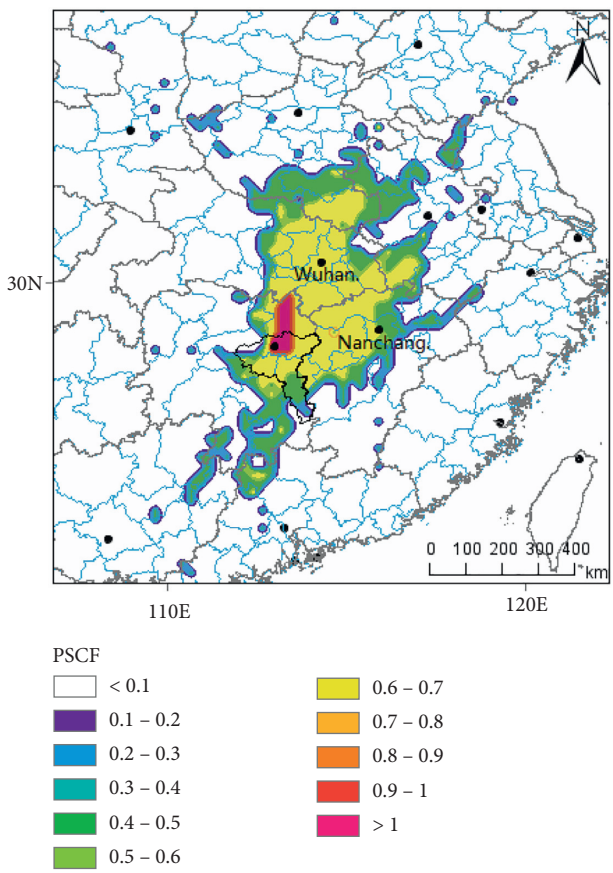


Figure 3: Contribution distribution of potential source areas of PM2.5 in winter of 2019 in Chang-Zhu-Tan region.

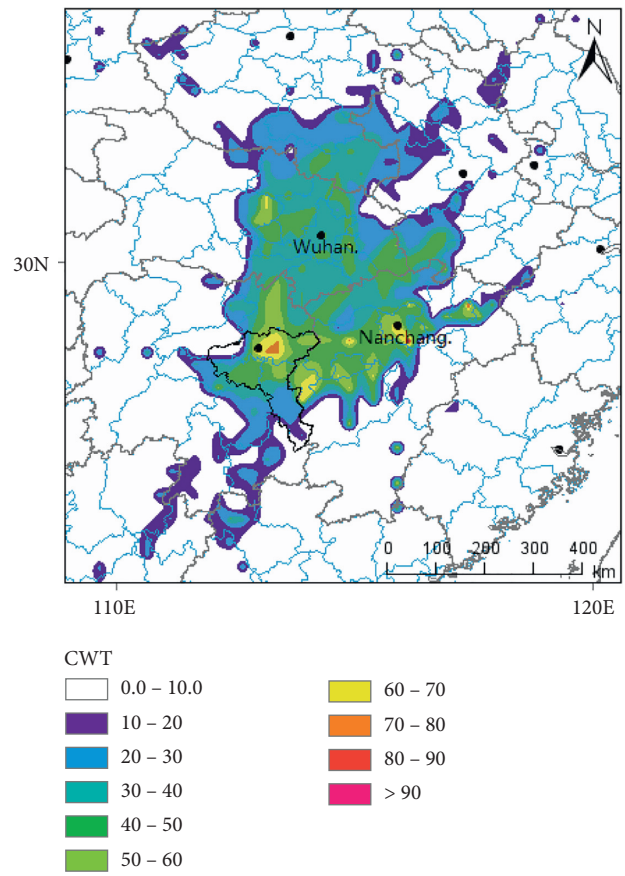


Figure 4: CWT contribution of PM2.5 in winter of 2019 in Chang-Zhu-Tan region.

The PSCF method calculates the proportion of pollution trajectories to all trajectories in a certain area to reflect the potential pollution impact of the region on the target area, which is only a conditional probability. On this basis, in order to quantitatively analyze the pollution level of different source regions, the CWT analysis method is used to calculate the weight concentration of each trajectory.

It is evident that the distribution characteristics of CWT and PSCF in winter in Chang-Zhu-Tan region are similar (Figure 4). The high-value range of CWT is concentrated in Hunan, Hubei, and Jiangxi. Unlike the distribution of PSCF high-value areas, CWT high-value areas are relatively scattered, which is not entirely centered on the study area. Hubei, Henan, Jiangxi, and Hunan are not connected to a large high-value region, so a more accurate source distribution is given. The main distribution areas are Chang-Zhu-Tan region and Yueyang city in Hunan province, Jingzhou city, and Xiaogan city in the central part of Hubei province, Ezhou region in the south of Wuhan City, Nanchang city in Jiangxi province, Pingxiang city in Jiangxi province bordering Zhuzhou city in Hunan province, central Jiangxi province (Xinyu and Ji'an city), and south of Jingdezhen in the northeast of Jiangxi. These areas are consistent with the key potential source areas analyzed by PSCF, and the contribution value of PM2.5 daily average mass concentration in the study area exceeds national secondary standard limit (75 $\mu g/m^3$). In addition, there are also a few small areas with a high contribution, such as Ganzhou city in the south of Jiangxi province and Xinyang city in the south of Henan province.

4. Discussion

In this paper, the potential source areas of PM2.5 in the study area are obtained by PSCF method, and the contribution weights of different source areas to PM2.5 concentration are obtained by CWT method. The distribution characteristics calculated by the two methods are close. Different from the distribution of PSCF high value area, the CWT high value area is relatively scattered and not completely centered on the study area. Compared with PSCF, the potential source areas determined by CWT method are more accurate, and this result may help to formulate pollution reduction strategies. There are three reasons that lead to the main potential sources of PM2.5 concentrated in the surrounding areas of Chan-Zhu-Tan area and the surrounding developed areas. Firstly, the industries in these areas are relatively developed, and the emission of air pollutants is high [24]. Secondly, the airflow through these areas moves slowly, and pollutants easily form accumulation. Thirdly, the pollutant diffusion effect is not good due to the short distance transmission, so it has a significant impact on the study area.

Due to seasonal heating and weather patterns with concentrated pollution in low altitude areas, the concentration of pollutants in winter in China is slightly higher than that in other seasons, and most areas cannot obtain long-term station data, so we chose the data of three months in winter for research. The estimated value of particulate pollution from the current study period may be similar to or slightly higher than the long-term average. However, because remote sensing data sets usually focus on annual or multiyear averages, this limits the ability to make direct comparisons.

The future work can be divided into two parts: (1) exploring seasonal changes and long-term trends combined with remote sensing data; (2) we can continue to establish the list of major pollutant emission sources in Chang-Zhu-Tan region and use various air quality models to simulate the complex process of atmospheric particulate matter diffusion and transformation. This will help to analyze the source of atmospheric particulate matter more scientifically and accurately.

5. Conclusion

Topography analysis shows that the northeast of Chang-Zhu-Tan area is the North China Plain and the Middle-Lower Yangtze River Plain, and there is no obvious mountain barrier for the airflow from the eastern ocean to central China. In the northwest direction, there are Qilian Mountains, Qinling Mountains, and Daba Mountains. In winter, Siberian cold air flows into the interior of China and even eastern China along the southeast trending Qilian Mountains. The south is hilly, and the airflow from the East China Sea and the South China Sea flows clockwise to central China. In the north, west, and southwest, there are Loess Plateau, Qinghai Tibet Plateau, and Yunnan Guizhou Plateau, which form natural barriers to block the airflow, so there is less airflow in these directions.

During the winter of 2019, as a whole, the trend of the airflow trajectories obtained by the backward trajectories clustering of the three cities in the study area was basically the same. The cluster trajectories were consistent with the leading wind directions in the meteorological analysis, and most of them were imported from the north by west, northeast, and south by west. Because of the significant influence of topography, the airflows were mostly imported along the mountain ranges. In all kinds of clustering airflow trajectories in winter, the average concentrations of pollutants carried by the airflow trajectories imported from eastern Hubei and northeastern Hunan were the highest, especially NO_2 and PM2.5, which exceed the national daily average concentration limits. The airflow trajectories were short, and the moving speeds of air masses were slow, which were affected by the short distance transmission of human-made pollution emissions in some areas, and the atmospheric diffusion condition was poor. In addition, the study period may be affected by the epidemic situation, factory shutdown, vehicle flow reduced, emissions reduced, and all kinds of air pollutants in the route area significantly reduced.

Data Availability

The data used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors would like to kindly thank Ms. Li (Qiaoyun Li), Ms. Wang (Ling Wang), and Ms. Zhang (Juan Zhang) for polishing this manuscript and thank Ms. Wang (Yaru Wang) and Ms Zhou (Hui Zhou) for their contribution to this study as jointly trained students. This work was supported by the National Natural Science Foundation of China (nos. 31971456 and 31600355) and the Hunan Province Special Project of Forestry Science and Technology Innovation of China (nos. XLK202103-2 and XLK201942).

References

[1] W. X. Wang, F. H. Chai, Z. H. Ren et al., "Process, achievements and experience of air pollution control in China since the founding of the People's Republic of China 70 years ago," *Research of Environmental Sciences*, vol. 32, no. 10, pp. 1621–1635, 2019.

[2] L. Yin, L. Wang, W. Huang, S. Liu, B. Yang, and W. Zheng, "Spatiotemporal analysis of haze in Beijing based on the multi-convolution model," *Atmosphere*, vol. 12, no. 11, p. 1408, 2021.

[3] Y. Liu, J. Tian, W. Zheng, and L. Yin, "Spatial and temporal distribution characteristics of haze and pollution particles in China based on spatial statistics," *Urban Climate*, vol. 41, Article ID 101031, 2022.

[4] S. D. Liu, J. N. Shi, and Y. Cheng, "Review of pollution characteristics of PM 2.5 in Chinese representative megacities," *Research of Environmental Sciences*, vol. 33, no. 2, pp. 243–251, 2020.

[5] B. Merete, B. Kelley, B. Murray, D. Christopher, and R. Ilona, "Saturation vapor pressures and transition enthalpies of low-volatility organic molecules of atmospheric relevance: from dicarboxylic acids to complex mixtures," *Chemical Reviews*, vol. 115, no. 10, pp. 4115–4156, 2015.

[6] M. L. Acosta-Urdapilleta, E. Villegas, E. Villegas, A. Estrada-Torres, M. Téllez-Téllez, and G. Díaz-Godínez, "Antioxidant activity and proximal chemical composition of fruiting bodies of mushroom, Pleurotus spp. produced on wheat straw," *Journal of Environmental Biology*, vol. 41, no. 5, pp. 1075–1081, 2020.

[7] A. Fulvio and K. H. Philip, "Source apportionment of the ambient PM 2.5 across St. Louis using constrained positive matrix factorization," *Atmospheric Environment*, vol. 46, no. 1, pp. 329–337, 2012.

[8] Y. Liu, Y. Yu, M. Liu et al., "Characterization and source identification of PM2.5-bound polycyclic aromatic hydrocarbons (PAHs) in different seasons from Shanghai, China," *Science of The Total Environment*, vol. 644, pp. 725–735, 2018.

[9] B. Moroni, S. Crocchianti, C. Petroselli et al., "Potential source contribution function analysis of long-range transported aerosols in the central mediterranean: a comparative study of

two background sites in Italy," *Rendiconti Lincei. Scienze Fisiche e Naturali*, vol. 30, no. 2, pp. 337–349, 2019.

[10] Y. Wang, H. Wang, and S. Zhang, "A weighted higher-order network analysis of fine particulate matter (PM2.5) transport in Yangtze River Delta," *Physica A: Statistical Mechanics and its Applications*, vol. 496, no. 3, pp. 654–662, 2018.

[11] Z. D. Zhang, T. J. Shao, X. G. Huang, and P. R. Wei, "Characteristics and potential sources of PM 2.5 pollution in Beijing-Tianjin-Hebei region in 2017," *Environmental Engineering*, vol. 38, no. 2, pp. 99–106, 2020.

[12] F. R. Deng, N. Kang, H. Kang, Y. C. Jiang, and Y. Xing, "Analysis of air pollution episodes over different cities in the Yangtze River Delta," *China Environmental Science*, vol. 38, no. 2, pp. 401–411, 2018.

[13] E. Potier, A. Waked, A. Bourin et al., "Characterizing the regional contribution to PM10 pollution over northern France using two complementary approaches: Chemistry transport and trajectory-based receptor models," *Atmospheric Research*, vol. 223, pp. 1–14, 2019.

[14] P. Praphatsorn and C. Somporn, "Identification of potential sources of PM 10 pollution from biomass burning in northern Thailand using statistical analysis of trajectories," *Atmospheric Pollution Research*, vol. 9, no. 6, pp. 1038–1051, 2018.

[15] K. Yang, Q. Li, M. Yuan et al., "Temporal variations and potential sources of organophosphate esters in PM2.5 in Xinxiang, North China," *Chemosphere*, vol. 215, pp. 500–506, 2019.

[16] C. L. Zhu, S. S. Meng, and R. G. Zhang, "Correlation analysis and spatial and temporal distribution characteris of main atmospheric pollutants in Xi'an," *Environmental Engineering*, vol. 35, no. 12, pp. 86–91, 2017.

[17] K. Peng, Y. Yang, J. Y. Zheng, S. S. Yin, Z. J. Gao, and X. B. Huang, "Emission factor and inventory of paved road fugitive dust sources in the Pearl River Delta region," *Acta Scientiae Circumstantiae*, vol. 33, no. 10, pp. 2657–2663, 2013.

[18] X. Zhou, T. J. Zhang, Z. Q. Li, Y. Tao, and F. T. Wang, "Particulate and gaseous pollutants in a petrochemical industrialized valley city, Western China during 2013–2016," *Environmental Science and Pollution Research*, vol. 25, no. 15, pp. 15174–15190, 2018.

[19] R. A. Rohde and R. A. Muller, "Air pollution in China: mapping of concentrations and sources," *PLoS One*, vol. 10, no. 8, Article ID e0135749, 2015.

[20] Y. Chen, S. Xie, and B. Luo, "Seasonal variations of transport pathways and potential sources of PM2.5 in Chengdu, China (2012-2013)," *Frontiers of Environmental Science & Engineering*, vol. 12, no. 1, p. 12, 2018.

[21] Y. Li, *Study on Land Use Change and its Effects on Eco-Environment in CZT Core Region*, Dissertation, Hunan Agricultural University, Changsha, Hunan, China, 2015.

[22] C. Yang, F. Gao, and M. Dong, "Energy efficiency modeling of integrated energy system in coastal areas," *Journal of Coastal Research*, vol. 103, no. 1, p. 995, 2020.

[23] J. Li, N. Chen, Y. Zhao, M. Liu, and W. Wang, "A catastrophic landslide triggered debris flow in China's Yigong: factors, dynamic processes, and tendency," *Earth Sciences Research Journal*, vol. 24, no. 1, pp. 71–82, 2020.

[24] L. He, Y. Chen, and J. Li, "A three-level framework for balancing the tradeoffs among the energy, water, and air-emission implications within the life-cycle shale gas supply chains," *Resources, Conservation and Recycling*, vol. 133, pp. 206–228, 2018.

Hindawi
Advances in Meteorology
Volume 2022, Article ID 1433835, 13 pages
https://doi.org/10.1155/2022/1433835

Research Article

Evaporation Rate Prediction Using Advanced Machine Learning Models: A Comparative Study

Zainab Abdulelah Al Sudani [1] **and Golam Saleh Ahmed Salem** [2]

[1] *Water Resources Department, College of Engineering, University of Baghdad, Baghdad, Iraq*
[2] *Department of Electrical and Electronic Engineering, Trust University, Nobogram Road, Barishal-8200, Bangladesh*

Correspondence should be addressed to Golam Saleh Ahmed Salem; dr.salem@trustuniversity.edu.bd

Received 14 November 2021; Revised 17 January 2022; Accepted 20 January 2022; Published 21 February 2022

Academic Editor: Upaka Rathnayake

Accurately estimating the amount of evaporation loss is necessary for scheduling and calculating irrigation water requirements. In this study, four machine learning (ML) modeling approaches, extreme learning machine (ELM), gradient boosting machine (GBM), quantile random forest (QRF), and Gaussian process regression (GPR), have been developed to estimate the monthly evaporation loss over two stations located in Iraq. Monthly climatical parameters have been used as an input variable for simulating the evaporation rate. Several statistical measures (e.g., mean absolute error (MAE), correlation coefficient (R), mean absolute percentage error (MAPE), and modified index of agreement (Md)), as well as graphical inspection, were used to compare the performances of the applied models. The results showed that the GBM model has much better performance in predicting monthly evaporation over two stations compared to other applied models. For the first case study which was in Diyala, the results showed a prediction enhancement in terms of MAE and RMSE by 7.17%, 21.01%; 16.51%, 15.74%; and 23.14%, 26.64%; using GBM compared to ELM, GPR, and QRF, respectively. However, for the second case study (in Erbil), the prediction enhancement was improved in terms of reduction of MAE and RMSE by 10.88%, 9.24%; 15.24%, 5%; and 16.06%, 15.76%; respectively, compared to ELM, GPR, and QRF models. The results of the proposed GMBM model can therefore assist local stakeholders in the management of water resources.

1. Introduction

In the hydrological cycle, evaporation plays a major role; therefore, monitoring evaporation is important for managing water resources, optimizing irrigation schedules, and modeling agricultural production [1,2]. Besides, evaporation rate has significant importance in studying climate change and global warming because this parameter dissipates a good proportion of the global precipitation [3–5]. The evaporation loss is influenced primarily by the vapor pressure gradient and the available heat energy, which are determined by the weather data like air temperature, relative humidity, wind speed, and solar radiation [6–8]. These variables are strongly associated with other aspects like the current season, time of day, geographical location, and sort of climate [9,10]. The evaporation process is therefore extremely nonlinear and complex.

For computing and evaluating evaporation, there are two procedures, direct and indirect [11]. Pan evaporation E_{pan} is considered as a well-known direct method used extensively for the estimation of evaporation rate. In particular, evaporimeters cannot be placed everywhere, especially in inaccessible regions where precise instrumentation is not possible [12]. Furthermore, the process of installing and maintaining this evaporation equipment in several regions is expensive [13]. However, the indirect method includes empirical equations used for measuring the evaporation rate [14]. These empirical equations can be established utilizing meteorological and hydrological parameters such as temperature, sunshine hour, wind speed, humidity, and rainfall [15,16]. Precise measurement of some of these meteorological factors requires advanced tools and skilled labor [17]. Often, instrument malfunctions, improper maintenance, and harsh weather conditions make it difficult to gauge these

data minus any errors, which is essential for the prediction of evaporation via empirical equations [18]. Thus, it would be problematic to project evaporation by gauging these factors incorrectly [19].

Thus, indirect systems of estimating evaporation by applying empirical equations are dependent on data and are also influenced by different assumptions. In other words, these approaches are considered as data-sensitive procedures and the accuracy of prediction would mainly depend on the data validity [20]. Additionally, such climatic data are generally scarce or hard to find at a particular hydrological station, and they tend to be discontinuous in certain places [21]. Evaporation is difficult to model through empirical techniques due to its extremely complex physical and nonlinear nature. In addition, an empirical model designed for a specific scenario might not perform well in another scenario, requiring recalibrations of the coefficients before execution. Several empirical models have been created by many researchers in literature to model evaporation loss [22]. The selection of the predictors is one of the main challenges for the nonlinear regression process. Therefore, creating a robust predictive model using empirical procedures is very difficult.

Many studies have been conducted to solve different water-resource problems employing different artificial intelligence (AI) approaches such as random forest (RF), support vector machine (SVM), extreme learning machine (ELM), feed-forward neural network (FFNN), extra-tree, Gaussian process regression (GPR), gradient boosting model (GBM), and quantile regression forest (QRF) [23–29]. Goyal et al. [30]. presented a study to estimate the daily evaporation loss over subtropical areas using different AI modeling approaches. The study used six meteorological parameters to establish the applied models. The findings of the study illustrated that the Adaptive Neurofuzzy Inference System (ANFIS) and least square support vector regression (LS-SVR) provide the best accuracy compared to the other used models. Another study was performed in [31] to estimate the evaporation loss of the Beysehir lake located in the southern part of Turkey. This study employed several machine learning approaches coupled with cross-validation technique to predict the monthly evaporation over that case study which is characterized as an arid and semiarid area. The study found that both ANN and SVR had a good prediction accuracy. Qasem et al. [32] developed a complicated model based on the incorporation of the ML models such as SVR and ANN with wavelet transforms (WT) for modeling the monthly rate of evaporation in arid and humid climates. The obtained results showed that the WT did not significantly enhance the prediction accuracy in some cases. Besides, the standard model (ANN) showed satisfactory accuracy in terms of predicting the evaporation rates. As ANN showed higher performance in prediction evaporation loss, it is significant to compare ANN with other machine learning methods such as RF and ELM. A study introduced by [33] provided a good comparison between the performances of ANN and random forest in the prediction of evaporation. The study's result proved that the RF has better performance than ANN as well as providing very accurate estimates. Furthermore, Althoff et al. [34] presented a study using different ML approaches to estimate the small dams' evaporation loss in Brazil. The findings of the study illustrated that the performance of RF was very satisfactory in the prediction of evaporation loss over small dams. Several other research evidenced the contribution of the AI models in simulating the catchment evaporation processes [35–37]. Recently, kernel-based models, fuzzy algorithms, and their hybrids with other algorithms have been successfully used for predicting evaporation [38]. However, developed gradient boosting models were rarely applied in modeling reference evapotranspiration worldwide. According to our knowledge, no study has focused on evaluating and comparing the capability of newly developed gradient boosting models for evaporation estimation in arid to semiarid climate zones of Iraq. Therefore, it is interesting to evaluate the performance of GBM and compare it with reliable AI models such as extreme learning machine (ELM), quantile regression forest (QRF), and Gaussian process regression (GPR) for estimating evaporation rate (E_p) in arid to semiarid climate zones of Iraq.

The contribution of this study is to determine the efficiency of the gradient boosting model (GBM) in estimating the evaporation rate (E_p) using data collected from two meteorological stations located in Iraq. The performance of GBM was compared with those of reliable AI models such as extreme learning machine (ELM), quantile regression forest (QRF), and Gaussian process regression (GPR). Furthermore, it is the first time to use GBM model for predicting the monthly evaporation loss related to several stations located in Iraq.

2. Data and Case Study

Iraq is geographically located in the Middle East and has almost two major climate zones, semiarid in the south and semihumid in the north [39]. The Iraqi region lacks sufficient water resources and suffers from droughts [40,41]. As temperatures rise in Iraq, surface water availability decreases, and groundwater levels in aquifers decrease. Iraq's hydrological cycle has been affected severely by evaporation, which currently depletes about 61% of its total precipitation [16,42]. Thus, it is very important to accurately predict the evaporation loss in Iraq. In this study, two case studies are selected to estimate the evaporation rate. The first case study is in Diyala state, while the second station is in Erbil state (see Figure 1). Diyala is located in the central part of the region, while Erbil is located in the northern region. The evaporation rate was predicted as function of six metrological parameters such as sunshine hours, minimum and maximum temperature, wind speed, rainfall, and relative humidity.

3. Methodology

3.1. Gaussian Process Regression. Rasmussen and Williams were the first to introduce the Gaussian process regression (GPR) [43]. This approach is a well-known and nonparametric method used for solving classification and regression

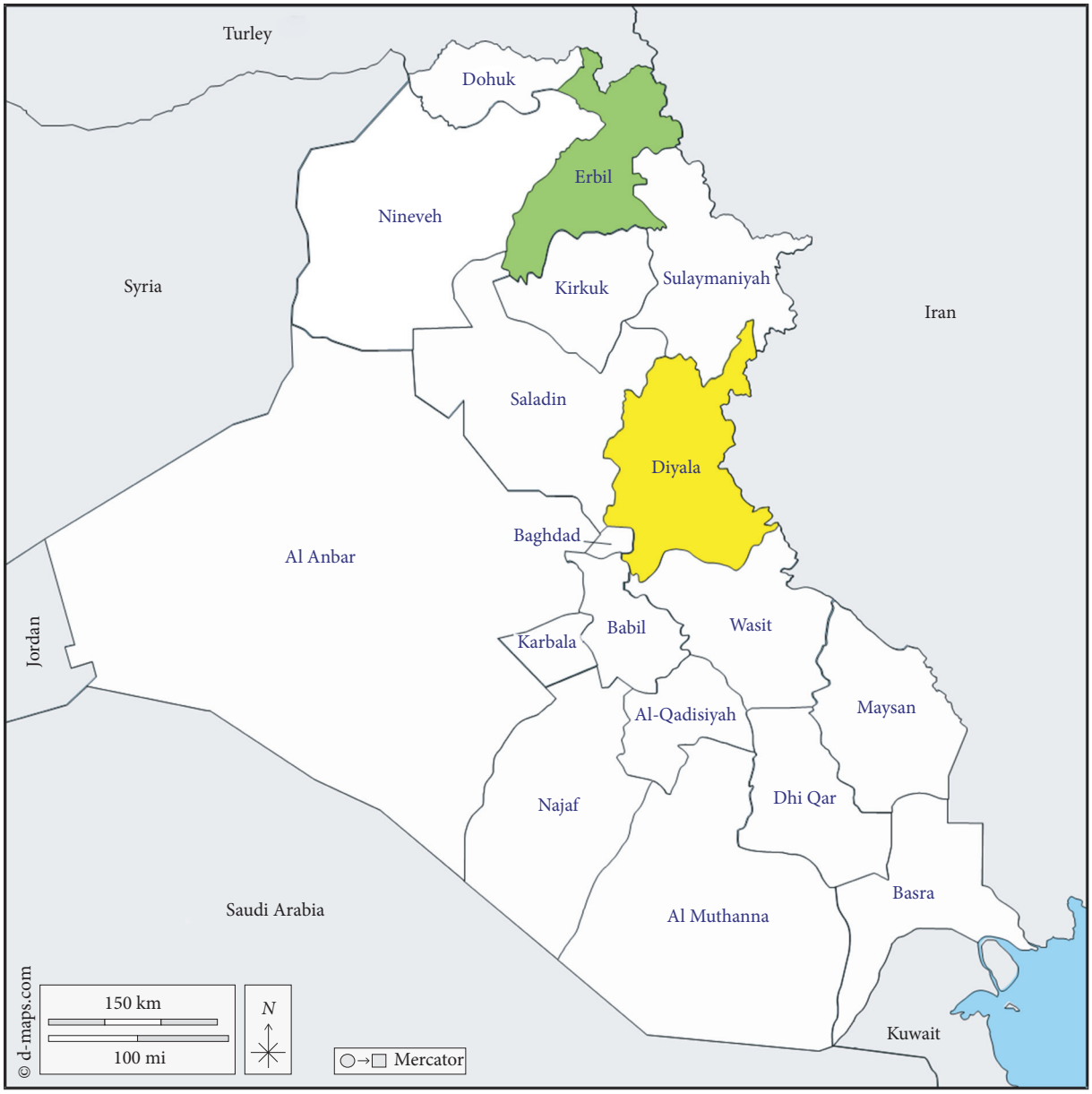


Figure 1: The locations of the selected region and the states.

problems. Furthermore, GPR model has been commonly employed to address several water resources concerts [44–47]. GPR combines Bayesian learning and kernel machines to form a principled and probabilistic approach to create a regression model. A model prediction's uncertainty can be directly outputted alongside the projected value [48].

In general, the mean and kernel function can be used to calculate a GPR [49]. According to this definition, GPR is an assemblage of random variables representing the value of function $f(t)$ at the given location (t). It can be expressed as follows:

$$\begin{aligned} f(t) &\sim GPR(m(t), k(t,t')), \\ m(t) &= E[f(t)], \\ k(t,t') &= E[(f(t)-m(t))(f(t')-m(t'))]. \end{aligned} \tag{1}$$

$f(t)$ is the prior distribution of the regression function, and $k(t)$, $an\ d\, m(t)$ are the kernel and function, respectively. By considering that the training set T includes input finite numbers in a matrix form $t_1, t_2, \ldots t_n$, the joint distribution of GPR is defined as follows:

$$p(f|TR) = N(f|M, K), \tag{2}$$

where $M(T)$ is the mean function which can be calculated by the mean function $m(t)$ as follows:

$$M(T) = \begin{bmatrix} m(t_1) \\ m(t_2) \\ \cdots \\ m(t_n) \end{bmatrix}. \tag{3}$$

Moreover, the kernel function $K(T,T)$ of the applied model can be determined by mean function $k(t,t')$ as follows:

$$K(T,T) = \begin{bmatrix} k(t_1, t_1') \dots (t_1, t_N') \\ \dots\dots \\ k(t_N, t_1') \dots (t_N, t_N') \end{bmatrix}. \quad (4)$$

In this study, the mean function is set to zero for simplicity to produce a widely used GPR prior. Besides, this technique has been widely used in previous studies [43,50]. Finally, (1) will be rewritten as follows:

$$f(t) \sim GPR(m(t) = 0, k(t,t')). \quad (5)$$

3.2. Extreme Learning Machine. Extreme learning machine (ELM) has the advantages of being a single hidden layer feed-forward neural network (FFNN) with good global search ability, simple structure, fast learning speed, and excellent generalization abilities [51]. There are two types of weights in the ELM: the input weights related to the hidden layer which are assigned randomly and the output weights which are attained by analysis and calculation [52]. In other words, unlike traditional neural networks, the ELM does not require iterative learning [53]. The outputs weights of the ELM can be easily computed by determining the generalized inverse of the output matrix of the hidden output weight values. The structure of the ELM is greatly simplified by this process. The training process of ELM is summarized by few steps as follows:.

(i) Input the training dataset, and select the ELM's structure (hidden nodes) and the activation function of the hidden layer (see Figure 2).

(ii) Calculate the H matrix (output of hidden layer) as follows:

$$H = (a_1, \dots, a_{l;} b_1, \dots, b_{l;} x_1, \dots, x_n) = \begin{bmatrix} g(a_1, b_1, x_1) g(a_l, b_l, x_n) \\ \dots \\ g(a_1, b_1, x_n) g(a_l, b_l, x_n) \end{bmatrix}_{l*n}. \quad (6)$$

$(a_{i,} b_i)$, $i = 1, 2, \dots, l$, are hidden nodes parameters which are randomly assigned.

(iii) Determine the output weight matrix (β):

$$\beta = H^+ T, \quad (7)$$

where T is the actual label vector of the training dataset and H^+ is Moore-Penrose generalized inverse matrix (H). $H^+ = (H^T H)^{-1} H^T$.

$$\beta = \begin{bmatrix} \beta_1{}^T \\ \cdot \\ \cdot \\ \beta_l{}^T \end{bmatrix}_{l*m}, \quad T = \begin{bmatrix} T_1{}^T \\ \cdot \\ \cdot \\ T_n{}^T \end{bmatrix}_{n*m}. \quad (8)$$

3.3. Quantile Random Forest (QRF). Random forest (RF) is an ensemble and supervised learning algorithm invented by Breiman [54]. The core concept of this approach is to integrate multiple trees through ensemble learning procedures. Furthermore, RF is a modified version of the Bagging algorithm with the basic idea that, for the original dataset, S_n are selected as a new data and S_n would be trained by using put back sampling method separately. The CART decision tree in RF is employed as a weak learner; however, for each tree is generated, the required number of features will be selected randomly from the original dataset labels. Thus, in a regression problem, the results of weak learners (T) are averaged to obtain the final model output. Averaging approach of RF has a significant importance in reducing the bias, as well as variance and correlation between trees [23].

The quantile random forest (QRF) is considered as an improved version of RF, applying quantile regression (QR) instead of averaging approach in calculating the final form of a target [55]. Furthermore, the QRF is considered a non-parametric approach enhanced by a solid theoretical foundation [56]. The conditional distribution of the QRF can be mathematically expressed as follows:

$$F(y|X = x) = P(Y \le y|X = x) E\left(l_{\{y_{i \le y}\}|} X = x\right). \quad (9)$$

$E(l_{\{y_{i \le y}\}|} X = x)$ in (8) is derived by taking mean value of the observations. With regard to QRF, $E(l_{\{y_{i \le y}\}|} X = x)$ is representing the weighted average value of all observations $l_{\{y_{i \le y}\}|}$.

$$\widehat{F}(y \mid X = x) = \sum_{i=1}^{n} W_i(x) l_{\{y_{i \le y}\}|}. \quad (10)$$

The steps below illustrate the QRF algorithm:

(i) The M decision tree $T(\theta_t)$, $t = 1, \dots k$, is created in random forests (RF) as well as taking into account the observations of each node related to a decision tree.

(ii) For $X = x$, o it will be repeated for all decision trees and then determine all observations of each decision tree. Finally, the weight $w_i(x, \theta_t)$ of each observation $i \in \{1, \dots, n\}$ is calculated by averaging the weights of tree decisions.

(iii) For all $y \in R$, calculate the estimate of the distribution function with (9) by using the weights obtained in step (2).

Figure 3 presents the flowchart of the QRF model.

3.4. Gradient Boosting Machine. Gradient boosting machine (GBM) model is one of the most famous supervised algorithms introduced as a robust technique to solve problems related to classification and regression [57]. Decision tree is a faster algorithm but it still suffers instability, so GBM is introduced to solve this serious problem [58–60]. Furthermore, GBM has combined the decision tress and boosting algorithms' advantages [61]. The GBM works mainly on the formulation of the gradient descent of

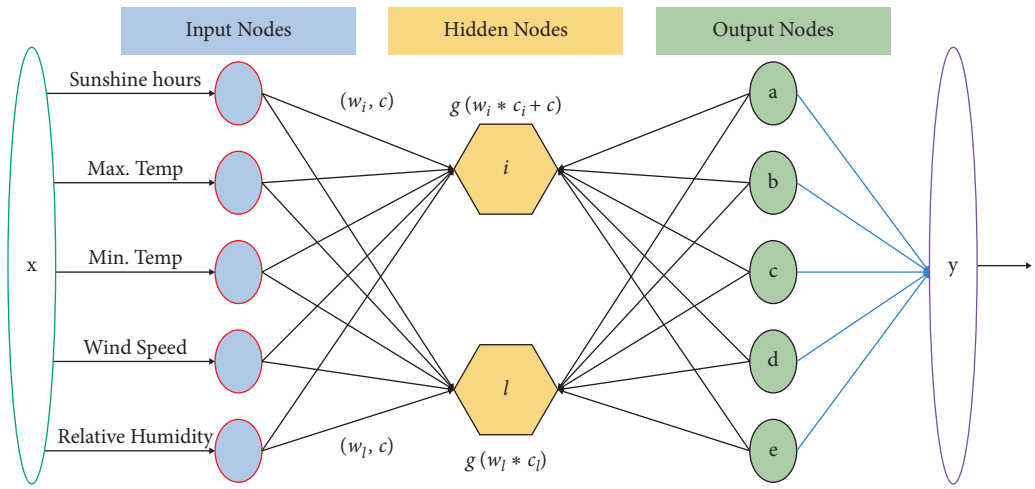


Figure 2: The basic structure of ELM model.

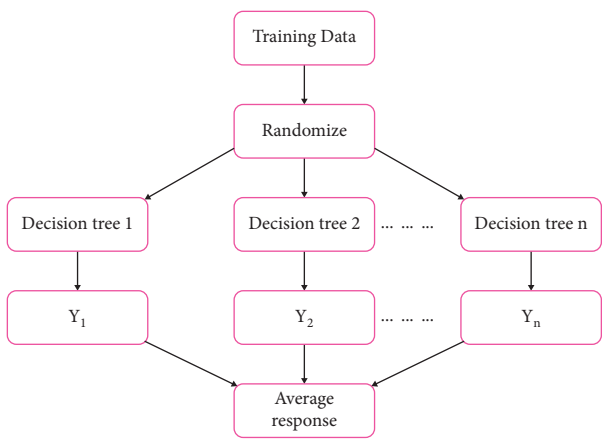


Figure 3: The flowchart of the quantile regression forest algorithm.

boosting technique and, hence, it is very useful for classification and regression problems [62]. The boosting structure is primarily a constructive scheme of ensemble formation that involves successively adding new weak base models that are trained according to the calculated error of the previous whole ensemble model for each iteration, and these base learners generate only a slightly lower error rate compared to random guessing. The boosting method family is based on a constructive strategy in which the learning mechanism will fit new models sequentially to produce a more precise estimation of the response variable. Figure 4 shows the structure of gradient boosting machine regression model.

The approach of the GBM model can be illustrated in several steps as follows:

(i) The GMB is initialized to minimize the loss function with a constant value.

(ii) The negative gradient of the cost function is estimated in each iterative training process as the residual value in x_i model (current one).

(iii) A new regression tree will be trained to fit the residual obtained from the second step.

(iv) In this step, the residual is updated and the current regression tree is added to the previous model.

(v) The algorithm of GBM is still iteratively trained and the maximum iterations number (selected by the user) is reached.

The mathematical expressions and brief description of applied GMB algorithm are shown below [63].

3.5. Statistical Evaluation Metrics. The four applied models have been compared and assessed to select the best models for predicting monthly evaporation. There are five statistical criteria, root mean square error (RMSE), mean absolute error (MAE), correlation coefficient (R), mean absolute percentage error (MAPE), and modified index of agreement (Md), which were used to assess the models' performances for training and testing phases. The mathematical expressions of these parameters are illustrated below [64]:

$$\begin{aligned}
\mathrm{MAE} &= \sum_{i=1}^{n} \frac{\left|\mathrm{EP}_{obi} - \mathrm{EP}_{smi}\right|}{n}, \\
\mathrm{RMSE} &= \sqrt{\sum_{i=1}^{n} \frac{\left(\mathrm{EP}_{obi} - \mathrm{EP}_{smi}\right)^2}{n}}, \\
\mathrm{MAPE} &= \frac{1}{n} \sum_{i=1}^{n} \frac{\left|\mathrm{EP}_{obi} - \mathrm{EP}_{smi}\right|}{\mathrm{EP}_{obi}}, \\
R &= \frac{\sum_{i=1}^{n} \left(\mathrm{EP}_{obi} - \overline{\mathrm{EP}_{ob}}\right)\left(\mathrm{EP}_{smi} - \overline{\mathrm{EP}_{sim}}\right)}{\sum_{i=1}^{n} \left(\mathrm{EP}_{obi} - \overline{\mathrm{EP}_{ob}}\right)^2 \sum_{i=1}^{n} \left(\mathrm{EP}_{smi} - \overline{\mathrm{EP}_{sim}}\right)}.
\end{aligned} \tag{11}$$

In the above equations, EP_{obi} and EP_{smi} are the actual and predictive monthly evaporation values at i – th record, respectively. $\overline{\mathrm{EP}_{ob}}$ and $\overline{\mathrm{EP}_{sim}}$ are the mean observed and predicted monthly evaporation values and n is the number of records Algorithm 1.

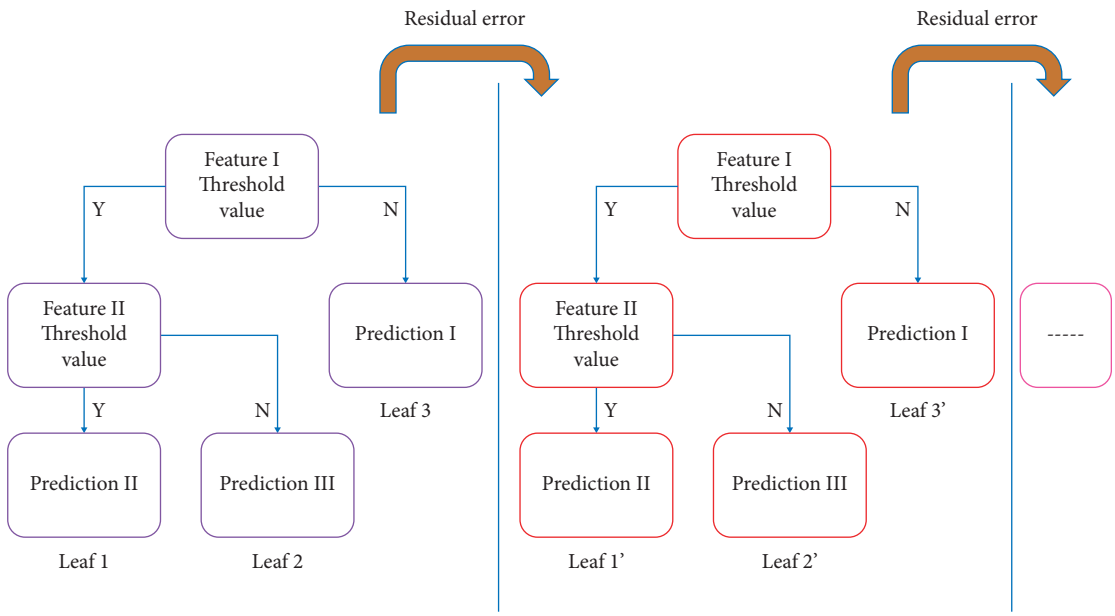


Figure 4: The structure of gradient boosting machine regression model.

Input: Train data
Data includes $T = \{(x_1, y_1), (x_2, y_2)\ldots..(\text{xn, yn}), x\&\, y \in R$
Loss function $L(y, f(x))$.
Output: regression tree $\widehat{f}(x)$;
(i) Initialization $f_{o\,x} = argmin_c \sum_{i=1}^{N} (y_i, c)$
(ii) $for\, m = 1\, To\, M$
(a) For i T N, calculating the residuals.$r_{mi} = -[\partial L(y, f(x_i))/\partial f(x_i)]_{f(x)=f_{m-1}(x)}$
(b) Train a regression tree in order to fit the computed residual (r_{mi}) and then obtain the leaf node area regarding m_{th} tree R_{mj}, $j = 1, 2, 3, .., J$;
(c) For $j = 1$ To J compute$C_{mj} = \text{argmin}_c \sum_{xi \in R_{mj}} L(y_i, f_{m-1}(x_i) + c)$
(d) Update the current model.$f_m(x) = f_{m-1}(x) + \sum_{j=1}^{J} c_{mj} I(x \in R_{mj})$
(iii) Obtain the final additive model $\widehat{f}(x) = f_M(x) = \sum_{m=1}^{M} \sum_{j=1}^{J} c_{mj} I(x \in R_{mj})$

Algorithm 1: Gradient boosting machine model.

4. Results and Discussion

In this study, four machine learning modeling approaches have been developed to select the best model for predicting monthly evaporation. The four models (RF, ELM, GBM, and GPR) are trained and validated using climate data collected from two different locations in Iraq. About seventy percent of available data were used for calibration and the other thirty percentage used for validating the predictive models. The used models in this study have been assessed by different statistical criteria as well as graphical presentations.

For the case study, the performances of the applied models through the training phase are summarized in Table 1. The given statistics showed that all the models provided a good similarity between predicted evaporation and predicted values except GPR ($R = O.938$, and Md = 0.967). Furthermore, it can be observed that the GBM generated fewer error forecasts compared to other models (MAE = 14.170, RMSE = 23.092, MAPE = 0.095, $R = 0.987$, and Md = 0.993). However, the performances of the ELM and RF models were very similar. However, it can be said that there was a slight advantage in favor of the ELM model. This model provided smaller values of MAE and MAPE compared to the ORF model. Table 2 provides a significant analysis of the models' performances through the training phase for the second case study. Based on the obtained results, the GBM model showed an excellent ability to predict the monthly evaporation, providing lowest estimated errors (MAE = 13.645, RMSE = 20.509, and MAPE = 0.058) and highest prediction accuracy ($R = 0.994$, WI = 9.997). The second and third best models were ELM and QRF, respectively. However, the GPR was considered the worst predicted model because it gave the highest values of RMSE, MAPE, and MAE. It can be concluded that, through the training stage, the GPR was noticed to have a poor accuracy of both case studies. However, the GBM model has a robust

Table 1: The evaluations of the predictive models through training phase: first case study in Diyala state.

Model	MAE	RMSE	MAPE	R	Md
ELM	26.803	44.588	0.170	0.951	0.973
GPR	32.157	49.524	0.231	0.938	0.967
QRF	27.095	43.214	0.186	0.953	0.975
GBM	14.170	23.092	0.095	0.987	0.993

Table 2: The evaluations of the predictive models through training phase: second case study in Erbil state.

Model	MAE	RMSE	MAPE	R	Md
ELM	31.595	45.222	0.123	0.968	0.983
GPR	37.719	49.512	0.184	0.961	0.980
QRF	26.690	39.933	0.103	0.975	0.987
GBM	13.645	20.509	0.058	0.994	0.997

performance in the simulation of the evaporation rate for both case studies according to the obtained statistical parameters.

To assess the prediction accuracy of the applied models for the two case studies, boxplot diagrams were established to visually show the similarity of the prediction values with the observed evaporation rates. The performances of the four applied models to predict the monthly evaporation rate for both cases studies are graphically illustrated in Figures 5 and 6, respectively. The clearest observation that can be reported was the inability of the GPR model to generate an acceptable accuracy of evaporation estimations. Moreover, this model could not provide a satisfactory prediction especially for higher and lower values of evaporation. However, both figures illustrated that the GBM was superior because the calculated median for that model was very close to the actual value. Additionally, it successfully managed to simulate the higher and lower values of evaporation compared to other models.

Although success has been attained in the monthly evaporation using the GBM model during the training phase, it is very essential to evaluate the proposed model with testing dataset. As is well known, the training results may provide misleading assessment because the model is trained using known input and third corresponding targets [65]. Besides, the testing phase is very crucial in assessing the quality of the predictive models and, hence, the models' abilities would be assessed very well in terms of generalization and avoiding overfitting [66].

The assessment process of the applied models through the testing phase for the first case study that was in Diyala state is exhibited in Table 3. The superiority of the GBM model in estimating the monthly evaporation compared to other models has been easily noted in the table. More specifically, the GBM model was found to produce a satisfactory estimate with RMSE of 28.478, MAE of 21.541, MAPE of 0.181, R of 0.976, and Md of 0.987. However, the QRF provided the worst prediction accuracy compared to the applied models. With respect to case study 2 which was in Erbil state, the performance of the GBM according

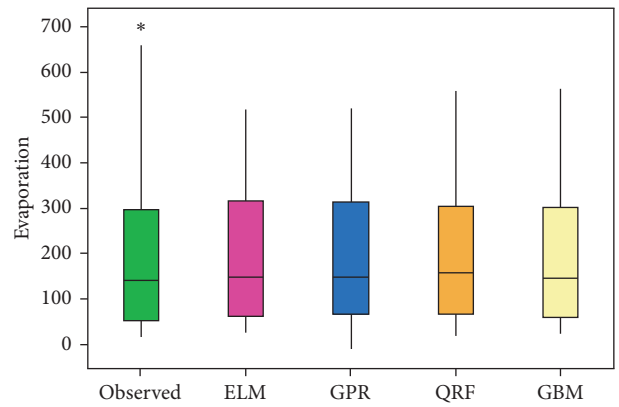


Figure 5: boxplot showing similarity between predictive and observed values for the first case study in Diyala state through training phase.

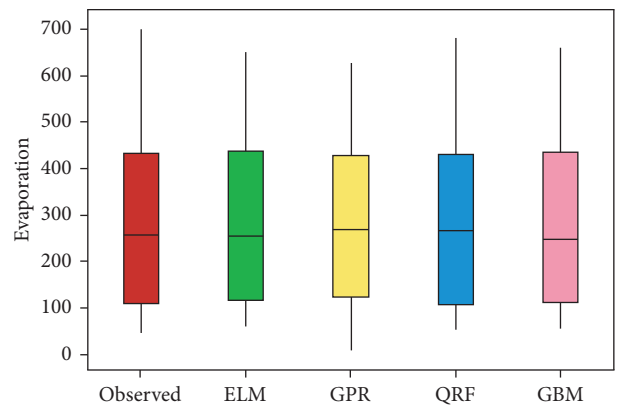


Figure 6: boxplot showing similarity between predictive and observed values for the second case study in Erbil state through training phase.

to Table 4 was also superior and provided fewer estimated errors (MAE = 26.368, RMSE = 35.345, and MAPE = 0.130) as well as higher values of R (0.985) and Md (0.989).

The reported results for both case studies showed that the GBM significantly outperformed the other machine learning models. The superiority of this model can be measured based on its capacity for reducing the MAE and RMSE for both stations during the testing phase (see Figure 7). The results showed for the case first case study a prediction enhancement in terms of MAE and RMSE by 7.17%, 21.01%; 16.51%, 15.74%; and 23.14%, 26.64%; during using GBM compared to ELM, GPR, and QRF, respectively. However, for the second case study in Erbil state, the prediction enhancement was improved in terms of reduction of MAE and RMSE by 10.88%, 9.24%; 15.24%, 5%; and 16.06%, 15.76%; respectively, compared to ELM, GPR, and QRF models.

The visualization assessment presented in Figures 8 and 9 proved that the estimated monthly evaporation rates for both stations by GBM through the testing phase were very close to the observed values. Moreover, the statistical parameters such as median and highest and lowest values

Table 3: The evaluations of the predictive models through testing phase: first case study in Diyala state.

Model	MAE	RMSE	MAPE	R	Md
ELM	23.204	36.053	0.171	0.963	0.978
GPR	25.801	33.799	0.216	0.966	0.982
QRF	28.026	38.818	0.200	0.959	0.974
GBM	21.541	28.478	0.181	0.976	0.987

Table 4: The evaluations of the predictive models through testing phase: second case study in Erbil state.

Model	MAE	RMSE	MAPE	R	Md
ELM	29.588	38.943	0.151	0.982	0.986
GPR	31.108	37.205	0.172	0.982	0.987
QRF	31.412	41.959	0.144	0.975	0.984
GBM	26.368	35.345	0.130	0.985	0.989

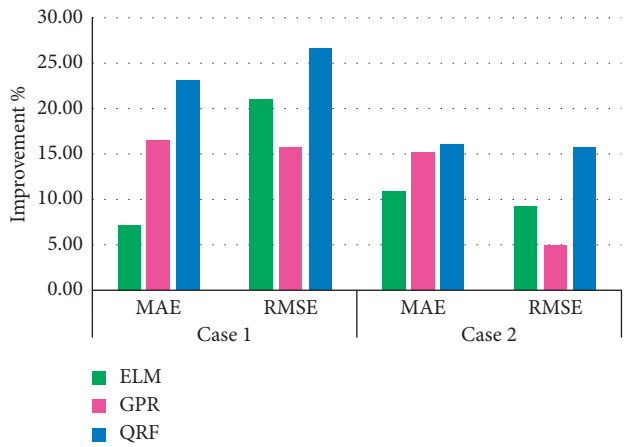


Figure 7: The superiority of the GBM over the ELM, GPM, and QRF in reducing the values of statistical parameters. Case 1 is in Diyala state, while case 2 is in Erbil state.

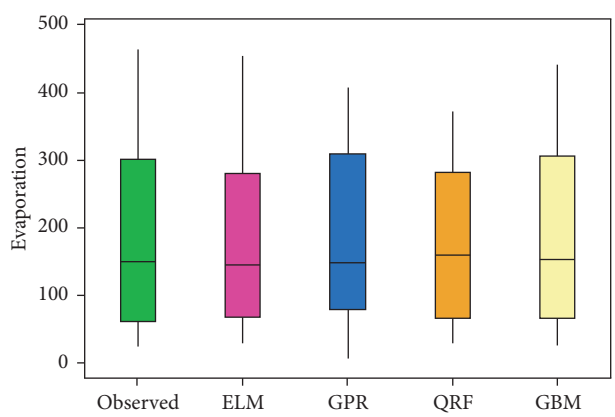


Figure 8: The boxplot showing similarity between predictive and observed values for the first case study in Diyala state through testing phase.

were noticed to be very similar to the actual values. However, these figures showed that the GPR model had a poor performance in both case studies compared to other models.

For further assessment, Taylor diagrams were created using the prediction values obtained from four models for both stations (see Figures 10 and 11). The advantage of

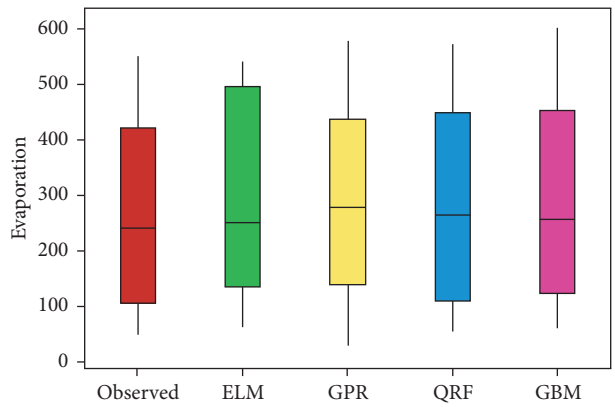


Figure 9: boxplot showing similarity between predictive and observed values for the second case study in Erbil state through testing phase.

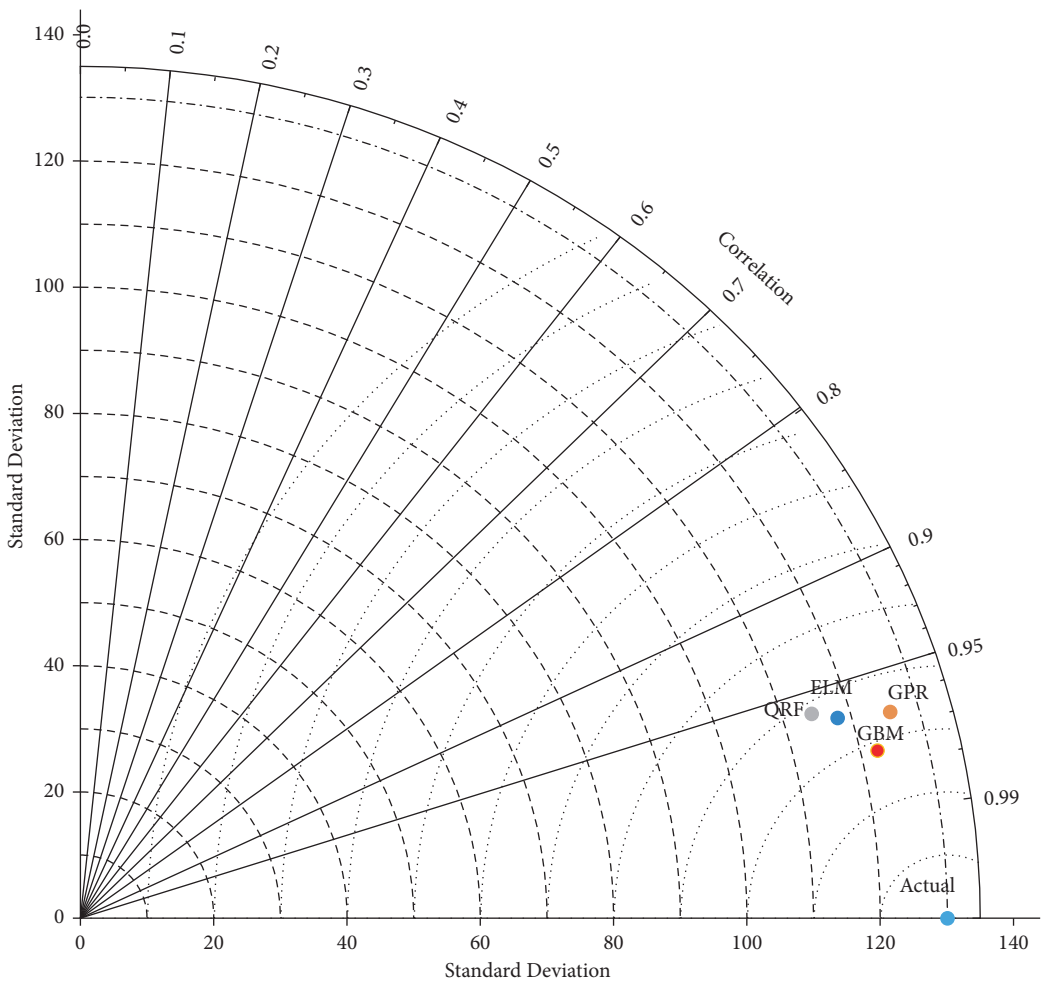


Figure 10: Taylor diagram was created to illustrate the similarity between observed and predicted values during the testing phase: case study 1 in Diyala state.

using Taylor diagram is to assess the comparable models with the actual data using three statistical parameters (standard deviation, root mean square error, and correlation coefficient). Besides, the equivalent evaporation rates obtained from each model and the actual values were assigned on a polar diagram. It can be seen from

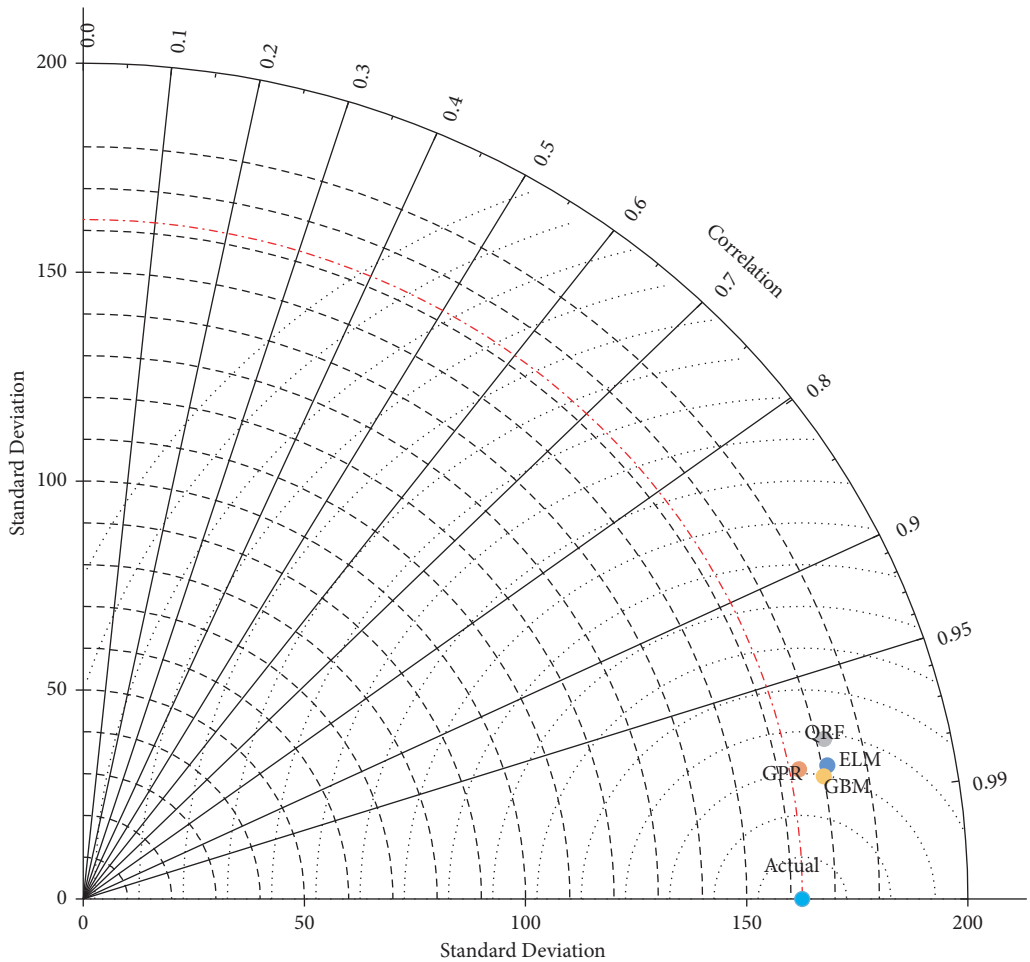


Figure 11: Taylor diagram was created to illustrate the similarity between observed and predicted values during the testing phase: case study 2 in Erbil state.

figures related to both stations that the location of the GBM model was closer to the actual values than other comparable models.

5. Conclusions

As the evaporation rate is a significant element in the hydrological cycle, its process in nature is very complicated and stochastic. In this paper, the capability of artificial intelligence models such as ELM, QRF, GBM, and GPR has been evaluated in the prediction of monthly evaporation over two stations located in Diyala and Erbil states, Iraq. The input parameters include metrological data such as sunshine hours, minimum and maximum temperature, wind speed, and relative humidity. The models were assessed using different statistical criteria as well as graphical plots. The findings of this study revealed that the GBM modeling approach has an excellent performance in the prediction of the monthly rate of evaporation over two stations with minimum forecasting errors. However, the QRF models showed the poorest performance compared with other applied models. All in all, the achieved results proved that the suggested predictive model (GBM) showed an optimistic technique for these regions; thus, it may assist local stakeholders in the management of water resources.

6. Recommendations

The recommendations for future research can be illustrated as follows:

(i) This study recommends the use of the adopted model GBM to estimate the monthly evaporation rates and investigate over several stations located in the middle and southern parts of Iraq. This study showed that the GBM model showed a good prediction accuracy in areas located in the eastern and northern parts of Iraq. Thus, it is very important to investigate the ability of this model in estimating evaporation in another regions.

(ii) The application of feature selection tool is very important to choose the most proper input variables, thus reducing the model complexity [13,67].

(iii) The GBM model is incorporated with novel bio-inspired algorithms for enhancing its performance prediction, thereby producing much accurate predictions [68–70].

Data Availability

The data are available upon request from the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] G. Konapala, A. K. Mishra, Y. Wada, and M. E. Mann, "Climate change will affect global water availability through compounding changes in seasonal precipitation and evaporation," *Nature Communications*, vol. 11, no. 3044, pp. 1–10, 2020.

[2] D. J. Short Gianotti, R. Akbar, A. F. Feldman, G. D. Salvucci, and D. Entekhabi, "Terrestrial evaporation and moisture drainage in a warmer climate," *Geophysical Research Letters*, vol. 47, no. 5, Article ID e2019GL086498, 2020.

[3] J. Shiri, "Evaluation of a neuro-fuzzy technique in estimating pan evaporation values in low-altitude locations," *Meteorological Applications*, 2018.

[4] S. Wang, J. Lian, Y. Peng, B. Hu, and H. Chen, "Generalized reference evapotranspiration models with limited climatic data based on random forest and gene expression programming in Guangxi, China," *Agricultural Water Management*, vol. 221, pp. 220–230, 2019.

[5] W. Ma, "Research progresses of flash evaporation in aerospace applications," *International Journal of Aerospace Engineering*, vol. 2018, Article ID 3686802, 15 pages, 2018.

[6] S. M. Vicente-Serrano, M. Bidegain, M. Tomas-Burguera et al., "A comparison of temporal variability of observed and model-based pan evaporation over Uruguay (1973-2014)," *International Journal of Climatology*, vol. 38, no. 1, pp. 337–350, 2018.

[7] J. Fan, B. Chen, L. Wu, F. Zhang, X. Lu, and Y. Xiang, "Evaluation and development of temperature-based empirical models for estimating daily global solar radiation in humid regions," *Energy*, vol. 144, pp. 903–914, 2018.

[8] W. Jing, Z. M. Yaseen, S. Shahid et al., "Implementation of evolutionary computing models for reference evapotranspiration modeling: short review, assessment and possible future research directions," *Engineering Applications of Computational Fluid Mechanics*, vol. 13, no. 1, pp. 811–823, 2019.

[9] Y. Yang, H. Su, and J. Qi, "A critical evaluation of the nonparametric approach to estimate terrestrial evaporation," *Advances in Meteorology*, vol. 2016, Article ID 5343718, 10 pages, 2016.

[10] O. Kisi, I. Mansouri, and J. W. Hu, "A new method for evaporation modeling: dynamic evolving neural-fuzzy inference system," *Advances in Meteorology*, vol. 2017, Article ID 5356324, 9 pages, 2017.

[11] C. M. Burt, A. J. Mutziger, R. G. Allen, and T. A. Howell, "Evaporation research: review and interpretation," *Journal of Irrigation and Drainage Engineering*, vol. 131, no. 1, pp. 37–58, 2005.

[12] Ö. Kişi, "Daily pan evaporation modelling using multi-layer perceptrons and radial basis neural networks," *Hydrological Processes*, vol. 23, no. 2, pp. 213–223, 2009.

[13] A. Malik, A. Kumar, S. Kim et al., "Modeling monthly pan evaporation process over the Indian central Himalayas: application of multiple learning artificial intelligence model," *Engineering Applications of Computational Fluid Mechanics*, vol. 14, no. 1, pp. 323–338, 2020.

[14] A. Malik, "Pan evaporation estimation in Uttarakhand and Uttar Pradesh States, India: Validity of an integrative data intelligence model," *Atmosphere*, 2020.

[15] H. Wang, H. Yan, W. Zeng, G. Lei, C. Ao, and Y. Zha, "A novel nonlinear Arps decline model with salp swarm algorithm for predicting pan evaporation in the arid and semi-arid regions of China," *Journal of Hydrology*, vol. 582, Article ID 124545, 2020.

[16] Z. M. Yaseen, A. M. Al-Juboori, U. Beyaztas et al., "Prediction of evaporation in arid and semi-arid regions: a comparative study using different machine learning models," *Engineering Applications of Computational Fluid Mechanics*, vol. 14, no. 1, pp. 70–89, 2019.

[17] R. Arunkumar and V. Jothiprakash, "Reservoir evaporation prediction using data-driven techniques," *Journal of Hydrologic Engineering*, vol. 18, no. 1, pp. 40–49, 2013.

[18] V. Nourani, M. Sayyah-Fard, M. T. Alami, and E. Sharghi, "Data pre-processing effect on ANN-based prediction intervals construction of the evaporation process at different climate regions in Iran," *Journal of Hydrology*, vol. 588, Article ID 125078, 2020.

[19] J. Tanny, "Evaporation from a small water reservoir: direct measurements and estimates," *Journal of Hydrology*, vol. 351, no. 1–2, pp. 218–229, 2008.

[20] L. Wang, O. Kisi, M. Zounemat-Kermani, and H. Li, "Pan evaporation modeling using six different heuristic computing methods in different climates of China," *Journal of Hydrology*, vol. 544, pp. 407–427, 2017.

[21] A. Guven and Ö. Kişi, "Daily pan evaporation modeling using linear genetic programming technique," *Irrigation Science*, vol. 29, no. 2, pp. 135–145, 2011.

[22] M. Allen, L. S. Pereira, D. Raes, and Smith, "Crop evaporation–Guidelines for computing crop water requirements–FAO Irrigation and drainage paper 56," *Food Agric. Organ. United Nations*, vol. 300, no. 9, p. 300, 1998.

[23] M. M. Hameed, M. K. Alomar, S. F. Mohd Razali et al., "Application of artificial intelligence models for evapotranspiration prediction along the southern coast of Turkey," *Complexity*, vol. 2021, Article ID 8850243, 20 pages, 2021.

[24] M. A. Ghorbani, R. C. Deo, S. Kim, M. Hasanpour Kashani, V. Karimi, and M. Izadkhah, "Development and evaluation of the cascade correlation neural network and the random forest models for river stage and river flow prediction in Australia," *Soft Computing*, vol. 24, no. 16, pp. 12079–12090, 2020.

[25] A. Ashrafzadeh, A. Malik, V. Jothiprakash, M. A. Ghorbani, and S. M. Biazar, "Estimation of daily pan evaporation using neural networks and meta-heuristic approaches," *ISH Journal of Hydraulic Engineering*, pp. 1–9, 2018.

[26] A. Elbeltagi, N. Azad, A. Arshad et al., "Applications of Gaussian process regression for predicting blue water footprint: case study in Ad Daqahliyah, Egypt," *Agricultural Water Management*, vol. 255, Article ID 107052, 2021.

[27] G. T. Patle, M. Chettri, and D. Jhajharia, "Monthly pan evaporation modelling using multiple linear regression and artificial neural network techniques," *Water Supply*, vol. 20, no. 3, pp. 800–808, 2019.

[28] A. S. Ponraj and T. Vigneswaran, "Daily evapotranspiration prediction using gradient boost regression model for irrigation planning," *The Journal of Supercomputing*, vol. 76, no. 8, pp. 5732–5744, 2019.

[29] A. Sharafati, R. Yasa, and H. M. Azamathulla, "Assessment of stochastic approaches in prediction of wave-induced pipeline scour depth," *Journal of Pipeline Systems Engineering and Practice*, vol. 9, no. 4, 2018.

[30] M. K. Goyal, B. Bharti, J. Quilty, J. Adamowski, and A. Pandey, "Modeling of daily pan evaporation in sub tropical climates using ANN, LS-SVR, Fuzzy Logic, and ANFIS," *Expert Systems with Applications*, vol. 41, no. 11, pp. 5267–5276, 2014.

[31] G. Tezel and M. Buyukyildiz, "Monthly evaporation forecasting using artificial neural networks and support vector machines," *Theoretical and Applied Climatology*, 2013.

[32] S. N. Qasem, S. Samadianfard, S. Kheshtgar et al., "Modeling monthly pan evaporation using wavelet support vector regression and wavelet artificial neural networks in arid and humid climates," *Engineering Applications of Computational Fluid Mechanics*, vol. 13, no. 1, pp. 177–187, 2019.

[33] A. Rakhee and A. Kumar, "Predictive modeling of pan evaporation using random forest algorithm along with features selection," in *Proceedings of the Confluence 2020 - 10th International Conference on Cloud Computing, Data Science and Engineering*, pp. 380–384, Uttarpradesh, India, January 2020.

[34] D. Althoff, R. Filgueiras, and L. N. Rodrigues, "Estimating small reservoir evaporation using machine learning models for the Brazilian savannah," *Journal of Hydrologic Engineering*, vol. 25, no. 8, Article ID 05020019, 2020.

[35] L. Qian, "A study of the conversion of different evaporation pans in South China based on the extreme learning machine model," *Hydrological Sciences Journal*, 2021.

[36] L. Dong, W. Zeng, L. Wu et al., "Estimating the Pan evaporation in northwest China by coupling CatBoost with bat algorithm," *Water*, vol. 13, no. 3, p. 256, 2021.

[37] M. Al-Mukhtar, "Modeling the monthly pan evaporation rates using artificial intelligence methods: a case study in Iraq," *Environmental Earth Sciences*, vol. 80, no. 1, 2021.

[38] A. Ghaemi, M. Rezaie-Balf, J. Adamowski, O. Kisi, and J. Quilty, "On the applicability of maximum overlap discrete wavelet transform integrated with MARS and M5 model tree for monthly pan evaporation prediction," *Agricultural and Forest Meteorology*, vol. 278, p. 107647, 2019.

[39] J. Chenoweth, "Impact of climate change on the water resources of the eastern Mediterranean and Middle East region: Modeled 21st century changes and implications," *Water Resource Research*, 2011.

[40] N. Al-Ansari, A. A. Ali, and S. Knutsson, "Present conditions and future challenges of water resources problems in Iraq," *Journal of Water Resource and Protection*, vol. 6, no. September, pp. 1066–1098, 2014.

[41] J. Lelieveld, "Climate change and impacts in the Eastern Mediterranean and the Middle East," *Climate Change*, 2012.

[42] O. T. Al-Taai and S. H. Hadi, "Analysis of the monthly and annual change of soil moisture and evaporation in Iraq," *Al-Mustansiriyah J. Sci.*vol. 29, no. 4, pp. 7–13, 2018.

[43] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, The MIT Press, Massachusets, MA, USA, 2005.

[44] R. C. Deo and P. Samui, "Forecasting evaporative loss by least-square support-vector regression and evaluation with genetic programming, Gaussian process, and minimax probability machine regression: case study of brisbane city," *Journal of Hydrologic Engineering*, vol. 22, no. 6, p. 05017003, 2017.

[45] M. Akbari, F. Salmasi, H. Arvanaghi, M. Karbasi, and D. Farsadizadeh, "Application of Gaussian process regression model to predict discharge coefficient of Gated Piano Key Weir," *Water Resources Management*, vol. 33, no. 11, pp. 3929–3947, 2019.

[46] P. Sihag, P. Jain, and M. Kumar, "Modelling of impact of water quality on recharging rate of storm water filter system using various kernel function based regression," *Modeling Earth Systems and Environment*, 2018.

[47] K. Roushangar and S. Shahnazi, "Prediction of sediment transport rates in gravel-bed rivers using Gaussian process regression," *Journal of Hydroinformatics*, 2019.

[48] B. T. Pham, H.-B. Ly, N. Al-Ansari, and L. S. Ho, "A comparison of Gaussian process and M5P for prediction of soil permeability coefficient," *Scientific Programming*, vol. 2021, Article ID 3625289, 13 pages, 2021.

[49] P. Koepernik and F. Pfaff, "Consistency of Gaussian process regression in metric spaces," *Journal of Machine Learning Research*, vol. 22, no. 244, pp. 1–27, 2021.

[50] Y. Wang and B. Chaib-draa, "An online Bayesian filtering framework for Gaussian process regression: application to global surface temperature analysis," *Expert Systems with Applications*, vol. 67, pp. 285–295, 2017.

[51] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1–3, pp. 489–501, 2006.

[52] G. Huang, G.-B. Huang, S. Song, and K. You, "Trends in extreme learning machines: a review," *Neural Networks*, vol. 61, pp. 32–48, 2015.

[53] A. M. Araba, Z. A. Memon, M. Alhawat, M. Ali, and A. Milad, "Estimation at completion in civil engineering projects: review of regression and soft computing models," *Knowledge-Based Engineering and Sciences*, vol. 2, no. 2, pp. 1–12, 2021.

[54] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[55] S. A. Gyamerah, P. Ngare, and D. Ikpe, "Probabilistic forecasting of crop yields via quantile random forest and Epanechnikov Kernel function," *Agricultural and Forest Meteorology*, vol. 280, Article ID 107808, 2020.

[56] N. Meinshausen, "Quantile regression forests," *Journal of Machine Learning Research*, vol. 7, no. Jun, pp. 983–999, 2006.

[57] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of Statistics*, 2001.

[58] J. Zhou, E. Li, S. Yang et al., "Slope stability prediction for circular mode failure using gradient boosting machine approach based on an updated database of case histories," *Safety Science*, vol. 118, pp. 505–518, 2019.

[59] Z. Zhou, L. Zhao, A. Lin et al., "Exploring the potential of deep factorization machine and various gradient boosting models in modeling daily reference evapotranspiration in China," *Arabian Journal of Geosciences*, vol. 13, no. 24, p. 1287, 2021.

[60] M. Xenochristou, C. Hutton, J. Hofman, and Z. Kapelan, "Water demand forecasting accuracy and influencing factors at different spatial scales using a gradient boosting machine," *Water Resources Research*, vol. 56, no. 8, p. e2019WR026304, 2020.

[61] S. R. Naganna, B. H. Beyaztas, N. Bokde, and A. M. Armanuos, "On the evaluation of the gradient tree boosting model for groundwater level forecasting," *Knowledge-Based Engineering and Science*, vol. 1, no. 1, pp. 48–57, 2020.

[62] A. Sharafati, S. B. H. S. Asadollah, and A. Neshat, "A new artificial intelligence strategy for predicting the groundwater level over the Rafsanjan aquifer in Iran," *Journal of Hydrology*, 2020.

[63] M. Gong, Y. Bai, J. Qin, J. Wang, P. Yang, and S. Wang, "Gradient boosting machine for predicting return temperature of district heating system: a case study for residential buildings in Tianjin," *Journal of Building Engineering*, vol. 27, Article ID 100950, 2020.

[64] Z. M. Yaseen, "An insight into machine learning models era in simulating soil, water bodies and adsorption heavy metals:

review, challenges and solutions," *Chemosphere*, vol. 277, Article ID 130126, 2021.

[65] H. Tao, "Training and testing data division influence on hybrid machine learning model process: application of river flow forecasting," *Complexity*, 2020.

[66] M. M. Hameed, M. K. AlOmar, W. J. Baniya, and M. A. AlSaadi, "Incorporation of artificial neural network with principal component analysis and cross-validation technique to predict high-performance concrete compressive strength," *Asian Journal of Civil Engineering*, vol. 22, no. 6, pp. 1019–1031, 2021.

[67] A. Malik, A. Kumar, and O. Kisi, "Daily Pan evaporation estimation using heuristic methods with gamma test," *Journal of Irrigation and Drainage Engineering*, vol. 144, no. 9, p. 04018023, 2018.

[68] H. Tao, A. A. Ewees, A. O. Al-Sulttani et al., "Global solar radiation prediction over North Dakota using air temperature: development of novel hybrid intelligence model," *Energy Reports*, vol. 7, pp. 136–157, 2021.

[69] N. Arya Azar, N. Kardan, and S. Ghordoyee Milan, "Developing the artificial neural network–evolutionary algorithms hybrid models (ANN–EA) to predict the daily evaporation from dam reservoirs," *Engineering Computation*, 2021.

[70] A. Malik, "Daily pan-evaporation estimation in different agro-climatic zones using novel hybrid support vector regression optimized by Salp swarm algorithm in conjunction with gamma test," *Engineering Applications of Computational Fluid Mechanics*, vol. 15, no. 1, pp. 1075–1094, 2021.

Hindawi
Advances in Meteorology
Volume 2022, Article ID 1860460, 21 pages
https://doi.org/10.1155/2022/1860460

Research Article

Short-Term Daily Univariate Streamflow Forecasting Using Deep Learning Models

Eyob Betru Wegayehu and Fiseha Behulu Muluneh

School of Civil and Environmental Engineering, Addis Ababa Institute of Technology, Addis Ababa University, Addis Ababa, Ethiopia

Correspondence should be addressed to Eyob Betru Wegayehu; eyob.betru@aait.edu.et

Received 9 December 2021; Revised 25 January 2022; Accepted 27 January 2022; Published 10 February 2022

Academic Editor: Upaka Rathnayake

Hydrological forecasting is one of the key research areas in hydrology. Innovative forecasting tools will reform water resources management systems, flood early warning mechanisms, and agricultural and hydropower management schemes. Hence, in this study, we compared Stacked Long Short-Term Memory (S-LSTM), Bidirectional Long Short-Term Memory (Bi-LSTM), and Gated Recurrent Unit (GRU) with the classical Multilayer Perceptron (MLP) network for one-step daily streamflow forecasting. The analysis used daily time series data collected from Borkena (in Awash river basin) and Gummera (in Abay river basin) streamflow stations. All data sets passed through rigorous quality control processes, and null values were filled using linear interpolation. A partial autocorrelation was also applied to select the appropriate time lag for input series generation. Then, the data is split into training and testing datasets using a ratio of 80 : 20, respectively. Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and coefficient of determination (R^2) were used to evaluate the performance of the proposed models. Finally, the findings are summarized in model variability, lag time variability, and time series characteristic themes. As a result, time series characteristics (climatic variability) had a more significant impact on streamflow forecasting performance than input lagged time steps and deep learning model architecture variations. Thus, Borkena's river catchment forecasting result is more accurate than Gummera's catchment forecasting result, with RMSE, MAE, MAPE, and R^2 values ranging between (0.81 to 1.53, 0.29 to 0.96, 0.16 to 1.72, 0.96 to 0.99) and (17.43 to 17.99, 7.76 to 10.54, 0.16 to 1.03, 0.89 to 0.90) for both catchments, respectively. Although the performance is dependent on lag time variations, MLP and GRU outperform S-LSTM and Bi-LSTM on a nearly equal basis.

1. Introduction

The science of streamflow forecasting is still one of the crucial research topics in hydrology. Accurate and reliable streamflow forecasting is vital for water resources planning, management, and disaster mitigation response authorities [1]. Streamflow forecasting can be classified into two. Firstly, short-term or real-time forecasting includes daily and hourly timestamps, widely applicable in flood management systems. Secondly, long-term forecasting usually contains weekly, monthly, and annual streamflow forecasting, crucial for reservoir operation, irrigation system management, and hydropower generation [2].

Generally, streamflow forecasting models can also be categorized into conceptual, physical-based, and data-driven models [3]. Conceptual models are lumped in nature and typically rely on empirical relationships among various hydrological variables. Due to the model's reliance on observed data, it is rarely applicable to data-scarce catchments. Hydrological processes can also be represented in physical models through mass, momentum, and energy conservation equations. These models may account for spatial variability, but since they are distributed in nature, they require a large amount of data on land use, slope, soil, and climate [4]. Lastly, data-driven models form a nonlinear input-output relationship without physical catchment information and

minimum data requirements. Hence, the popularity of data-driven models is exploding these days with the advancement of computational capability and data set availability [5].

Zhang et al. [6] classified the data-driven approach as conventional and Artificial Intelligence (AI) based models. Conventional data-driven models such as Auto Regressive Moving Average with the exogenous term (ARMAX), Multiple Linear Regressive (MLR), and Auto-Regressive Integrated Moving Average (ARIMA) are easy to implement [6]. However, the nonlinearity of hydrological processes was left out of the equations. On the other hand, AI-based data-driven models can detect nonlinearity and perform better in streamflow forecasting. As a result, machine learning models have become a hot research topic these days.

AI-based data-driven streamflow forecasting models can be univariate when the model's input and output are designed with a single time series variable. Univariate forecasting models are straightforward to train using sparse data and provide ease of inference when evaluating forecast performance. Due to the complexity of agrometeorological data, it is simpler and more efficient to forecast the variables individually [7]. On the other hand, multivariate models are designed with multiple variables such as precipitation, temperature, evaporation, and other variables as input and a streamflow variable as output [6]. Thus, in data-scarce catchments with a limited amount of data, the application of univariate modelling is more feasible and has received wide attention in recent years [3, 6, 8, 9].

Wide variates of classical and deep learning models are present in the literature for time series forecasting, which includes Artificial Neural Network (ANN), Support Vector Machine (SVM), Fuzzy Logic (FL), Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), Adaptive Neuro-Fuzzy Inference System (ANFIS), and Genetic Programming (GP). However, because of the nonlinearity present in streamflow time series, the forecasting performance of these models is usually debatable [3, 10]. Under one-step and multistep-ahead forecast scenarios, Suradhaniwar et al. [7] compared the performance of Machine Learning (ML) and Deep Learning (DL) based time series forecasting algorithms. They also evaluated recursive one-step forward forecasts using walk-forward validation. Finally, Seasonal Auto-Regressive Integrated Moving Average (SARIMA) and Support Vector Regression (SVR) models outperform their DL-based counterparts: Neural Network (NN), Recurrent Neural Network (RNN), and Long Short-Term Memory (LSTM) with fixed forecast horizons.

ANN (MLP) is the most widely used classical machine learning architecture in hydrology [11]. Cui et al. [12] demonstrated that when used for hourly river flow forecasting, the new generation of ANN or Emotional Neural Network (ENN) models outperformed the Multivariate Adaptive Regression Splines (Mars), Minimax Probability Machine Regression (MPMR), and Relevance Vector Machine (RVM) models. Yaseen et al. [2] also conducted a detailed review of literature from high impacted journals in the time frame of 2000–2015 on the state-of-the-art application of Artificial Neural Network (ANN), Support Vector Machine (SVM), Fuzzy Logic (FL), Evolutionary Computation (EC), and Wavelet-Artificial Intelligence (W-AI) for streamflow forecasting. The same study was concluded by stating that time series preprocessing, input variable selections, and time scale choices are the critical parameters for high-performing forecasting models.

RNN is the popular type of deep learning architecture that is optimized for time series analysis. However, it has drawbacks, such as vanishing and exploding gradients. Hochreiter and Schmidhuber [13] introduced the improved RNN version or LSTM, a popular time series model for longtime step forecasting. Recently, various fields of study have been experimenting with these models [14–20]. Moreover, Cho et al. [21] firstly introduced GRU as a simplified version of the LSTM model. GRU merges short-term and long-term memory cells into a single gate with reasonably good performance and fast running time [22]. Lara Benítez et al. [23] evaluated the accuracy and efficiency of seven popular deep learning architectures: Multilayer Perceptron (MLP), Elman Recurrent Neural Network (ERNN), Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), Echo State Network (ESN), Convolutional Neural Network (CNN), and Temporal Convolutional Network (TCN). Additionally, they constructed over 38000 distinct models to search for optimal architecture configuration and training hyperparameters, with LSTM achieving the lowest weighted absolute percentage error followed by GRU.

Even though we found multiple effective forecasting models with GRU and LSTM in different fields, specifically in hydrology, the accuracy of these models must be further fine-tuned with different data processing techniques and data input variations [24–27]. We can restructure the previous time step data series as a predictor input variable for univariate time series forecasting and the current or next step as the output variable. However, the number of input time steps selected is difficult to decide without prior knowledge. Hence, studying the effect of lagged time selection in streamflow forecasting is mandatory for to obtain accurate models. Lagged variables in univariate streamflow forecasting are significant factors that vary the model's performance positively or negatively and hold temporal dependency information as a predictor variable [28]. Tyralis and Papacharalampous [29] concluded that using a low number of recent predictor variables achieves higher accuracy for time series forecasting using Random Forest (RF) algorithm. It is vital to expand this finding to other popular deep learning models and to different climatic conditions.

Papacharalampous et al. [30] tested 20 one-step ahead univariate time series forecasting methods with extensive time series data. They concluded the study by suggesting that the most and least accurate approaches for one-step-ahead forecasting, such as the importance of time series length on the performance of various forecasting methods, are well addressed. Furthermore, the same study underlined that the machine learning model's optimization heavily relies on hyperparameter optimization and lagged variable selection. Torres et al. [31] also identified research gaps in various fields by analyzing the most successful deep learning architectures

that predict time series effectively and highlighting MLP, RNN, GRU, and Bi-LSTM architectures in particular.

In the present study, we compared the different forms of LSTM architectures, S-LSTM, Bi-LSTM, and GRU, with the classical MLP network to forecast a single step streamflow amount with the available records of daily streamflow data. To the best of our knowledge, the LSTM has been mainly studied for monthly multivariate time series and not for daily univariate streamflow forecasting. Even though machine learning can model hydrological forecasting efficiently, researchers should further examine the impact of suitable input variables and model parameters selection on model accuracy very carefully [32].

2. Study Area and Data

Two river basin subcatchments in Ethiopia were selected for this study: (a) Gummera subcatchment in Abay River basin (Figure 1(a)), (b) Borkena subcatchment in Awash River basin (Figure 1(b)).

2.1. Borkena Catchment (Awash River Basin/Ethiopia). Borkena River originates at Kutaber woreda or the conjunction of Abay and Awash River basins (Ethiopia). The Awash River basin is usually classified into three main catchments, Lower Awash, Middle Awash, and Upper Awash. Borkena River is found in Lower Awash with its different tributaries, including Berbera River, Arawutie River, Abasharew/Wuranie River, Abba Abdela/Desso River, Worka River, and Leyole River. The total length of this river is estimated at around 165 km. Borkena River catchment hosts major cities, including Kombolcha, Dessie, and Kemissie. The study area streamflow outlet is at Kombolch station, and the catchment covers an area of 1709 km^2, bounded from 39° 30′E to 40° 0′E to 10° 15′N to 11° 30′N. Moreover, the area's elevation varies from 1775 m to 2638 m above sea level. The rainfall pattern of this catchment is unimodal, where 84% of the rainfall is happening from July to September [33].

2.2. Gummera Catchment (Abay River Basin/Ethiopia). The second case study area is the Gummara subbasin, one of the main tributaries of Lake Tana in the Abay River basin. The lake is located in the north-western highlands at 12° 00′N and 37° 15′E and collects runoff from more than 40 rivers. The lake receives water from several major rivers, including Gilgel Abay in the south, Megech River in the north, and Ribb and Gummara in the east. Small river streams from the lake's western side drain into the lake. Gummara River originates from the Guna mountains southeast of Debre Tabor at an altitude of approximately 3250 m.a.s.l. The Gummara catchment covers a total area of about 1592 km^2. Many small intermittent and perennial rivers and springs flow into the mainstream Gummara River. The catchment's topography is undulating, ranging from 1788 m.a.s.l. to 3750 m.a.s.l.

2.3. Data. Daily streamflow time series of two hydrometrological stations were collected from Ethiopia's Ministry of Water, Irrigation, and Energy (MoWIE). Then, these data were used to forecast single-step streamflow. At Borkena station, 6575 available streamflow data series were collected over the time window of January 1, 1972, to December 31, 1989. Similarly, at Gummera station, 9496 streamflow time series values from January 1, 1981, to December 31, 2006, were collected. A total of 866 null values are examined in the time series (i.e., 658 at Borkena and 208 at Gummara). The options for filling these gaps range from simple interpolation to complex statistical methods [34]. The method chosen depends on the length and season of missing data, availability of hydrometeorological data, climatic region, and length of previous observations. The sample mean or subgroup mean can be used to fill in missing values in daily streamflow data. However, replacing missing values with sample means may cause underestimating variance and incorrect subgroup identification [35]. Instead, the linear interpolation method is quick and straightforward to use, and it may be sufficient for data with a small gap [36]. Thus, we implemented linear interpolation in this study, and since the majority of the null values were stacked at low flows, the interpolation is acceptable [37–40].

After passing these rigorous quality control processes, the raw data were then split into training and testing datasets using a ratio of 80 : 20, respectively. Accordingly, different sizes of single overlapping step sliding windows were used to rebuild the input time series into a supervised learning format. The subsets were further standardized using a standard scalar approach. Figure 2 shows the descriptive statistics and the corresponding plots of split data for both catchments.

3. Methods

This study compared three types of recurrent network architectures (GRU, Bi-LSTM, and S-LSTM) with the classical neural network architecture (MLP).

3.1. Deep Learning Models. Deep learning models are usually distinguished from nondeep machine learning models by the depth of the network layer or by the number of stacked neuron layers and designed architectures. Nondeep learning models usually do not accurately learn the advanced nonlinearity present in the input and output variables [41]. In contrast, deep learning models are widely applied in different tasks, including processing, analyzing, designing, estimating, filtering, and detection tasks [42]. The popular deep learning models applied in different fields of studies are Multilayer Perceptron (MLP), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM), Radial Basis Function Networks (RBFN), and Generative Adversarial Network (GAN) [21, 25, 43–46]. The time series analysis models used in this study were specifically chosen, and they are briefly discussed in detail in the following sections.

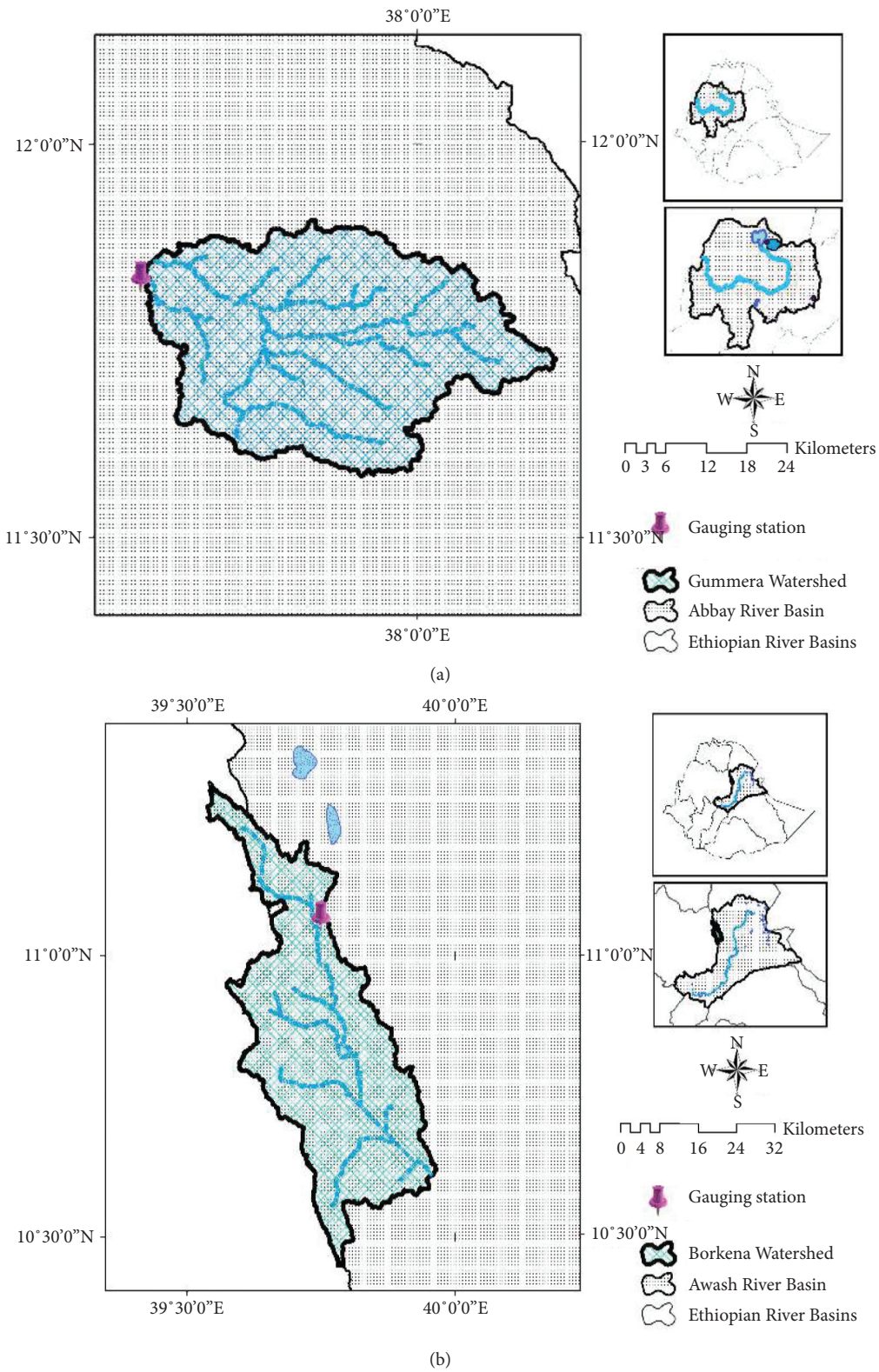
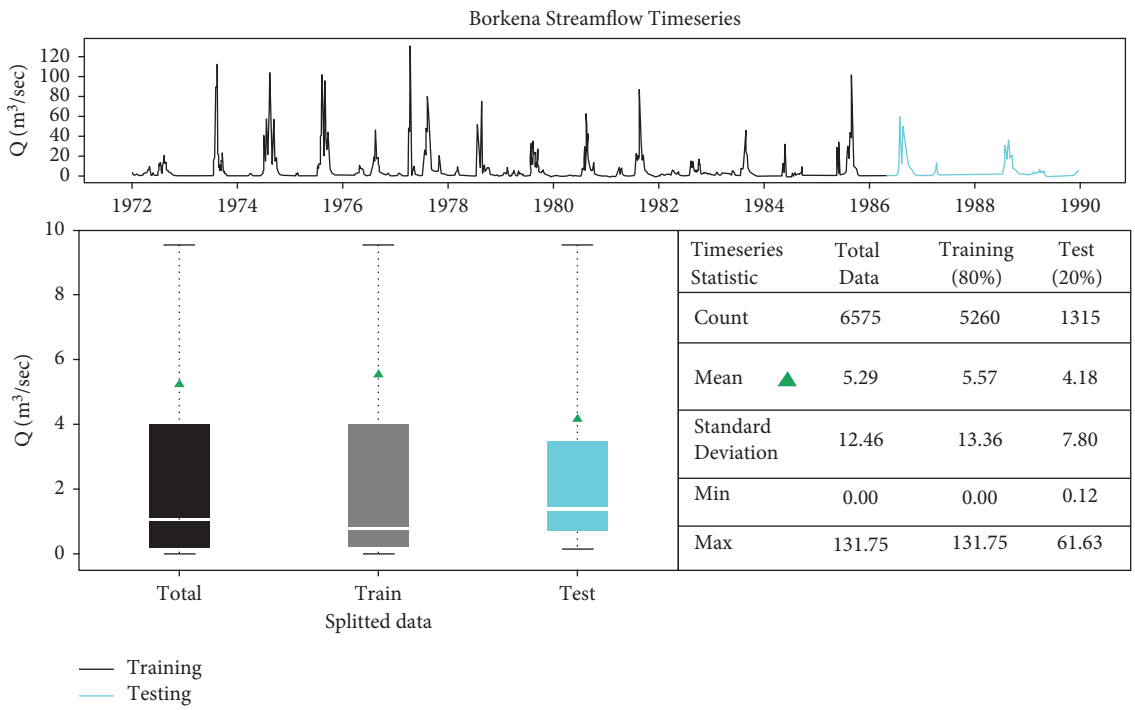


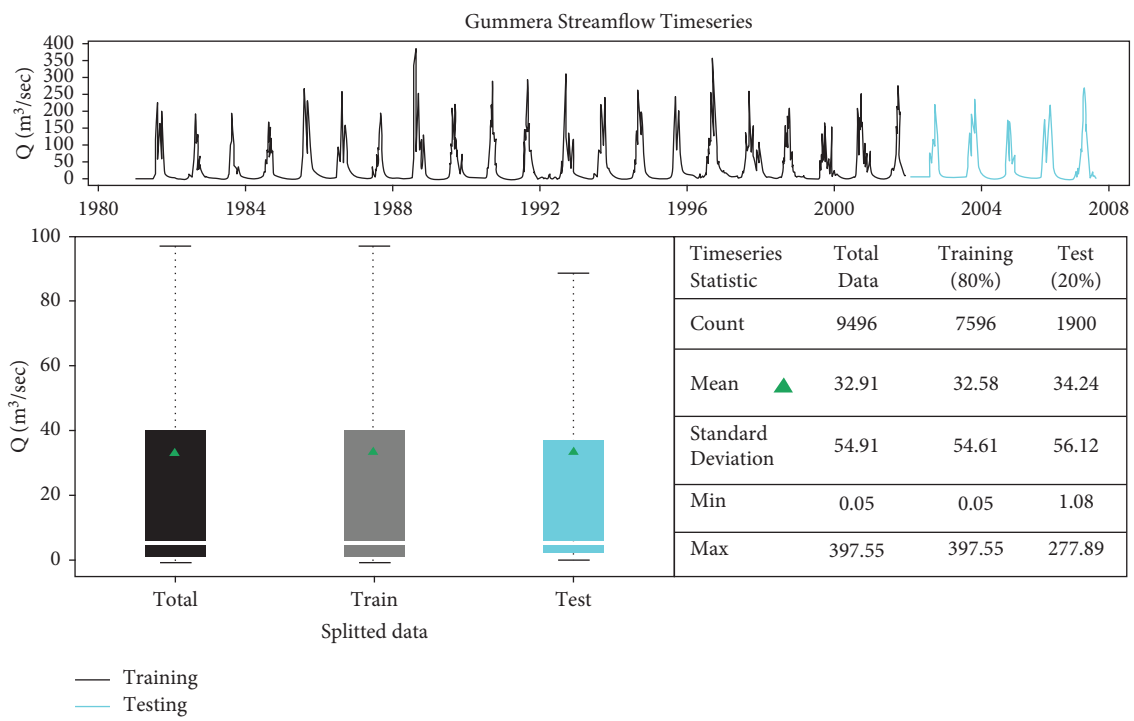
Figure 1: The location of the study areas. (a) Gummera, (b) Borkena.

3.1.1. Multilayer Perceptron (MLP). ANN or feed-forward multilayer perceptron (MLP) is an immensely used architecture in the hydrological literature [47]. Perceptron operates artificially, replicating our brain processing system and passing on different mathematical and probabilistic operations [48]. MLP contains three main layers: input layer, hidden layer, and output layer. The network becomes deep and can extract higher-order statistics by adding more



Timeseries Statistic	Total Data	Training (80%)	Test (20%)
Count	6575	5260	1315
Mean	5.29	5.57	4.18
Standard Deviation	12.46	13.36	7.80
Min	0.00	0.00	0.12
Max	131.75	131.75	61.63

(a)



Timeseries Statistic	Total Data	Training (80%)	Test (20%)
Count	9496	7596	1900
Mean	32.91	32.58	34.24
Standard Deviation	54.91	54.61	56.12
Min	0.05	0.05	1.08
Max	397.55	397.55	277.89

(b)

Figure 2: Descriptive statistics and the corresponding box plots of split data. (a) Borkena station, (b) Gummera station.

hidden layers [44]. Three-layered MLP is common in hydrological time series modelling.

A typical diagram of one node (j^{th}) ANN is displayed in Figure 3. Conditional to the layer location, the series of inputs form an input vector $X = (x_1, x_i, \ldots, x_n)$. The equivalent series of weights fit to each input form a weight vector $W_j = (w_{1j}, w_{ij}, \ldots, w_{nj})$. The output ($Y_j$) for the node j is calculated using the value of a function (f) with the inner product of the input vector (X) and weight vector (W_j) minus bias (b_j) [49]. The stated operation is displayed as follows:

$$Y_j = f\left(X \cdot W_j - b_j\right). \quad (1)$$

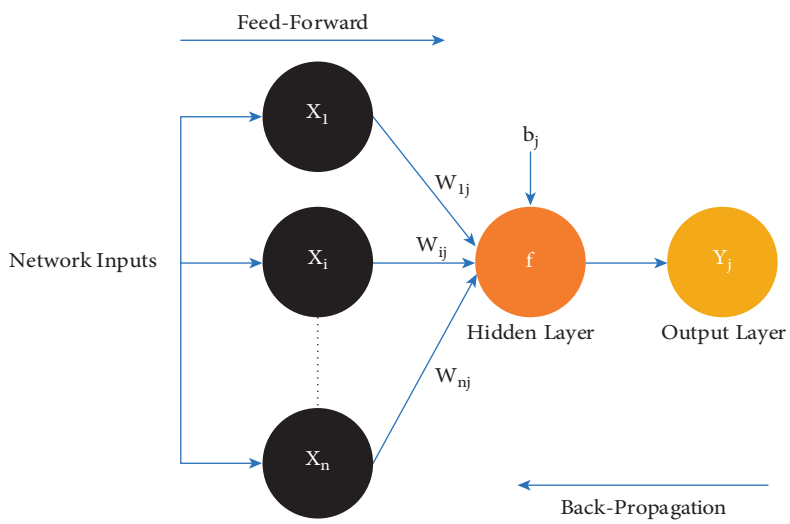


Figure 3: Schematic diagram of MLP for node j.

The activation function (f) helps decide a neuron activates a process or not, and these are the few commonly used activation functions: Rectified Linear Unit (ReLu), Leaky ReLu, Sigmoid, Hyperbolic Tangent (Tanh), and Softmax [50]. Even though ANN has been enormously applied in hydrological modelling for the past decades, its performance to capture extreme events is doubtful for complex problems such as rainfall-runoff processes [47].

3.1.2. Long Short-Term Memory (LSTM). Long Short-Term Memory (LSTM) networks are unique to MLP because the networks have recurrent connections; the information from previous long time-step memories is used to formulate forecasting. Overcoming vanishing and exploding gradients makes LSTM more popular in sequence and time series analysis than the traditional Recurrent Neural Network (RNN) [51]. LSTM networks have memory cells and three gates: input, forget, and output. These gates are responsible for the network to save, forget, pay attention, or pass the information to other cells [27]. Figure 4 displays the typical LSTM memory cell with three gated layers, and the detailed mathematical formulation for the network components is described as follows:

$$f_t = \sigma\left(u_f x_t + w_f h_{t-1} + b_f\right), \tag{2}$$

$$i_t = \sigma\left(u_i x_t + w_i h_{t-1} + b_i\right), \tag{3}$$

$$o_t = \sigma\left(u_o x_t + w_o h_{t-1} + b_o\right), \tag{4}$$

$$c_{et} = \tanh\left(w_c x_t + u_c h_{t-1} + b_c\right), \tag{5}$$

$$c_t = f_t * c_{t-1} + i_t * c_{et}, \tag{6}$$

$$h_t = \tanh\left(c_t\right) * o_t, \tag{7}$$

where f_t (equation (2)) is a forget gate that has a responsibility to choose the information to reject from the cell, i_t (equation (3)) is an input gate that can decide on the input values to update the memory state, and o_t (equation (4)) is an output gate that can decide the output value after analyzing the input and memory of the cell. The weight matrices w_i, w_f, w_o, and w_c correspond to the input gate, forget gate, output gate, and cell gate units, respectively. While u_i, u_f, u_o, and u_c weight matrices map the hidden layer output gates, b_i, b_f, b_o, and b_c are the bias vectors of the input gate, forget gate, output gate, and cell gate units, respectively. Moreover, c_t (equation (6)) and h_t (equation (7)) are a memory cell and hidden state [24].

3.1.3. Bidirectional LSTM (Bi-LSTM). Bi-LSTM is also the other option for getting the most out of RNN by stepping through the input time steps forward and backward. Although Bi-LSTMs were developed for speech recognition, the use of bidirectional input sequences is one of the principal options for sequence prediction nowadays. The hidden layer of the Bi-LSTM model needs to save two values, in which h_t involves the forward calculation and h'_t involves the backward calculation. The final output value Y_t is obtained by combining the outputs of the forward layers and backward layers' outputs [52].

Each point in the input sequence of the output layer is provided with the complete past and future contextual information. There is no information flow between the forward and backward hidden layers, ensuring that the expanded graph is acyclic [53]. Figure 5 displays the structure of bidirectional LSTM architecture.

3.1.4. Gated Recurrent Unit (GRU). GRU is the newer generation of LSTM that merges input and forget gates into the update gate. Hence, it has fewer parameters, faster running time, and debatable performance than LSTM [24, 26, 27, 40]. Update and reset gates are the two gates available in GRU. Update gate renews the current memory, which enables the memorization of valuable information; on the contrary, the reset gate clears the current memory to

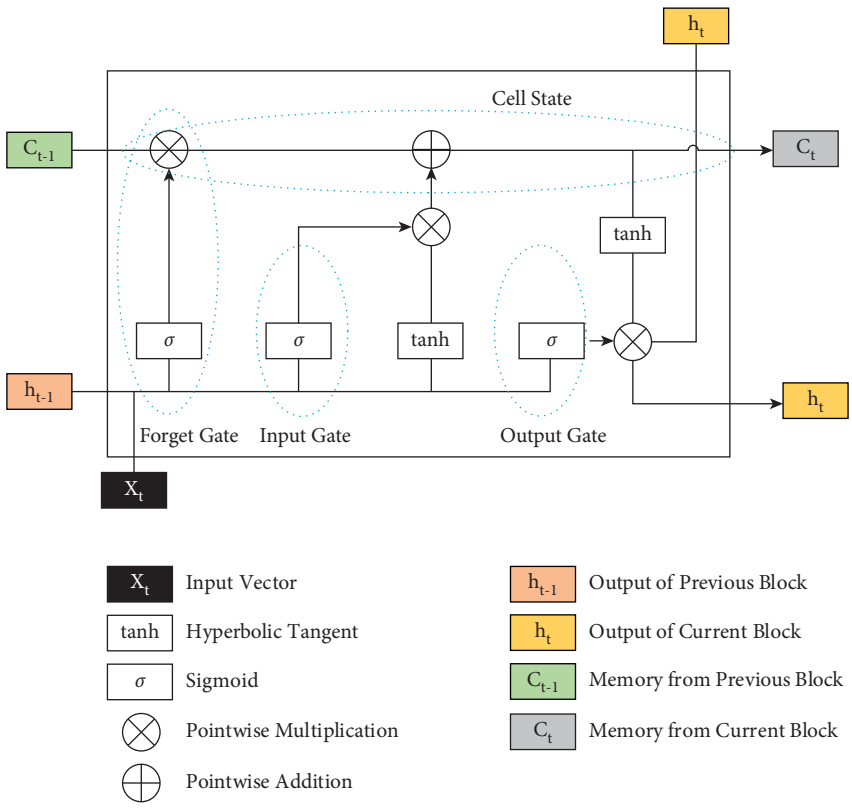


FIGURE 4: LSTM memory cell with three gated layers: forget gate f_t, input gate i_t, and output gate o_t, controlling the activation of cells c_{t-1} and c_t [38].

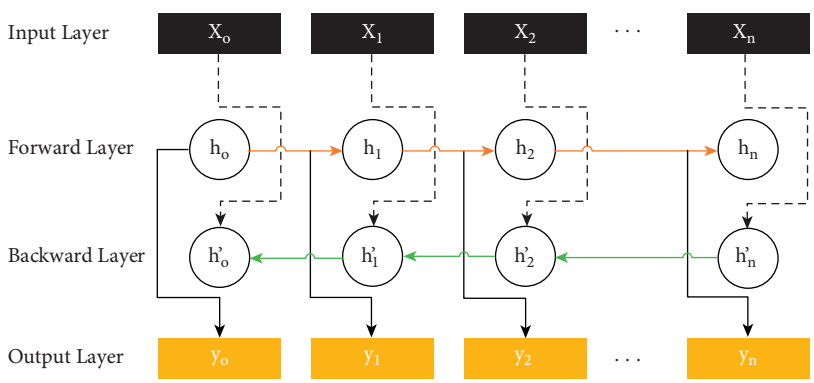


FIGURE 5: Bidirectional LSTM architecture [39].

forget invaluable information at any time step. Figure 6 shows the structure of the GRU network, and the detailed equations of the hidden units are described as follows:

$$Z_t = \sigma(W_z X_t + U_z h_{t-1} + b_z), \tag{8}$$

$$r_t = \sigma(W_r X_t + U_r h_{t-1} + b_r), \tag{9}$$

$$\widehat{h}_t = \tanh(W_h X_t + (r_t \odot h_{t-1}))U_h + b_h, \tag{10}$$

$$h_t = (1 - Z_t) \odot h_{t-1} + Z_t \odot \widehat{h}_t, \tag{11}$$

where X_t is the input vector, Z_t (equation (8)) is the update gate vector, r_t (equation (9)) is the reset gate vector, h_t (equation (11)) is the output vector, $\widehat{h}_t$ (equation (10)) is a candidate activation vector, W, U, and b are parameter matrices, and the sign $\odot$ denotes Hadamard product.

3.2. Main Model Development. Optimizing deep learning models require decisions on a combination of large hyperparameters, including the number of layers, number of units, batch size, epochs, and learning rate [54]. A random search can produce an infinite number of hyperparameter

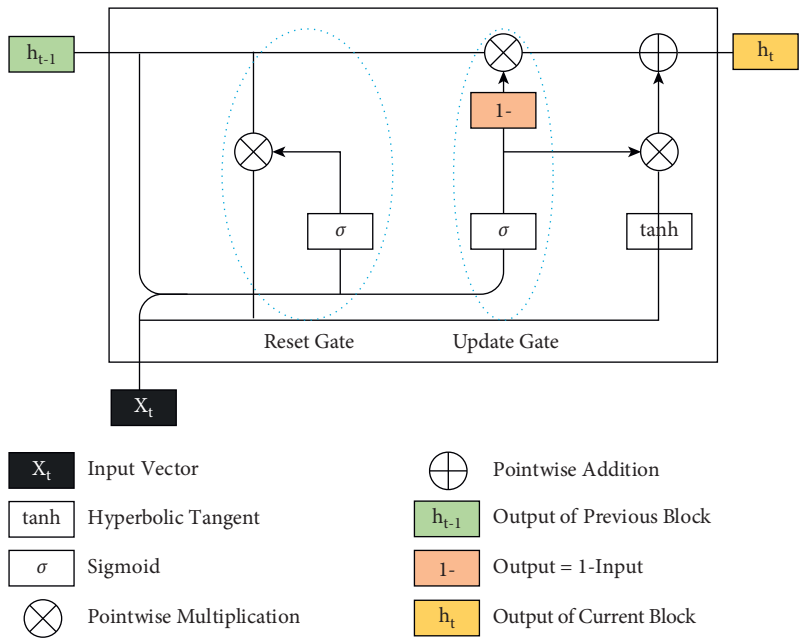


Figure 6: The structure of gated recurrent unit (GRU) network [38].

combinations with a median cost from the four major hyperparameter optimization techniques: trial-and-error, grid, random, and probabilistic approach [31]. Hence, in this study, we use a computationally efficient randomized search method called Keras Tuner developed by the Google team to search random combinations of parameters for optimized performance. The detailed flow chart of the proposed methodology is presented here in Figure 7.

The proposed models applied a double fully connected hidden layer. The minimum and the maximum numbers of neurons at each hidden layer were set with prior experience. Then, the output layer is linked to a dense layer with a single output neuron with a linear activation function. The network is compiled with Adam optimizer and mean squared error loss function, and the details of hyperparameter value ranges and choices are listed in Table 1. Moreover, the following paragraph discusses each hyperparameter optimized using Keras tuner.

3.2.1. Activation Function. In deep learning models, the activation function defines the output from the inputs each node receives. In our case, we applied Rectified Linear Units (ReLU) in all layers except the output layer.

3.2.2. Learning Rate. In deep learning, the learning rate is one of the hyperparameters that decides the step size at each time the model progresses to a minimum loss function. Hence, it is crucial to optimize the learning rate properly; otherwise, the model may converge slowly with too small learning rate or diverge from the optimal error points with large learning rate values [55]. This study sets three values ($1e-2$, $1e-3$, or $1e-4$) to choose using Keras Tuner.

3.2.3. Number of Epochs. The other hyperparameter that decides how much time the deep learning algorithm will learn with the entire training samples. When we set more epochs, the weight in the model will get more chance to update itself. The loss curve goes through different stages such as underfitting, optimal state, or overfitting; even though there are no strict rules for these hyperparameter configurations, we set the minimum (10) and maximum (100) values for optimization using Keras Tuner.

3.2.4. Number of Batch Sizes. The batch size is a sample size cluster processed before the model updates itself.

3.2.5. Drop Out. The dropout layer is a hyperparameter that prevents overfitting and enhances training performance. Hence, at each iteration, the dropout freezes a fraction of the neuron from training, and it is defined on a range of 0 to 1.

Different open-source Python libraries were used for model development, such as Tensorflow, Keras, Scikit-Learn, Matplotlib for visualization, Statsmodels for performance evaluation. Moreover, the simulation was also conducted on a computer with Processor: Intel(R) Core (TM) i7-6500U CPU 2.50 GHz and RAM: 8 Gigabytes memory.

3.3. Input Time Lag Selection. To reprocess the highly correlated time series delay that was decomposed and used as a single input to a deep learning neural network, the autoregressive model using the partial autocorrelation function (pacf) was applied. Equation (12) shows the autoregressive model A.R. (p).

$$x_t = \varphi_1 x_{(t-1)} + \varphi_2 x_{(t-2)} + \cdots + \varphi_p x_{(t-p)} + \varepsilon_t, \quad (12)$$

where φ is the autoregressive parameter, x is the observation at time t, and ε is the weighted noise at time t. The autoregressive model explores the correlation between current and past values [56]. As shown in Figure 8, the partial autocorrelation of daily streamflow time series with a 95% confidence interval, the time delay of 4 days, was considered for both case study areas in this study.

3.4. Performance Measures. The following performance measures were used to evaluate the accuracy of the model developed: coefficient of determination (R^2), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) [57].

(i) Coefficient of Determination (R^2)

$$R^2 = \frac{n\left(\sum Q_{\text{obs}} * Q_{\text{sim}}\right) - \left(\sum Q_{\text{obs}}\right) * \left(\sum Q_{\text{sim}}\right)}{\sqrt[1/2]{\left[n\left(\sum Q_{\text{obs}}^2\right) - \left(\sum Q_{\text{obs}}\right)^2\right] * \left[n\left(\sum Q_{\text{sim}}^2\right) - \left(\sum Q_{\text{sim}}\right)^2\right]}}. \tag{13}$$

(ii) Root Mean Square Error (RMSE)

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{N}\left(Q_{\text{obs}}^t - Q_{\text{sim}}^t\right)^2}{N}}. \tag{14}$$

(iii) Mean Absolute Error (MAE)

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n}\left|Q_{\text{obs}}^t - Q_{\text{sim}}^t\right|. \tag{15}$$

(iv) Mean Absolute Percentage Error (MAE):

$$\text{MAPE} = \frac{100\%}{n}\sum_{t=1}^{n}\left|\frac{Q_{\text{obs}}^t - Q_{\text{sim}}^t}{Q_{\text{obs}}^t}\right|, \tag{16}$$

where Q_{obs} = discharge observed, Q_{sim} = discharge simulated, and n = number of observations. The range of R^2 lies between 0 and 1, representing, respectively, no correlation and a perfect correlation between observed and simulated values, whereas, for RMSE, MAE, and MAPE, the better performance is reached when we are close to 0. If $R^2 > 0.90$, the simulation is very acceptable; if $0.60 < R^2 < 0.90$, the simulation is acceptable; and if $R^2 < 0.60$, the simulation is unacceptable [58].

4. Results and Discussion

Previous analyses revealed a wide range of results, which varied according to the type of deep learning architecture used, the degree of climatic variability, and the timescale used. The GRU model outperforms both the extreme learning machine (ELM) and the least-squares support vector machine (LSSVM) on monthly univariate streamflow data from the Shangjingyou and Fenhe reservoir stations in the upper reaches of the Fenhe River [59]. Sahoo et al. [19] also demonstrated the superiority of LSTM over RNN on univariate daily discharge data from the Basantapur gauging station in India's Mahanadi River basin. Suradhaniwar et al. [7] demonstrated that SARIMA and SVR models outperform NN, LSTM, and RNN models when hourly averaged univariate time series data is used. Even though the best model generalization is complex, case-based analysis is the most effective method for determining which model best fits a given situation [60]. For the first time in this study, we attempted to organize the findings around three distinct themes: model variability, lag time variability, and time series characteristics (climatic variability).

The following section discusses the performance of four selected models under four different input time lag scenarios; in total, 16 experimental results are presented. Tables 2 and 3 show the optimized hyperparameter values for both stations and all scenarios using Keras Tuner. Additionally, Tables 4 and 5 illustrate performance metrics in terms of RMSE, MAE, MAPE, R^2, and TTPE (sec). In the testing phase, single-step streamflow forecasting results demonstrated very acceptable performance for both case study areas.

4.1. Model Variability. We used four evaluation matrices to investigate the effect of model variability on the accuracy of single-step streamflow forecasting. Then, using box plots, we could visualize the spread of prediction error (m^3/s). Additionally, we plotted a bar chart (Figure 9) with different prediction error classes to identify the class limit with the highest error concentration. As a result, as shown in Table 4, GRU's model has a slight performance advantage over MLP and S-LSTM for Borkena station's lag time ($T+2$ and $T+3$). Whereas, for the Gummera catchment (Table 5), MLP outperformed GRU and S-LSTM in terms of performance increment in lag time ($T+1$ and $T+3$). Prediction error box plots and bar chart graphs (Figures 10 and 11) were used to investigate these high-performing architectures further. Hence, the prediction error of GRU is typically concentrated in small ranges (0 to 0.5 m^3/s) for Borkena and (0 to 2.5 m^3/s) for the Gummera catchment. Moreover, as shown in Tables 4 and 5, when considering computational speed, MLP demonstrated the quickest training time per epoch, followed by S-LSTM, GRU, and Bi-LSTM.

4.2. Time Series Characteristics (Climatic Variability). The other major issue affecting deep learning model performance is time series characteristics. As a result, in this study, the four-evaluation metrics displayed Borkena's river catchment forecasting result is more accurate than Gummera's catchment with RMSE, MAE, MAPE, R^2, ranging between (0.81 to 1.53, 0.29 to 0.96, 0.16 to 1.72, 0.96 to 0.99) and (17.43 to 17.99, 7.76 to 10.54, 0.16 to 1.03, 0.89 to 0.90) for both catchments, respectively. The possible cause for this performance variation between the two catchments is the significant natural streamflow time series variability in the Borkena catchment. Furthermore, the spread of prediction error (m^3/sec) in Figures 10 and 11 shows that the error class limit for most cases is smaller in the Borkena station than in the Gummera station.

4.3. Lag Time Variability. In univariate streamflow forecasting, lagged variables are the other significant factors that hold temporal information and affect model performance. Taylor diagram in Figure 12 shows the forecasting ability of

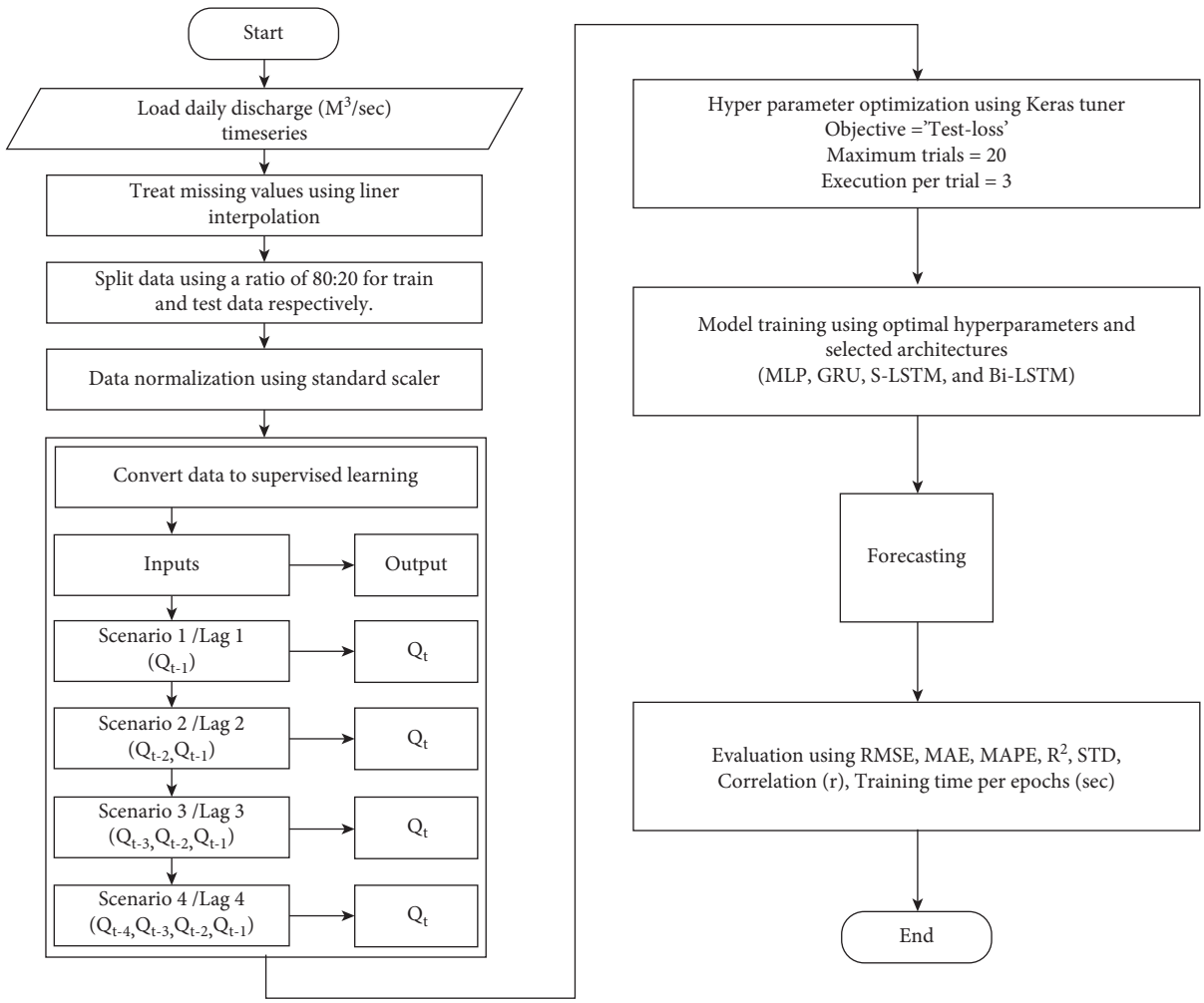


Figure 7: Data analysis flow chart proposed for the study.

Table 1: Model hyperparameter choices or value ranges for optimization by keras tuner.

N°	Hyperparameters	Value Ranges**			Choices	Default
		Min	Max	Step		
1	MLP, LSTM, Bi-GRU+LSTM, or GRU layer 1 units	5	40	5	*	*
2	Dropout 1	0.0	0.3	0.1	*	0.2
3	MLP, LSTM, Bi-LSTM, or GRU layer 2 units	5	40	5	*	*
4	Learning rate	*	*	*	$1e-2$, $1e-3$ or $1e-4$	*
5	Number of epochs	10	100	10	*	*
6	Number of batch sizes	10	100	10	*	*

**Value ranges or choices for optimization by keras tunner: (objective = "test loss," max trials = 20, executions per trial = 3). *Not applicable.

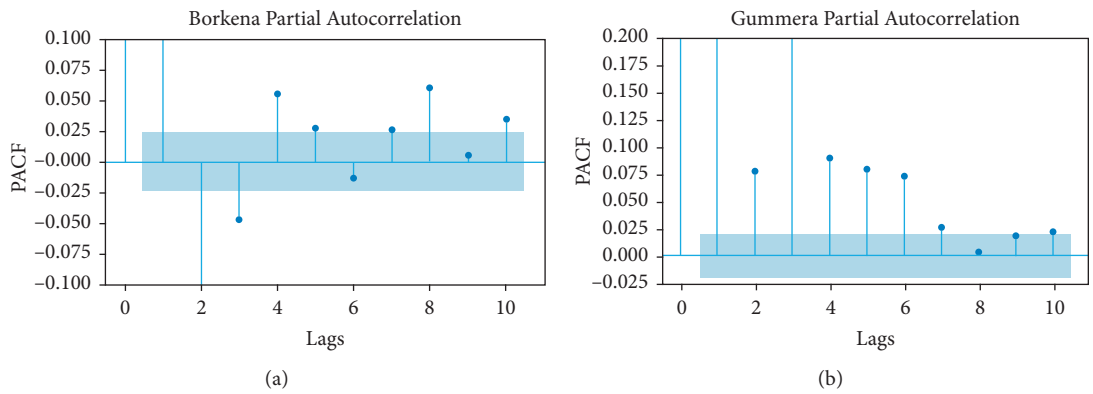


Figure 8: Partial autocorrelation of the daily time series for both catchments (lag units are in days, and the blue shadow lines indicate 95% confidence intervals). (a) Borkena, (b) Gummera.

Table 2: Keras tuner optimized hyperparameter values for Borkena station with its MSE score.

Hyperparameters	MLP				S-LSTM				Bi-LSTM				GRU			
	$T+1$	$T+2$	$T+3$	$T+4$	$T+1$	$T+2$	$T+3$	$T+$	$T+1$	$T+2$	$T+3$	$T+4$	$T+1$	$T+2$	$T+3$	$T+4$
Layer_1_units	35	10	40	35	25	35	40	40	35	40	25	20	25	30	15	10
Dropout_1	0.0	0.0	0.0	0.1	0.1	0.0	0.2	0.2	0.1	0.2	0.0	0.1	0.1	0.0	0.1	0.1
Layer_2_units	10	20	25	15	35	25	30	10	20	30	20	30	20	25	40	25
Learning rate	0.0001	0.0001	0.0001	0.001	0.0001	0.0001	0.001	0.001	0.0001	0.001	0.001	0.001	0.0001	0.001	0.0001	0.001
Number of epochs	90	100	100	40	60	80	20	30	70	20	50	70	50	60	40	50
Number of batch sizes	20	20	30	30	30	40	20	10	40	20	20	30	50	30	10	90
Score (MSE)	0.099	0.099	0.096	0.097	0.099	0.099	0.098	0.096	0.099	0.099	0.097	0.096	0.099	0.099	0.099	0.097

Table 3: Keras tuner optimized hyperparameter values for Gummera station with its MSE score.

Hyperparameters	MLP				S-LSTM				Bi-LSTM				GRU			
	$T+1$	$T+2$	$T+3$	$T+4$	$T+1$	$T+2$	$T+3$	$T+4$	$T+1$	$T+2$	$T+3$	$T+4$	$T+1$	$T+2$	$T+3$	$T+4$
Layer_1_units	25	25	10	25	20	30	15	25	10	25	15	15	5	40	25	15
Dropout_1	0.1	0.0	0.0	0.0	0.0	0.0	0.2	0.2	0.0	0.0	0.1	0.1	0.0	0.0	0.0	0.2
Layer_2_units	15	40	30	25	30	10	35	10	5	20	10	20	30	20	25	15
Learning rate	0.0001	0.0001	0.0001	0.001	0.001	0.001	0.01	0.001	0.001	0.001	0.01	0.001	0.01	0.01	0.0001	0.001
Number of epochs	70	90	80	50	60	50	50	80	60	90	50	30	80	40	50	30
Number of batch sizes	60	60	70	60	80	40	30	50	50	40	20	10	20	30	50	40
Score (MSE)	0.019	0.011	0.012	0.015	0.019	0.011	0.013	0.015	0.018	0.011	0.013	0.013	0.018	0.012	0.013	0.016

the proposed models with the observed test data for both case study areas. The diagram is designed on a two-dimensional scale: the standard deviation on the polar axis, root mean square error, and correlation coefficient on the radial axis. It shows that, irrespective of deep learning models, forecasting with a lag time of four gives us a time series closest to the standard deviation of the actual test observations. Moreover, Figures 13 and 14 also display the

Table 4: Performance comparison of the proposed models for different input time lags in Borkena station.

Borkena	$T+1$					$T+2$					$T+3$					$T+4$				
	RMSE	MAE	MAPE	R^2	TTPE * (sec)	RMSE	MAE	MAPE	R^2	TTPE* (sec)	RMSE	MAE	MAPE	R^2	TTPE * (sec)	RMSE	MAE	MAPE	R^2	TTPE* (sec)
MLP	1.09	0.47	0.85	0.98	0.24	0.89	0.38	0.76	0.99	0.16	1.06	0.35	0.19	0.98	0.12	**0.81**	**0.29**	**0.34**	**0.99**	0.29
GRU	1.15	0.64	0.84	0.98	1.29	**0.85**	**0.31**	**0.16**	**0.99**	0.79	**0.91**	**0.35**	**0.18**	**0.98**	0.99	1.35	0.69	1.41	0.97	1.31
S-LSTM	1.07	**0.38**	**0.29**	**0.98**	0.30	0.87	0.36	0.26	0.99	0.54	1.02	0.64	0.99	0.98	0.81	1.53	0.96	1.72	0.96	0.58
Bi-LSTM	1.06	0.39	0.31	0.98	0.88	0.90	0.46	0.76	0.99	1.55	1.12	0.51	0.95	0.98	1.55	0.98	0.44	0.53	0.98	5.3

*TTPE(training time per epochs). The bold values indicate the best performance score for each time lag.

Table 5: Performance comparison of the proposed models for different input time lags in Gummera station.

Gummera	$T+1$					$T+2$					$T+3$					$T+4$				
	RMSE	MAE	MAPE	R^2	TTPE * (sec)	RMSE	MAE	MAPE	R^2	TTPE* (sec)	RMSE	MAE	MAPE	R^2	TTPE * (sec)	RMSE	MAE	MAPE	R^2	TTPE* (sec)
MLP	17.67	**7.84**	**0.18**	**0.90**	0.88	17.68	7.87	0.19	0.90	0.75	**17.43**	**7.76**	**0.18**	**0.90**	**0.61**	17.56	9.73	0.78	0.90	0.95
GRU	17.66	8.39	0.41	0.90	0.51	17.76	**7.82**	**0.16**	**0.90**	1.43	17.47	8.29	0.35	0.90	1.24	17.99	10.54	1.03	0.89	0.53
S-LSTM	17.66	8.29	0.39	0.90	1.54	17.71	8.04	0.22	0.90	0.75	17.69	8.51	0.32	0.90	2.99	17.69	**8.41**	**0.35**	**0.90**	3.31
Bi-LSTM	17.63	7.98	0.27	0.90	0.92	17.93	8.39	0.29	0.89	2.34	17.83	9.18	0.56	0.89	2.77	17.56	8.56	0.39	0.90	1.84

*TTPE(training time per epochs).The bold values indicate the best performance score for each time lag.

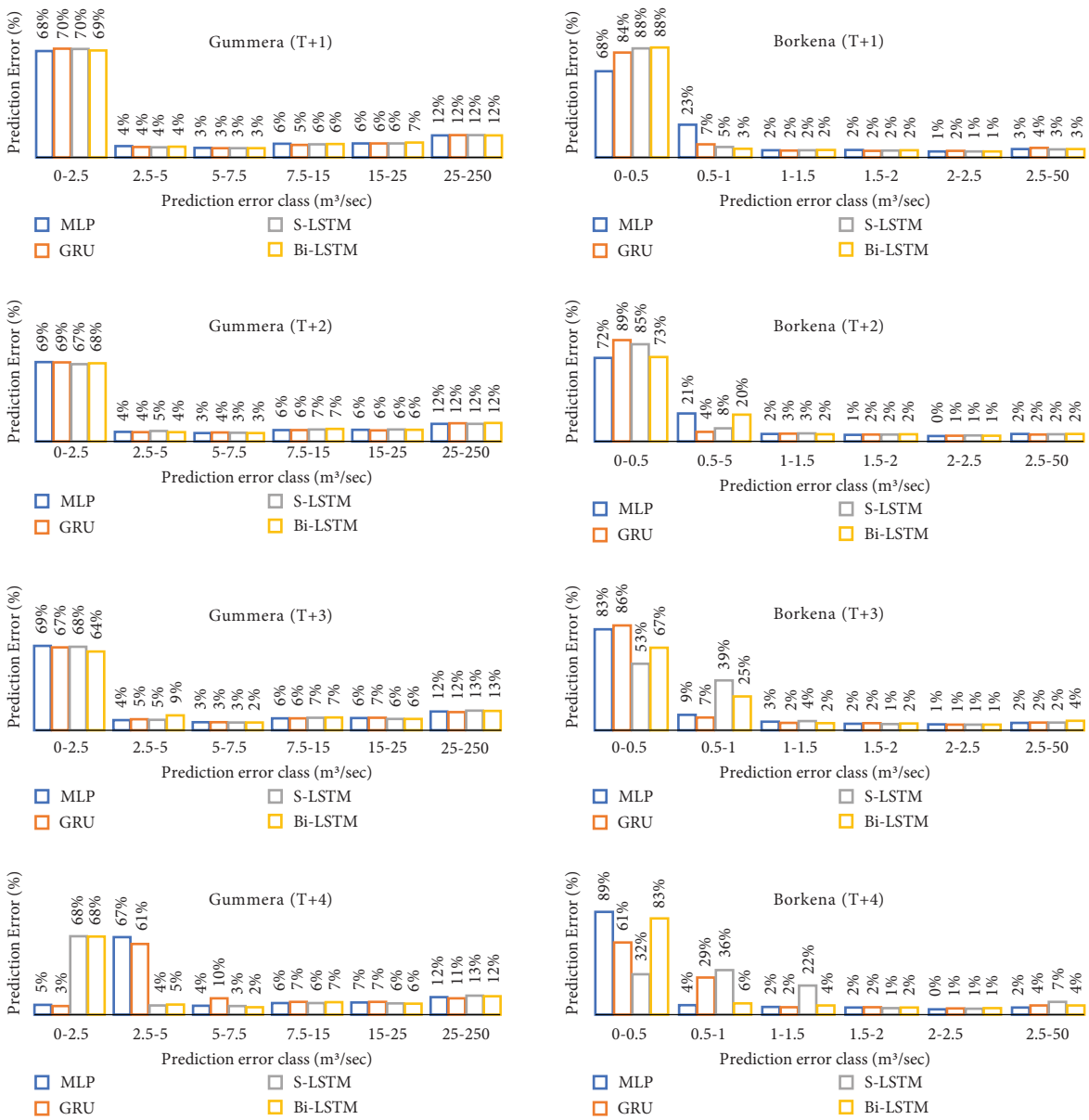


Figure 9: Bar chart showing the frequency (%) of absolute prediction errors |PE| (m^3/sec) with different class limits for the proposed four models and input lagged time variables in both catchments.

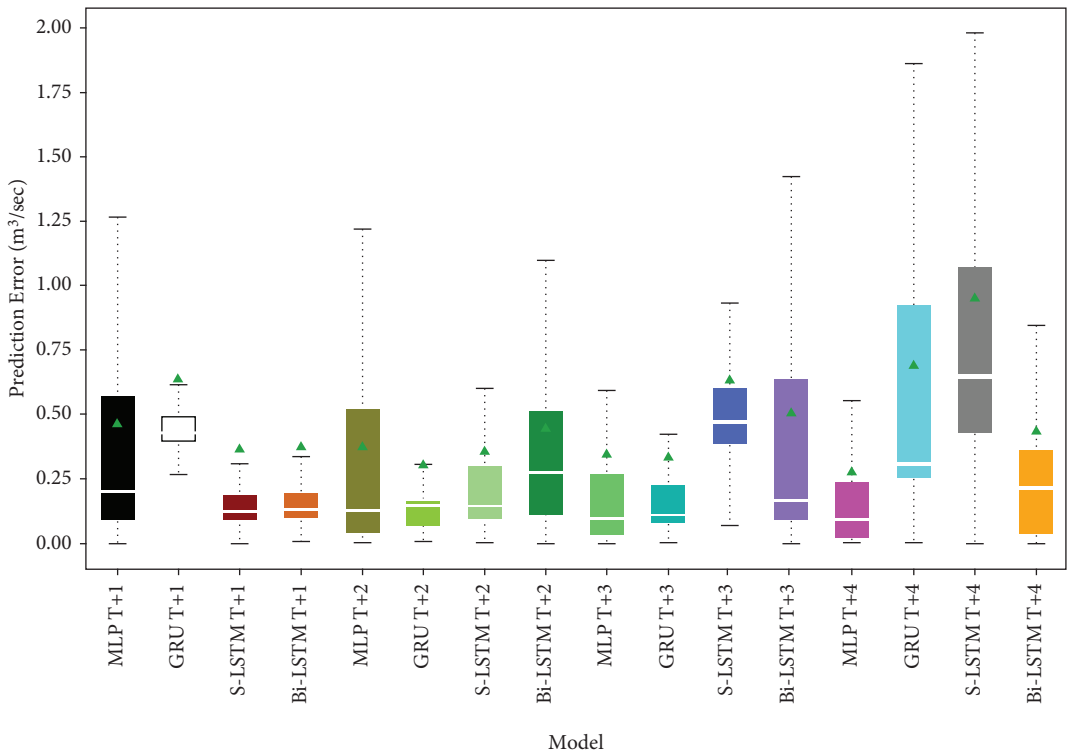


Figure 10: Spread of prediction error (m^3/s) or box plot for the proposed models during the test period (Borkena station).

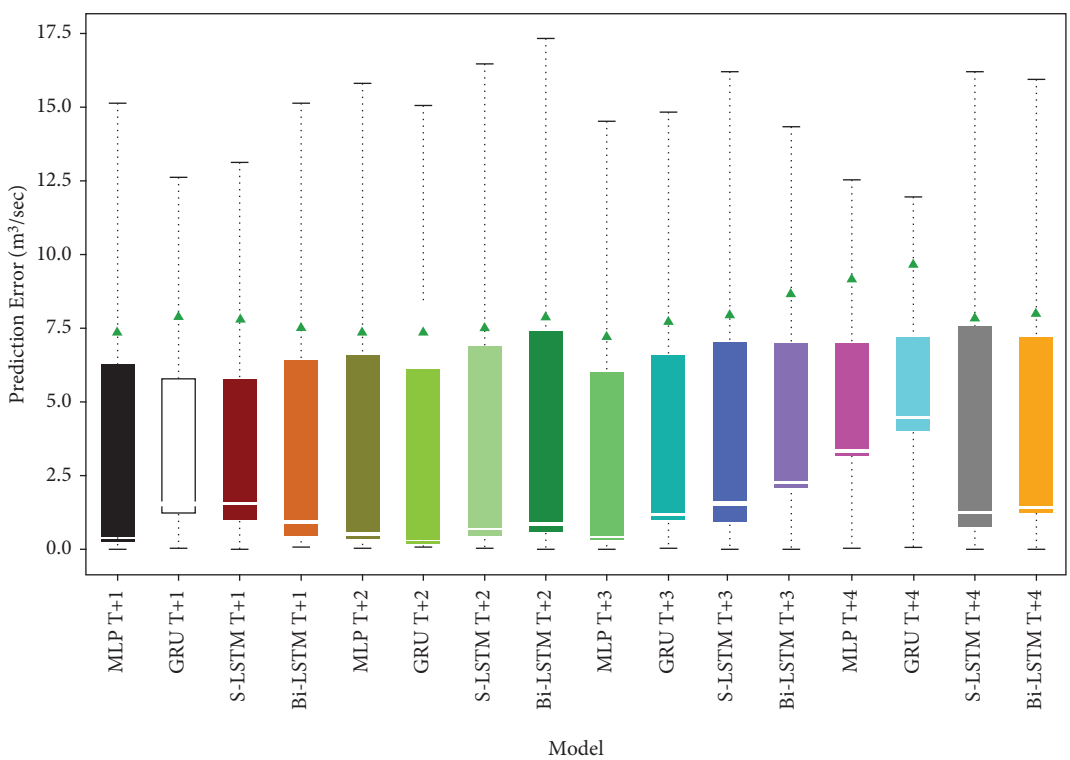
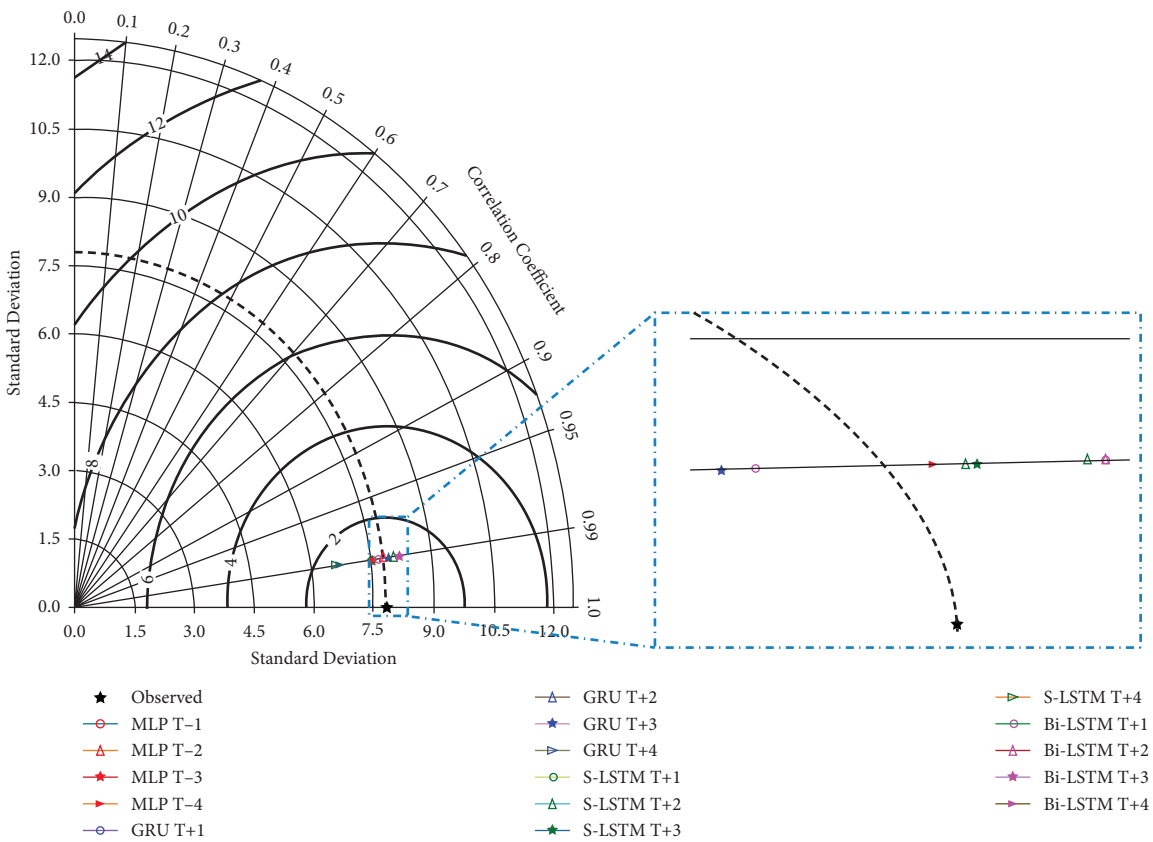
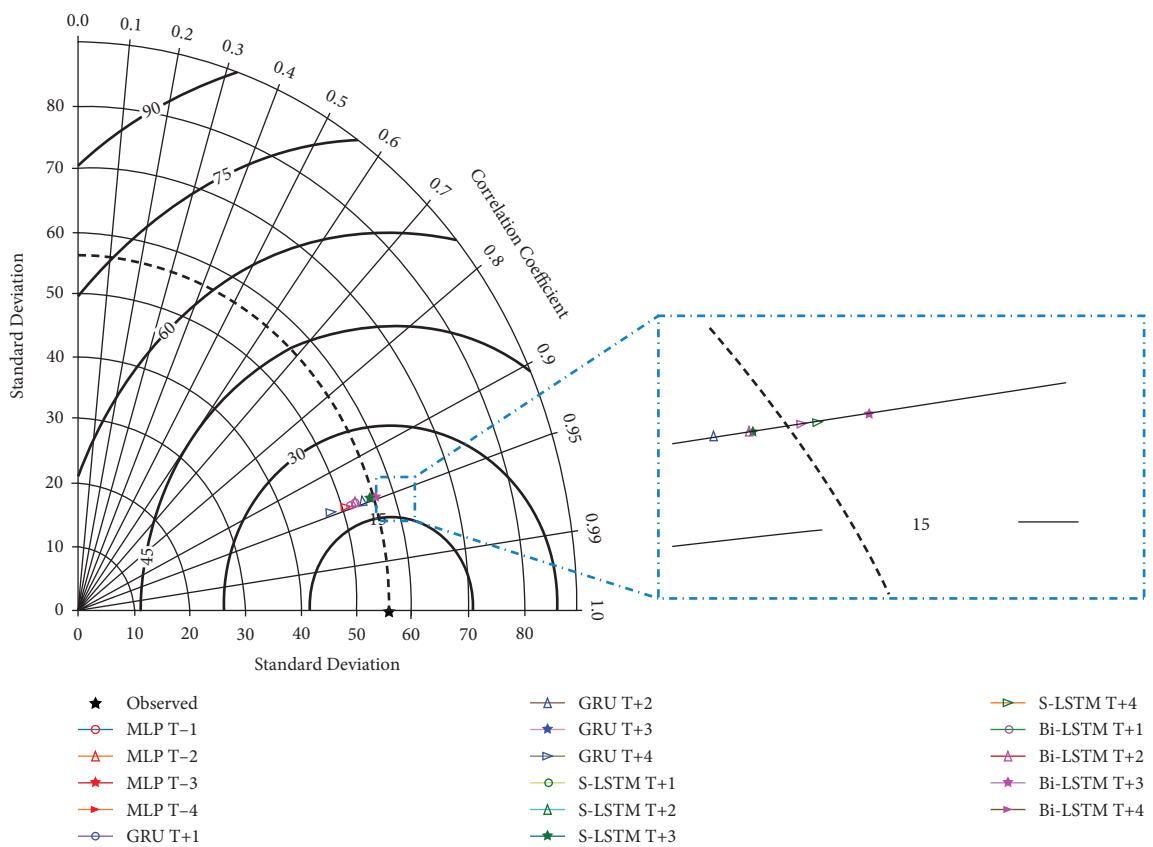


Figure 11: Spread of prediction error (m^3/s) or box plot for the proposed models during the test period (Gummera station).



(a)



(b)

Figure 12: Taylor diagram displays the standard deviations, root mean square error, and correlation coefficient between observed and predicted streamflow for the proposed four models and lagged time variables (Q m^3/s). (a) Borkena, (b) Gummera.

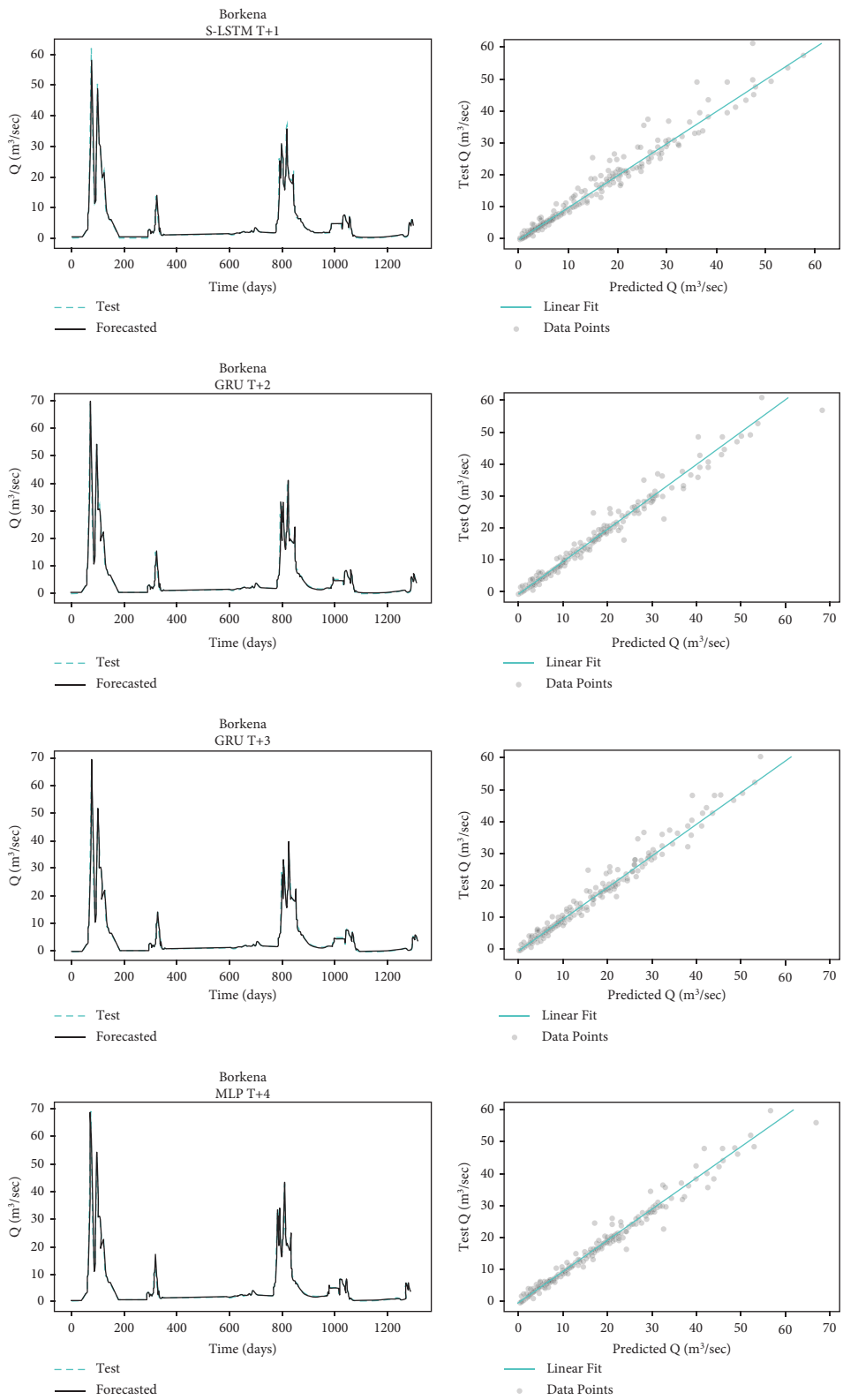


Figure 13: Comparison of true values and predicted values of the optimized high score deep learning model for each time lag (Borkena station).

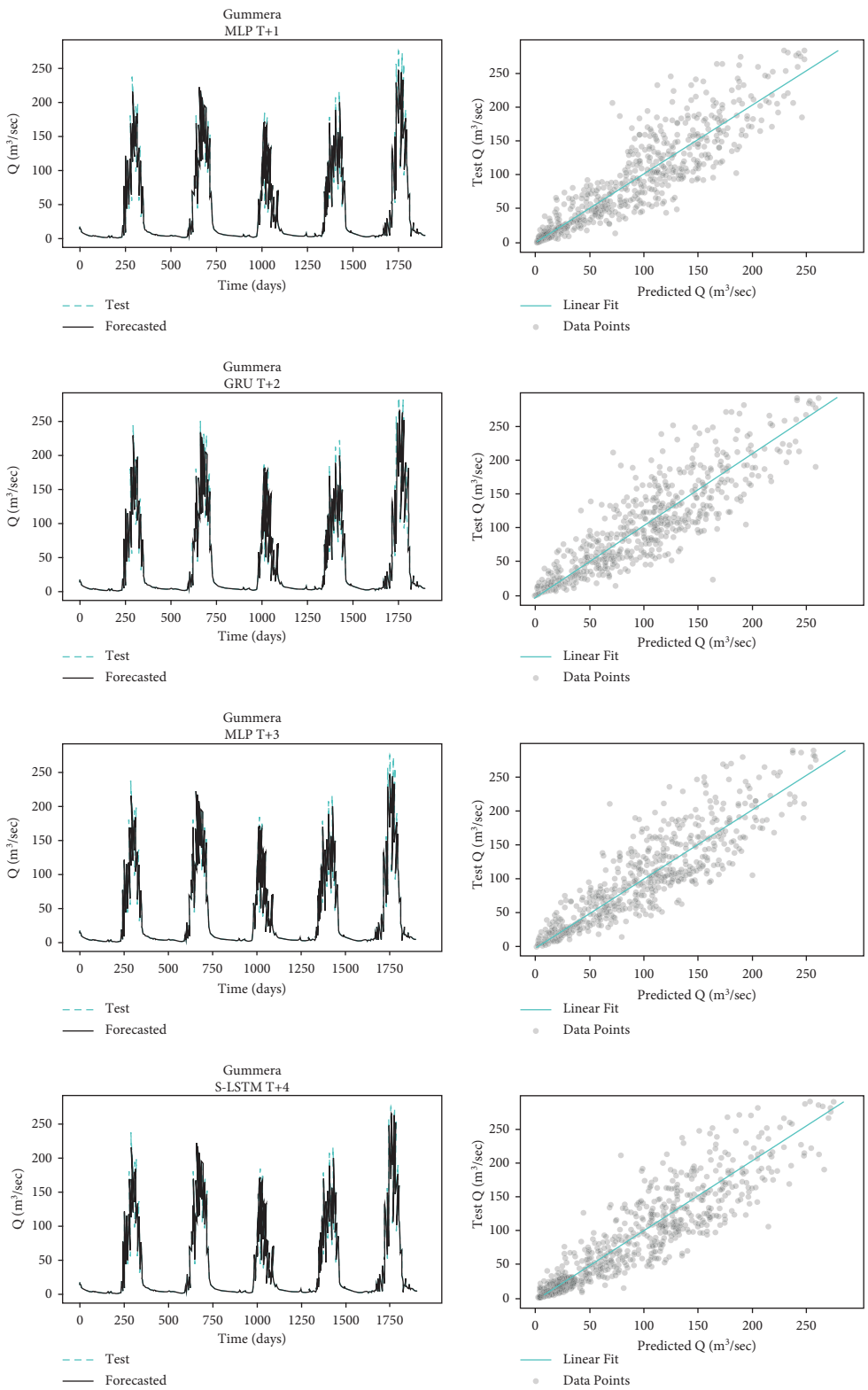


Figure 14: Comparison of true values and predicted values of the optimized high score deep learning model for each time lag (Gummera station).

actual values and predicted values of the optimized high score deep learning model for each time lag and both case study catchments.

5. Conclusions

This study showed a comparative analysis of different deep learning algorithms for one-step daily streamflow forecasting at two subcatchment stream flow outlets. MLP, S-LSTM, Bi-LSTM, and GRU are the four algorithms used in this study. The study clearly showed the impacts of climatic (time series characteristics) and lagged time variability on the performance of different proposed deep learning models. The following key points will elaborate on the outcome of this research.

(i) Deep learning models have excellent potential in forecasting short-term daily streamflow in different time series characteristics.

(ii) The performance of deep learning models for short-term streamflow forecasting varies with time series characteristics and input time lag variations. Moreover, the Borkena station has more significant natural streamflow variability than the Gummera station, which is also reflected in the model results. Hence, this study showed that catchment response variability impacts deep learning model performance.

(iii) MLP and GRU outperform S-LSTM and Bi-LSTM on a nearly equal basis for single-step short-term streamflow forecasting in both case study areas. However, the performance is relative to the lagged time variations.

(iv) Catchment characteristics had a high impact on the performance of streamflow forecasting than deep learning model architectures and lagged time variations.

(v) The study also showed that classical MLP could almost equally perform with S-LSTM and GRU deep learning networks on a small amount of streamflow time series data.

Future research may further expand this study's findings to other climatic regions, hybrid deep learning model architectures, hyperparameter tuning, and lagged time selection methods. We must also investigate the effects of large input variability on deep learning models for univariate streamflow forecasting in all its implications. As part of our future work, we plan to implement an ensemble learning approach to simulate streamflow from remote sensing-derived data products (precipitation and vegetation indexes) using a combination of neural networks, decision trees, and boosting algorithms.

Data Availability

The corresponding author's raw metrological and hydrological data sets used for the Borkena and Gummera watershed are available upon request. However, authorization letters are required from the Ethiopian National Metrological Agency (NMA) and Ethiopian Ministry of Water, and Energy (MoWE) (https://mowe.gov.et/)

Conflicts of Interest

The authors declare that they have no conflicts of interest for this article.

Acknowledgments

The authors acknowledge the Ethiopian Ministry of Water, and Energy (MoWE) for hydrological data and the Ethiopian National Metrological Agency (NMA) for metrological data.

References

[1] P. Sharma and D. Machiwal, "Streamflow forecasting," in *Advances in Streamflow Forecasting*, pp. 1–50, Elsevier, Amsterdam, Netherlands, 2021.

[2] Z. M. Yaseen, A. El-shafie, O. Jaafar, H. A. Afan, and K. N. Sayl, "Artificial intelligence based models for streamflow forecasting: 2000-2015," *Journal of Hydrology*, vol. 530, pp. 829–844, 2015.

[3] Z. M. Yaseen, W. H. M. W. Mohtar, A. M. S. Ameen et al., "Implementation of univariate paradigm for streamflow simulation using hybrid data-driven model: case study in tropical region," *IEEE Access*, vol. 7, pp. 74471–74481, 2019.

[4] R. K. Jaiswal, S. Ali, and B. Bharti, "Comparative evaluation of conceptual and physical rainfall-runoff models," *Applied Water Science*, vol. 10, no. 1, Article ID 48, 2020.

[5] M. Ghaith, A. Siam, Z. Li, and W. El-Dakhakhni, "Hybrid hydrological data-driven approach for daily streamflow forecasting," *Journal of Hydrologic Engineering*, vol. 25, no. 2, Article ID 04019063, 2020.

[6] Z. Zhang, Q. Zhang, and V. P. Singh, "Univariate streamflow forecasting using commonly used data-driven models: literature review and case study," *Hydrological Sciences Journal*, vol. 63, no. 7, pp. 1091–1111, 2018.

[7] S. Suradhaniwar, S. Kar, S. S. Durbha, and A. Jagarlapudi, "Time series forecasting of univariate agrometeorological data: a comparative performance evaluation via one-step and multi-step ahead forecasting strategies," *Sensors*, vol. 21, no. 7, Article ID 2430, 2021.

[8] A. Danandeh Mehr and M. J. S. Safari, "Genetic programming for streamflow forecasting," in *Advances in Streamflow Forecasting*, pp. 193–214, Elsevier, Amsterdam, Netherlands, 2021.

[9] A. Essien and C. Giannetti, "A deep learning framework for univariate time series prediction using convolutional LSTM stacked autoencoders," in *Proceedings of the IEEE International Symposium on Inovations in Intelligent Systems and Applications (INISTA)*, pp. 1–6, Sofia, Bulgaria, July 2019.

[10] Z. M. Yaseen, S. O. Sulaiman, R. C. Deo, and K.-W. Chau, "An enhanced extreme learning machine model for river flow forecasting: state-of-the-art, practical applications in water resource engineering area and future research direction," *Journal of Hydrology*, vol. 569, pp. 387–408, 2019.

[11] O. Oyebode and D. Stretch, "Neural network modeling of hydrological systems: a review of implementation techniques," *Natural Resource Modeling*, vol. 32, no. 1, Article ID 12189, 2019.

[12] F. Cui, S. Q. Salih, B. Choubin, S. K. Bhagat, P. Samui, and Z. M. Yaseen, "Newly explored machine learning model for river flow time series forecasting at Mary River, Australia," *Environmental Monitoring and Assessment*, vol. 192, no. 12, Article ID 761, 2020.

[13] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[14] Y. Bai, N. Bezak, B. Zeng, C. Li, K. Sapač, and J. Zhang, "Daily runoff forecasting using a cascade long short-term memory model that considers different variables," *Water Resources Management*, vol. 35, no. 4, pp. 1167–1181, 2021.

[15] D. Couta, Y. Zhang, and Y. Li, "River flow forecasting using long short-term memory," *DEStech Transactions on Computer Science and Engineering*, vol. 16, 2019.

[16] A. Z. U. Din, Y. Ayaz, M. Hasan, J. Khan, and M. Salman, "Bivariate short-term electric power forecasting using LSTM network," in *Proceedings of the 2019 International Conference on Robotics and Automation in Industry (ICRAI)*, pp. 1–8, Rawalpindi, Pakistan, Octomber 2019.

[17] M. Jain, S. Manandhar, Y. H. Lee, S. Winkler, and S. Dev, "Forecasting precipitable water vapor using LSTMs," in *Proceedings of the 2020 IEEE USNC-CNC-URSI North American Radio Science Meeting (Joint with AP-S Symposium)*, Montreal, QC, Canada, July 2020.

[18] M. Rahimzad, A. Moghaddam Nia, H. Zolfonoon, J. Soltani, A. Danandeh Mehr, and H.-H. Kwon, "Performance comparison of an lstm-based deep learning model versus conventional machine learning algorithms for streamflow forecasting," *Water Resources Management*, vol. 35, 2021.

[19] B. B. Sahoo, R. Jha, A. Singh, and D. Kumar, "Long short-term memory (LSTM) recurrent neural network for low-flow hydrological time series forecasting," *Acta Geophysica*, vol. 67, no. 5, pp. 1471–1481, 2019.

[20] L. Yan, J. Feng, and T. Hang, "Small watershed stream-flow forecasting based on LSTM," *Advances in Intelligent Systems and Computing*, vol. 935, pp. 1006–1014, 2019.

[21] K. Cho, B. van Merrienboer, C. Gulcehre et al., "Learning phrase representations using RNN encoder-decoder for statistical machine translation," 2014, https://arxiv.org/abs/1406.1078.

[22] M. M. Hussain, S. H. Bari, I. Mahmud, and M. I. H. Siddiquee, "Application of different artificial neural network for streamflow forecasting," in *Advances in Streamflow Forecasting*, pp. 149–170, Elsevier, Amsterdam, Netherlands, 2021.

[23] P. Lara-Benítez, M. Carranza-García, and J. C. Riquelme, "An experimental review on deep learning architectures for time series forecasting," *International Journal of Neural Systems*, vol. 31, no. 3, Article ID 2130001, 2021.

[24] K. A. Althelaya, E.-S. M. El-Alfy, and S. Mohammed, "Stock market forecast using multivariate analysis with bidirectional and stacked (LSTM, GRU)," in *Proceedings of the 2018 21st Saudi Computer Society National Computer Conference (NCC)*, pp. 1–7, Riyadh, Saudi Arabia, April 2018.

[25] S. Kumar, L. Hussain, S. Banarjee, and M. Reza, "Energy load forecasting using deep learning approach LSTM and GRU in spark cluster," in *Proceedings of the 2018 Fifth International Conference on Emerging Applications of Information Technology (EAIT)*, pp. 1–4, Kolkata, India, January 2018.

[26] F. Shahid, A. Zameer, and M. Muneeb, "Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM," *Chaos, Solitons, and Fractals*, vol. 140, Article ID e110212, 2020.

[27] P. T. Yamak, L. Yujian, and P. K. Gadosey, "A comparison between ARIMA, LSTM, and GRU for time series forecasting," in *Proceedings of the 2019 2nd International Conference on Algorithms, Computing and Artificial Intelligence*, pp. 49–55, New York, NY, USA, December 2019.

[28] G. Papacharalampous, H. Tyralis, and D. Koutsoyiannis, "Univariate time series forecasting of temperature and precipitation with a focus on machine learning algorithms: a multiple-case study from Greece," *Water Resources Management*, vol. 32, no. 15, pp. 5207–5239, 2018.

[29] H. Tyralis and G. Papacharalampous, "Variable selection in time series forecasting using random forests," *Algorithms*, vol. 10, no. 4, Article ID 114, 2017.

[30] G. Papacharalampous, H. Tyralis, and D. Koutsoyiannis, "One-step ahead forecasting of geophysical processes within a purely statistical framework," *Geoscience Letters*, vol. 5, no. 1, Article ID 12, 2018.

[31] J. F. Torres, D. Hadjout, A. Sebaa, F. Martínez-Álvarez, and A. Troncoso, "Deep learning for time series forecasting: a survey," *Big Data*, vol. 9, no. 1, pp. 3–21, 2021.

[32] W. Niu and Z. Feng, "Evaluating the performances of several artificial intelligence methods in forecasting daily streamflow time series for sustainable water resources management," *Sustainable Cities and Society*, vol. 64, Article ID 102562, 2021.

[33] F. F. Abera, S. Arega, and B. H. Gedamu, "Climate change induced precipitation and temperature effects on water resources: the case of Borkena watershed in the highlands of Wollo, central Ethiopia," *Water Conserv Sci Eng*, vol. 5, no. 1, pp. 53–66, 2020.

[34] F. D. Mwale, A. J. Adeloye, and R. Rustum, "Infilling of missing rainfall and streamflow data in the Shire River basin, Malawi-a self organizing map approach," *Physics and Chemistry of the Earth, Parts A/B/C*, vol. 50-52, pp. 34–43, 2012.

[35] B. Mesta, O. B. Akgun, and E. Kentel, "Alternative solutions for long missing streamflow data for sustainable water resources management," *International Journal of Water Resources Development*, vol. 37, no. 5, pp. 882–905, 2021.

[36] F. B. Hamzah, F. Mohd Hamzah, S. F. Mohd Razali, and H. Samad, "A comparison of multiple imputation methods for recovering missing data in hydrological studies," *Civil Engineering Journal*, vol. 7, no. 9, pp. 1608–1619, 2021.

[37] A. Gnauck, "Interpolation and approximation of water quality time series and process identification," *Analytical and Bioanalytical Chemistry*, vol. 380, no. 3, pp. 484–492, 2004.

[38] D. K. Yawson, V. M. Kongo, and R. K. Kachroo, "Application of linear and nonlinear techniques in river flow forecasting in the Kilombero River basin, Tanzania," *Hydrological Sciences Journal*, vol. 50, no. 5, Article ID 796, 2005.

[39] A. K. Fleig, L. M. Tallaksen, H. Hisdal, and D. M. Hannah, "Regional hydrological drought in north-western Europe: linking a new regional drought area index with weather types," *Hydrological Processes*, vol. 25, no. 7, pp. 1163–1179, 2011.

[40] S. Gao, Y. Huang, S. Zhang et al., "Short-term runoff prediction with GRU and LSTM networks without requiring time step optimization during sample generation," *Journal of Hydrology*, vol. 589, Article ID 125188, 2020.

[41] C. Shen, "A transdisciplinary review of deep learning research and its relevance for water resources scientists," *Water Resources Research*, vol. 54, no. 11, pp. 8558–8593, 2018.

[42] S. ardabili Faizollahzadeh, A. Mosavi, and M. Dehghani, "Deep learning and machine learning in hydrological processes, climate change and earth systems: a systematic review," in *Engineering for Sustainable Future*Springer, Berlin, Germany, 2019.

[43] N. Hayatbini, B. Kong, K.-L. Hsu et al., "Conditional generative adversarial networks (cGANs) for near real-time precipitation estimation from multispectral GOES-16 satellite imageries-PERSIANN-cGAN," *Remote Sensing*, vol. 11, no. 19, 2019.

[44] S. Nacar, M. A. Hınıs, and M. Kankal, "Forecasting daily streamflow discharges using various neural network models and training algorithms," *KSCE Journal of Civil Engineering*, vol. 22, no. 9, pp. 3676–3685, 2018.

[45] A. Sahoo, S. Samantaray, and D. K. Ghose, "Stream flow forecasting in mahanadi river basin using artificial neural networks," *Procedia Computer Science*, vol. 157, pp. 168–174, 2019.

[46] O. B. Sezer, M. U. Gudelek, and A. M. Ozbayoglu, "Financial time series forecasting with deep learning: a systematic literature review: 2005–2019," *Applied Soft Computing*, vol. 90, Article ID 106181, 2020.

[47] W. Wang, P. H. A. J. M. V. Gelder, J. K. Vrijling, and J. Ma, "Forecasting daily streamflow using hybrid ANN models," *Journal of Hydrology*, vol. 324, no. 1, pp. 383–399, 2006.

[48] R. K. Mishra, G. Y. S. Reddy, and H. Pathak, "The understanding of deep learning: a comprehensive review," *Mathematical Problems in Engineering*, vol. 2021, Article ID e5548884, 15 pages, 2021.

[49] ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, "Artificial neural networks in hydrology I: preliminary concepts," *Journal of Hydrologic Engineering*, vol. 5, no. 2, pp. 115–123, 2000.

[50] O. Sharma, "A new activation function for deep neural network," in *Proceedings of the 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, pp. 84–86, Faridabad, India, February 2019.

[51] E. B. Wegayehu and F. B. Muluneh, "Multivariate streamflow simulation using hybrid deep learning models," *Computational Intelligence and Neuroscience*, vol. 2021, Article ID e5172658, 16 pages, 2021.

[52] J. Yin, Z. Deng, A. V. M. Ines, J. Wu, and R. Eeswaran, "Forecast of short-term daily reference evapotranspiration under limited meteorological variables using a hybrid bidirectional long short-term memory model (Bi-LSTM)," *Agricultural Water Management*, vol. 242, Article ID 106386, 2020.

[53] Q. Zou, Q. Xiong, Q. Li, H. Yi, Y. Yu, and C. Wu, "A water quality prediction method based on the multi-time scale bidirectional long short-term memory network," *Environmental Science and Pollution Research*, vol. 27, no. 14, pp. 16853–16864, 2020.

[54] J. F. Torres, D. Gutiérrez-Avilés, A. Troncoso, and F. Martínez-Álvarez, "Random hyper-parameter search-based deep neural network for power consumption forecasting," *Advances in Computational Intelligence*, pp. 259–269, Springer, Berlin, Germany, 2019.

[55] J. Konar, P. Khandelwal, and R. Tripathi, "Comparison of various learning rate scheduling techniques on convolutional neural network," in *Proceedings of the 2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*, pp. 1–5, Bhopal, India, February 2020.

[56] S. Aljahdali, A. Sheta, and H. Turabieh, "river flow forecasting: a comparison between feedforward and layered recurrent neural network," in *Innovation in Information Systems and Technologies to Support Learning Research*, pp. 523–532, Springer, Berlin, Germany, 2020.

[57] S. Kumar, T. Roshni, and D. Himayoun, "A comparison of emotional neural network (ENN) and artificial neural network (ANN) approach for rainfall-runoff modelling," *Civil Engineering Journal*, vol. 5, no. 10, pp. 2120–2130, 2019.

[58] H. Apaydin, H. Feizi, M. T. Sattari, M. S. Colak, S. Shamshirband, and K.-W. Chau, "Comparative analysis of recurrent neural network architectures for reservoir inflow forecasting," *Water*, vol. 12, no. 5, Article ID 1500, 2020.

[59] X. Zhao, H. Lv, S. Lv, Y. Sang, Y. Wei, and X. Zhu, "Enhancing robustness of monthly streamflow forecasting model using gated recurrent unit based on improved grey wolf optimizer," *Journal of Hydrology*, vol. 601, Article ID 126607, 2021.

[60] M. B. Gunathilake, C. Karunanayake, A. S. Gunathilake et al., "Hydrological models and artificial neural networks (ANNs) to simulate streamflow in a tropical catchment of Sri Lanka," *Applied Computational Intelligence and Soft Computing*, vol. 2021, Article ID e6683389, 9 pages, 2021.