Research Article
Multivariate Regression Analysis and Statistical Modeling for Summer Extreme Precipitation over the Yangtze River Basin, China

Tao Gao1,2,3 and Lian Xie1,3

1 Key Laboratory of Marine Environment and Ecology (Ocean University of China), Ministry of Education of China, Qingdao 266100, China
2 Department of Resources and Environmental Sciences, Heze University, Heze 274000, China
3 Department of Marine, Earth and Atmospheric Sciences, North Carolina State University, Raleigh, NC 27695, USA

Correspondence should be addressed to Tao Gao; tgao.oc@gmail.com

Received 22 November 2013; Accepted 4 December 2013; Published 8 January 2014

Academic Editor: Huiwang Gao

Copyright © 2014 T. Gao and L. Xie. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Extreme precipitation is likely to be one of the most severe meteorological disasters in China; however, studies on the physical factors affecting precipitation extremes and corresponding prediction models are not accurately available. From a new point of view, the sensible heat flux (SHF) and latent heat flux (LHF), which have significant impacts on summer extreme rainfall in Yangtze River basin (YRB), have been quantified and then selections of the impact factors are conducted. Firstly, a regional extreme precipitation index was applied to determine Regions of Significant Correlation (RSC) by analyzing spatial distribution of correlation coefficients between this index and SHF, LHF, and sea surface temperature (SST) on global ocean scale; then the time series of SHF, LHF, and SST in RSCs during 1967–2010 were selected. Furthermore, other factors that significantly affect variations in precipitation extremes over YRB were also selected. The methods of multiple stepwise regression and leave-one-out cross-validation (LOOCV) were utilized to analyze and test influencing factors and statistical prediction model. The correlation coefficient between observed regional extreme index and model simulation result is 0.85, with significant level at 99%. This suggested that the forecast skill was acceptable although many aspects of the prediction model should be improved.

1. Introduction

Temporal and spatial variations in extreme precipitation events often result in serious impacts on human society and ecological environment. And higher frequency of these extremes poses vast catastrophic consequences, including floods, landslides, and urban waterlog (e.g., [1, 2]). In recent years numerous disastrous floods have been documented worldwide, for example, the intense flash flooding occurred in Minnesota, Wisconsin, in the United States, in June, 2012 [3], and the extreme rainfall in Beijing, China, in July, 2012 [4]; all those events have caused devastating social impacts. Moreover, in the context of global climate change, previous studies have suggested that many regions over the world would experience more frequent extreme precipitation with the enhancement of anthropogenic greenhouse gas and aerosol emissions (e.g., [5–8]). Therefore, projection of seasonal variations in precipitation extremes on local and regional scale is overwhelmingly important for reducing casualties and property losses as well as water resource management.

However, the two major current methods, for dynamical model and statistics, show low operational skills for forecasting local extreme precipitation. On the basis of dynamical prediction systems, general and regional circulation models (GCMs and RCMs) are useful to reproduce large-scale circulation, whereas they cannot very well characterize heavy rainfall features within 50 km, because local extreme precipitation is strongly influenced by land-air contrasts or topography, which are not well represented by coarse resolution models (e.g., [9, 10]). Another commonly used prediction method, statistical downscaling, is to establish an empirical
relationship between GCM-based quantitative predictions on large-scale and local climatic variable (e.g., [11, 12]). Therefore, the prediction results by using statistical downscaling methodology are based on the output of climatic model to some extent, which also cannot capture accurate changes of extreme precipitation in the future. Consequently, the multiple regression, being applied to forecast regional heavy rainfall by utilizing historical rain gauge station dataset, has drawn more and more attention in recent decades.

In China, Fan et al. [13] forecasted the summer mean precipitation in the midlower reaches of Yangtze River basin (YRB) using regression model based on observed dataset and suggested that the statistical forecasting model had the prediction scores as 60%–70%. Nevertheless, it should be noted that the trends of seasonal mean rainfall were not always consistent with the ones for extreme precipitation in many regions. For instance, Alexander et al. [6] analyzed the global variations in extreme precipitation events based on large-scale observations and reported that there were indications for increases in frequency and intensity of precipitation extremes even in the regions where mean rainfall decreased in midlatitudes of the Northern Hemisphere. For this reason, prediction of extreme precipitation in YRB is necessary and practical considering complex climatic conditions and dense population as well as rapid development of economy within the region. However, previous studies were not available for forecasting extreme precipitation in YRB based on the statistical model using observed dataset; this is the motivation for us to conduct the present research to establish statistical prediction model with respect to extreme precipitation in YRB.

2. Data and Methodology

2.1. Study Area and Data. The Yangtze River is located between 91°E–122°E and 25°N–35°N with a total drainage area of 1,808,500 km² (Figure 1). Originating from Qinghai-Tibet Plateau in southwest China, the Yangtze River flows into East China Sea in Shanghai. This is the longest river in China and the third longest river over the world. The YRB exhibits distinct changes of precipitation and is vulnerable to climate change (e.g., [14, 15]). Frequent floods have occurred in history, leading to substantial losses in economy and lives in YRB. For example, the disastrous floods in 1998 triggered a death toll of around 3000 lives and direct economic losses of more than 250 billion Yuan RMB (40 billion US dollars) [16]. Meanwhile, previous investigations and studies (e.g., [17, 18]) suggested that the majority of flood hazards in YRB were directly due to above-normal rainfall that occurred in summer (June–August), owing to the impacts of the East Asian summer monsoon (EASM). Therefore, summer extreme precipitation events in YRB are considered in the present study to analyze their correlations with multiphysical factors. The extreme precipitation is defined to be that summer total precipitation when daily rainfall >95th percentile on the basis of the period 1961–1990 [6].

The observed daily precipitation data from 140 rain gage stations in YRB during the period 1967–2010 are provided by Climate Data Center (CDC) of National Meteorological Information Center, China Meteorological Administration (CMA). The quality control procedures have been conducted by CDC [19]. The daily reanalysis data for SHF, LHF, and SST are provided by the National Centers for Environmental Prediction and National Center for Atmospheric Research (NCEP/NCAR) [20].

2.2. Methodology. In order to combine stations over the whole basin without inducing a bias toward any station or other subregional stations, it is inappropriate to average all station precipitation amount over YRB. The regional precipitation index proposed by Kraus [21] was adopted in this study. Firstly, standardization is carried out for summer precipitation extremes calculated on the basis of extreme definition (>95% percentile) at every station according to the following formula:

\[ Z = \frac{x - \mu}{\sigma}, \]

where \( \mu \) is the mean of summer extreme precipitation and \( \sigma \) is the standard deviation of the time series. All standardized precipitation extremes over the YRB are then averaged in each year, producing a regional summer extreme precipitation index for every year at each station.

Spatial distributions of correlation coefficients between regional summer extreme precipitation index and SHF, LHF, and SST are analyzed on the global ocean scale, and the regions where significance level of the correlation coefficients exceeds 95% are marked by dash areas as shown in Figure 2. It is noteworthy that the regional summer extreme indices in upper, middle, and lower reaches of YRB are also considered to calculate correlation coefficients, respectively. Nevertheless, their spatial distribution displays no significant differences compared to basin regional index (Figures not shown), although the rainfall in three reaches is dominated by different atmospheric circulations. Thus, the regional summer extreme index for the whole basin is utilized in the present study. This is also consistent with the methods applied by Sun et al. [22]. As a consequence, three key regions, Eastern Pacific (EP), Central Equatorial Pacific (CEP), and Northwest Indian Ocean (NWIO), are selected (Figure 3) based on distribution of shaded areas for different time series. The time series of SHF, LHF, and SST during the period of 1967–2010 are then selected in the three key regions, respectively.

Multiple stepwise regression is used to select impact factors and establish statistical prediction model. The focus of stepwise regression would be the question of what the best combination of independent (predictor) variables would be to forecast the dependent (predicted) variable. In a stepwise regression, predictor variables are entered into the regression equation one at a time based upon statistical criteria. At each step in the analysis the predictor variable that contributes the most to the prediction equation in terms of increasing the multiple correlation is entered first. This process is continued only if additional variables add anything statistically to the regression equation. More detailed information about this method can be found in Myers [23]. After the establishment
Figure 1: The location of Yangtze River basin and the location of the meteorological stations.

Figure 2: The spatial distribution of significant correlation coefficients between regional summer extreme precipitation indices in YRB and in-phase LHF, SHF, and SST over the global ocean. Shaded areas denote the values for correlation coefficient significance which exceed the 95% confidence level. (a), (b), and (c) represent LHF, SHF, and SST, respectively.
of statistical prediction model, the method of leave-one-out [24] is utilized to test the effect of the forecast model. For example, one year summer extreme precipitation is taken out in each experiment during 1967–2010, and this year is forecasted by utilizing the rest of the years.

3. The Physical Process of Atmospheric Circulation Variability Associated with the Summer Extreme Precipitation in YRB

Prediction of summer extreme precipitation in YRB was improved slowly in the past decades due to the complexity of the interannual and decadal variability of summer precipitation and diverse impact factors. Selection of physical factors is vitally important for forecasting summer precipitation extremes, while the major factors adopted in the prediction model in many previous studies are far from comprehensive. In this study, many possible influencing factors are considered in the beginning, and then the multiple stepwise regression is conducted to remove the factors that are not very suitable to the model one by one based on Akaike's information criterion (AIC) [25]. Related researches (e.g., [4, 26]) have reported that several circulation processes exhibit significant relationships with summer precipitation extremes in China or YRB, including EASM, Qinghai-Tibet Plateau snow cover, atmospheric circulation in high latitude, El Niño-Southern oscillations (ENSO) cycle, and other dynamic conditions. Changes in these circulations may obviously affect the frequency and intensity of extreme precipitation in YRB.

3.1. LHF, SHF, and SST. Latent and sensible heat fluxes are known to enhance the energy in the lower tropospheric layers which are strongly associated with the formation of precipitation [27]. Pauling et al. [28] and Bichet et al. [29] also suggested that dynamic and thermodynamic effects would trigger significant variations in precipitation extremes in many regions over the world. Dynamic effects are usually linked to large-scale circulation variability or teleconnection patterns, whereas thermodynamic effects are associated with variations in atmospheric humidity and stratification. Meehl et al. [30] indicated that greater thermodynamic effects would increase water vapor, especially in subtropical regions, which was greatly attributed to occurrences of extreme precipitation. Similar conclusions were also investigated in south China and YRB (e.g., [31]). Large latent and sensible heat fluxes over the subtropical North Pacific were conducive to the strengthening of the North Pacific Subtropical High (NPSH), which, in turn, favored the persistence of a quasi-stationary (Mei-yu) front in the YRB after June, and resulted in more heavy rainfall in July and August. Figures 2(a) and 2(b) confirmed that there were some key marine regions where the LHF and SHF had a significant in-phase correlation with the summer extreme precipitation in YRB.

SST anomalies can modulate the variability in global and regional climate as well as extreme precipitation events (e.g., [32]). Diaz et al. [33] indicated that SST might become a dominant contributor to regional extreme precipitation events to some extent. In recent years numerous studies have been conducted to analyze the SST effects on precipitation extremes in China. For example, Yao et al. [34] analyzed the relationship between SST over northwest Pacific Ocean and rainfall in China by utilizing regional air-sea coupled model (RegCM3-POM) and reported that SST and the upper ocean were modulated by the atmosphere through turbulent air-sea heat, momentum, and moisture fluxes, but the modified oceanic state generated further changes in the atmosphere, which would lead to frequent occurrences of heavy precipitation in East Asia. Wang and Yang [35] indicated there was an evident linkage between the summer precipitation in YRB and tropical Pacific SST on the interannual time scales. Furthermore, Figure 2(c) also confirmed that the global SST had significant influence on summer extreme precipitation in YRB. Accordingly, the correlation coefficients between regional summer extreme precipitation index and LHF, SHF, and SST in preceding periods of March–May, February–April, January–March, and the last year’s winter (December-January-February, DJF) were also calculated, respectively (Figures not shown).

Based on the aforementioned analyses, three key Regions of Significant Correlation (RSC), Eastern Pacific (EP), Central Equatorial Pacific (CEP), and Northwest Indian Ocean (NWIO), are selected as shown in Figure 3. The time series for LHF, SHF and SST in each key RSC during period 1967–2010 are selected, respectively. Thereafter, the correlation coefficients between regional summer extreme precipitation index and three selected factors (LHF, SHF, and SST) were calculated. However, only three factors, LHF in winter over CEP, SST during March–May in EP, and SST during June–August in NWIO, displayed significant correlation with significant level exceeding 99% (Table 1). Others were below 95% significant level. The physical causes underlying this phenomenon deserve further study, while selection of impact factors on the basis of distribution of correlation coefficients is the major work for us, and the analyses for deep-seated physical mechanism are beyond the scope of this study. Thus, the above three factors are selected to be the predictors from the perspective of air-sea interaction for statistical prediction model.

3.2. Winter Tibet Plateau Snow. Numerous previous research papers have investigated the possible influences of winter Tibet Plateau (TP) snow on the subsequent summer rainfall and atmospheric circulation over YRB or central China. Wu and Qian [36] reported that there was a significantly positive correlation between winter TP snow and variations of subsequent summer rainfall in YRB. Duan et al. [37] further pointed out that the thermal forcing over TP in winter, which was contributed from the snow cover within the region, played a considerable role in the modulation of the EASM and corresponding precipitation patterns. Zhao et al. [38] analyzed the physical mechanism of interaction between winter TP snow and subsequent summer rainfall in YRB and indicated that the enhancement of winter snow cover over TP tended to increase the local soil moisture content and cool
Table 1: The correlation coefficients and corresponding significant level between regional summer extreme precipitation index (95% percentile) and the initial impact factors selected in the present study.

<table>
<thead>
<tr>
<th>Indices</th>
<th>ISM</th>
<th>AO</th>
<th>NAO</th>
<th>PNA</th>
<th>Winter snow</th>
<th>LHF</th>
<th>SST</th>
<th>SST</th>
<th>ENSO</th>
<th>AAO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Periods/month</td>
<td>August</td>
<td>Nov</td>
<td>June</td>
<td>July</td>
<td>DJF</td>
<td>DJF</td>
<td>March–May</td>
<td>June–August</td>
<td>March</td>
<td>May–June</td>
</tr>
<tr>
<td>R</td>
<td>−0.28</td>
<td>−0.30</td>
<td>−0.29</td>
<td>0.26</td>
<td>−0.48</td>
<td>−0.41</td>
<td>0.38</td>
<td>0.44</td>
<td>0.34</td>
<td>0.41</td>
</tr>
<tr>
<td>P</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.08</td>
<td>0.0008</td>
<td>0.006</td>
<td>0.01</td>
<td>0.002</td>
<td>0.02</td>
<td>0.006</td>
</tr>
</tbody>
</table>

The correlation coefficient between regional extreme index and AO index is −0.3, with 95% significance level. Thus, AO should be considered to be a predictor of the precipitation extremes in YRB.

3.3. Antarctic Oscillation and Arctic Oscillation. Thompson and Wallace [39] demonstrated that Antarctic Oscillation (AAO) was a dominant annual mode of the Southern Hemisphere (SH) extratropical circulation, which represented primarily a large-scale seesaw oscillation of atmospheric circulations in the midhigh latitudes of SH. Many recent studies have noted that this SH circulations also had close linkages with the summer rainfall in YRB, especially for AAO. Fan and Wang [40] analyzed the relationships between AAO and Mei-yu over YRB as well as the frequency of typhoon in western North Pacific and discussed the possible mechanisms for this far-reaching teleconnection. Sun et al. [22] revealed that the positive-phase AAO was concurrently according to the strong convection activities in the region of Maritime Continent through anomalous meridional circulations along central South Pacific and two meridional teleconnection wave train patterns; this anomalous convection propagated northward along with the seasonal cycle and then changed the Western Pacific subtropical High (WPSH) in the subsequent seasons, leading to an impact on the summer precipitation over YRB. The correlation coefficient between regional extreme index and AAO index is 0.41 with significance level exceeding 99% (Table 1). Therefore, AAO should be considered to be a predictor for the precipitation extremes in YRB.

Compared to AAO, Arctic Oscillation (AO) is a leading mode of the extratropical Northern Hemisphere (NH) circulation. Thompson and Wallace [41] indicated that many climate system components were significantly affected by winter AO, and consequently AO exerted significantly wide-ranging effects on extreme weather/climate events over the NH during November–April. Gong et al. [42] suggested that AO triggered significant influences on the large-scale EASM rainfall, including the variations in summer precipitation in YRB. Ju et al. [43] demonstrated that the high-index polarity of AO during the last two decades played an important role in the interdecadal decrease of land-sea contrast anomalies in NH, which finally resulted in the weakening of EASM circulation; accordingly, the Mei-yu would stay longer over YRB. Gong et al. [44] further revealed that AO could impact the EASM via generating tropical air-sea feedback over the western North Pacific on the basis of observation evidence and numerical simulations. The correlation coefficient between regional extreme index and AO index is −0.3, with the 95% significance level. Thus, AO is also selected as a predictor of the precipitation extremes in YRB.

3.4. ENSO, Indian Summer Monsoon, North Atlantic Oscillation, and Pacific North American Pattern. Although the SST in key RSCs was selected in the above analyses, the ENSO is one important factor for the EASM variability. Besides, the multiple stepwise regression adopted to select final factors

![Figure 3: The defined key regions that have significantly affected summer extreme precipitation in YRB according to correlation coefficients over the global ocean. The solid line indicates three key regions selected in the present study. EP: Eastern Pacific; CEP: Central Equatorial Pacific; NWIO: Northwest Indian Ocean. Dashed boxes represent insignificant impact.](image-url)
for prediction model in the present study can automatically examine correlation among different impact factors. Thus, ENSO is also considered to analyze the correlation with summer rainfall in YRB. Wu et al. [45] suggested that the development and migration of rain-belt over East China were associated with the ENSO-related tropical heating anomalies over western Pacific and South Asia, although the impacts of ENSO on the EASM were discrepancies along with different stages of the ENSO cycle. Tong et al. [46] further indicated that ENSO episodes were in good teleconnection with floods in YRB and significant correlation appeared at 5.04- or 10–12-year periods. The possible physical mechanism might be that ENSO influenced the EASM through strengthening the subtropical high over western Pacific region. Table 1 demonstrates that the correlation coefficient between regional extreme index and ENSO index during 1967–2010 is 0.34, with the significant level being at 98%.

Indian Summer Monsoon (ISM) blows from sea to land after crossing the Indian Ocean, Arabian Sea, and Bay of Bengal and brings abundant precipitation to YRB and south China during a year. One of the three main water vapor transports to China, the southwesterly flow which is also known as the Somali jet, towards the Indian Peninsula and Bay of Bengal associated with the ISM, then arrives in south or even north China. Furthermore, the strong ISM can strengthen moisture transport by enhancing Somali cross-equatorial flow and bring more rainfall for China. The correlation between ISM and regional extreme index is also significant (Table 1). Additionally, the relationships between summer precipitation and North Atlantic Oscillation (NAO) and Pacific North American Pattern (PNA) were also investigated in many studies (e.g., [47, 48]), while the corresponding correlation coefficients in the same way in this study are shown in Table 1. Several indices adopted in the analyses above are available at http://www.cpc.ncep.noaa.gov/. The time scales of selected indices are from 1967 to 2010 expect for TP with a period of 1966–2009.

4. Selection of Predictors and Validation of the Forecast Model

From the perspective of physical mechanisms, many possible impact factors are identified, including direct and far-reaching teleconnection physical factors, respectively (Table 1). Because the differences among the magnitudes of selected indices are relatively remarkable, the indices are normalized before the multiple stepwise regression is conducted to improve forecasting accuracy of statistical prediction model. In addition, the periods or time scales for the selected factors are also shown in Table 1. In the process of stepwise regression, the optimal subset regression is applied and the smallest statistic of AIC is selected as a criterion. The PNA was taken out at first step; nevertheless, after taking NAO out in the second choice, the value of AIC would increase whether any predictor was removed. Therefore, the physically based statistical prediction model was obtained via multiple stepwise regression analyses:

\[
Y = -0.43 - 0.20x_1 + 0.12x_2 + 0.34x_3 - 0.20x_4 \\
+ 0.31x_5 + 0.18x_6 - 0.21x_7 + 0.16x_8, \tag{2}
\]

where \( Y \) denotes regional summer extreme precipitation index, \( x_1 \) denotes ISM, \( x_2 \) denotes AO, \( x_3 \) denotes winter TP snow, \( x_4 \) denotes LHF in winter, \( x_5 \) denotes SST during March–May over EP, \( x_6 \) denotes SST during June–August over NWIO, \( x_7 \) denotes ENSO, and \( x_8 \) denotes AAO. The significant levels of coefficients for the statistical prediction model (2) were all exceeding 95%; besides, the coefficient of TP was exceeding 99%.

In order to make a validation of the prediction model, the method of leave-one-out was utilized during 1967–2010. A good agreement between the interannual variations of forecasting and observed regional summer extreme precipitation indices is illustrated in Figure 4. The correlation coefficient between the observed and modeled extreme precipitation is 0.85, with significant level being at 99%. Figure 4 also demonstrates that the statistical forecast model can successfully reconstruct the large magnitudes of extreme precipitation years in 1969, 1980, 1991, 1998, and 2010, with the forecasting values being close to the observed values. Accordingly, the model can also successfully reproduce the smaller magnitudes of extreme precipitation years in 1971, 1978, 1992, and 2006 during the period of 1967–2010. Therefore, the statistical prediction model can reasonably forecast the mainly larger and smaller magnitudes of summer extreme precipitation over YRB.

5. Conclusions

The Yangtze River basin (YRB) is highly populated, accounting for nearly one-third of the national population and a significant share of the gross domestic product in China, coupled with comprehensive social infrastructures. It is therefore of practical importance to accurately forecast variations
in precipitation extremes over YRB. Although prediction of extreme precipitation cannot reduce the occurrences of floods in YRB, the protective measures can be prepared via predicting in advance, thereby reducing economic losses and casualties.

In the present study, from the viewpoint of air-sea interaction, prediction of summer extreme precipitation by utilizing the multiple stepwise regression is proposed. The results of validation during 1967–2010 using the method of leave-one-out show that the statistical prediction model can capture the major characteristics of the interannual variations in summer precipitation extremes in YRB. It can successfully reconstruct larger and smaller magnitudes of extreme precipitation, respectively. This work also quantifies the time series indices of LHF, SHF, and SST as well as winter TP snow in key regions by analyzing the spatial distribution of correlation coefficients with regional summer extreme precipitation over YRB. Therefore, it provides a practical approach to forecast extreme precipitation based on rain gauge station dataset from a new perspective.

In addition, the year-to-year variability of summer extreme precipitation in YRB is influenced by many factors, while the selected factors that are favorable to the variations in summer extreme precipitation over YRB in this study are far from comprehensive; for example, the latest research shows that the ozone depletion over the South Pole has affected the extreme daily precipitation in the austral summer [49]. Thus, more decisive predictors should be added into the statistical prediction model with further study. However, our efforts indicate that the summer extreme precipitation in YRB can be predictable although a lot of work should be done in the future, and this meaningful exploration is very important to forecast the variations in regional extreme precipitation.

Conflict of Interests
The authors declare that they have no conflict of interests.

Acknowledgments
This study is supported by the Taishan Scholar Fund (no. 3000-841112013) awarded by Shandong Province and National Natural Science Foundation of China (no. 41210008). The authors thank He Meng at Ocean University of China for providing the useful help. They are also grateful to NCEP/NCAR for making the relevant data available.

References


