Research Article

The Prediction Algorithm and Characteristics Analysis of Kuroshio Sea Surface Temperature Anomalies

Dawei Shi,[1,2] Chao Li [ID],[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

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 $1.0° \times 1.0°$ [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 $1° \times 1°$.

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).

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).
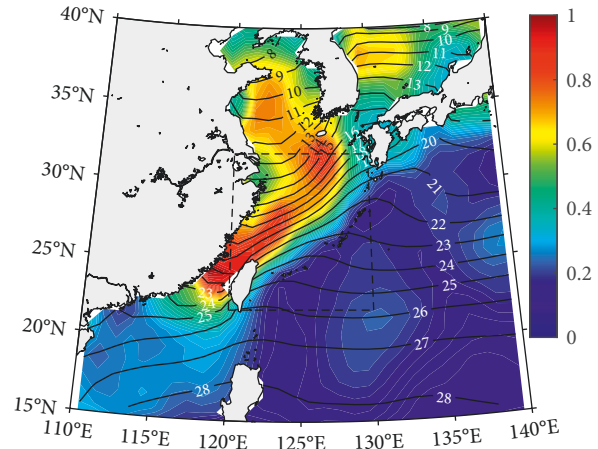

Figure 1: Distribution of climatic mean (isoline, unit °C) and mean-square error (colored area) of Spring SST.

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.

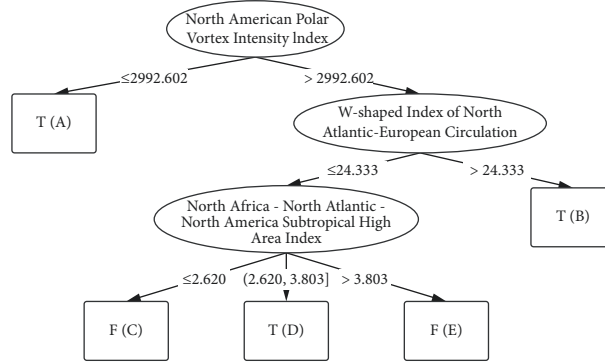|	Total sample (1951–2015)	Training set sample (1951–2004)	Test set sample (2005–2015)
High SST	11	7	4
No high SST	44	37	7

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.
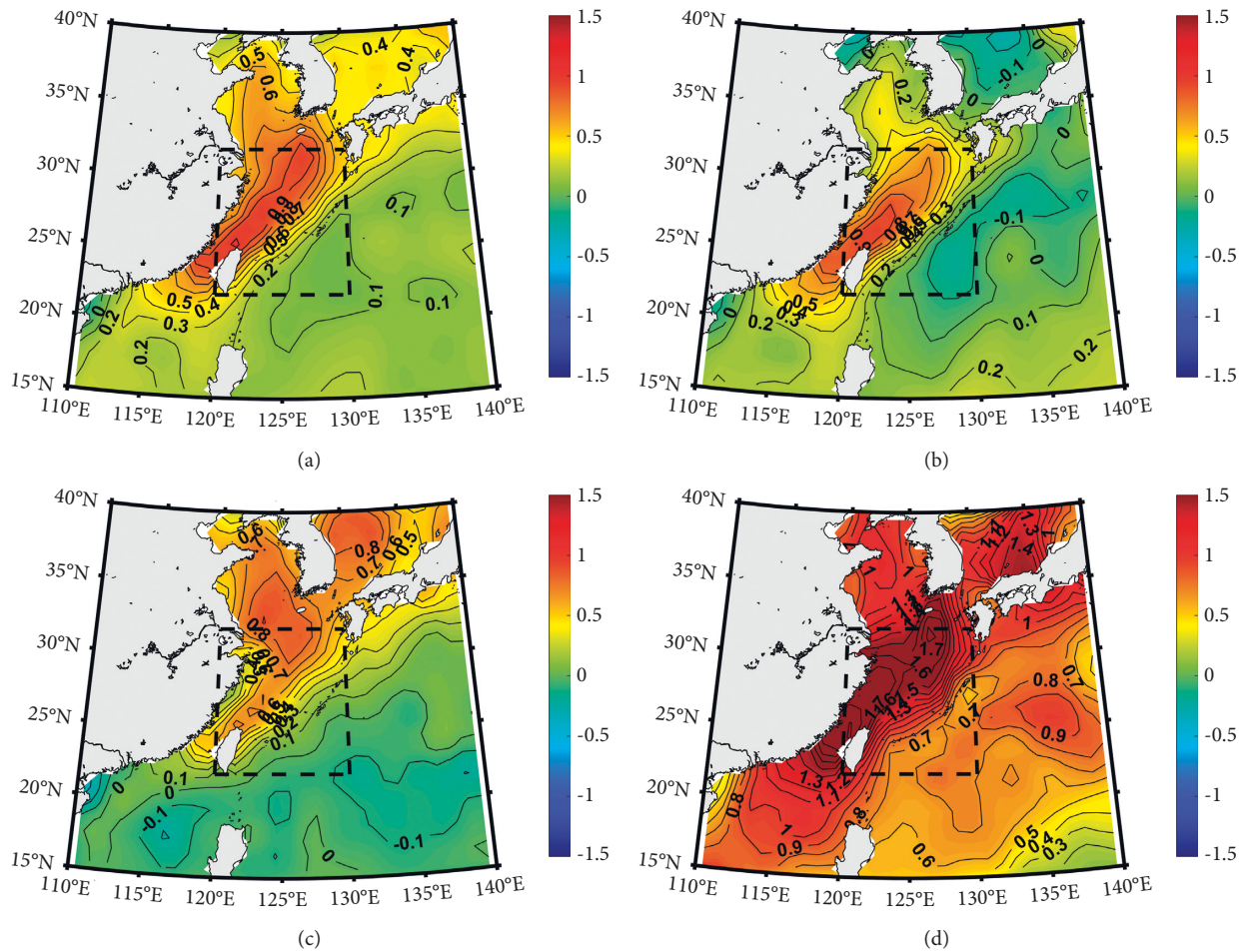
FIGURE 3: Distribution of Spring SST anomalies in Kuroshio Area. (a) Strong Annual Climate Mean; (b) Type A Anomaly; (c) Type B Anomaly; (d) Type D Anomaly.

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
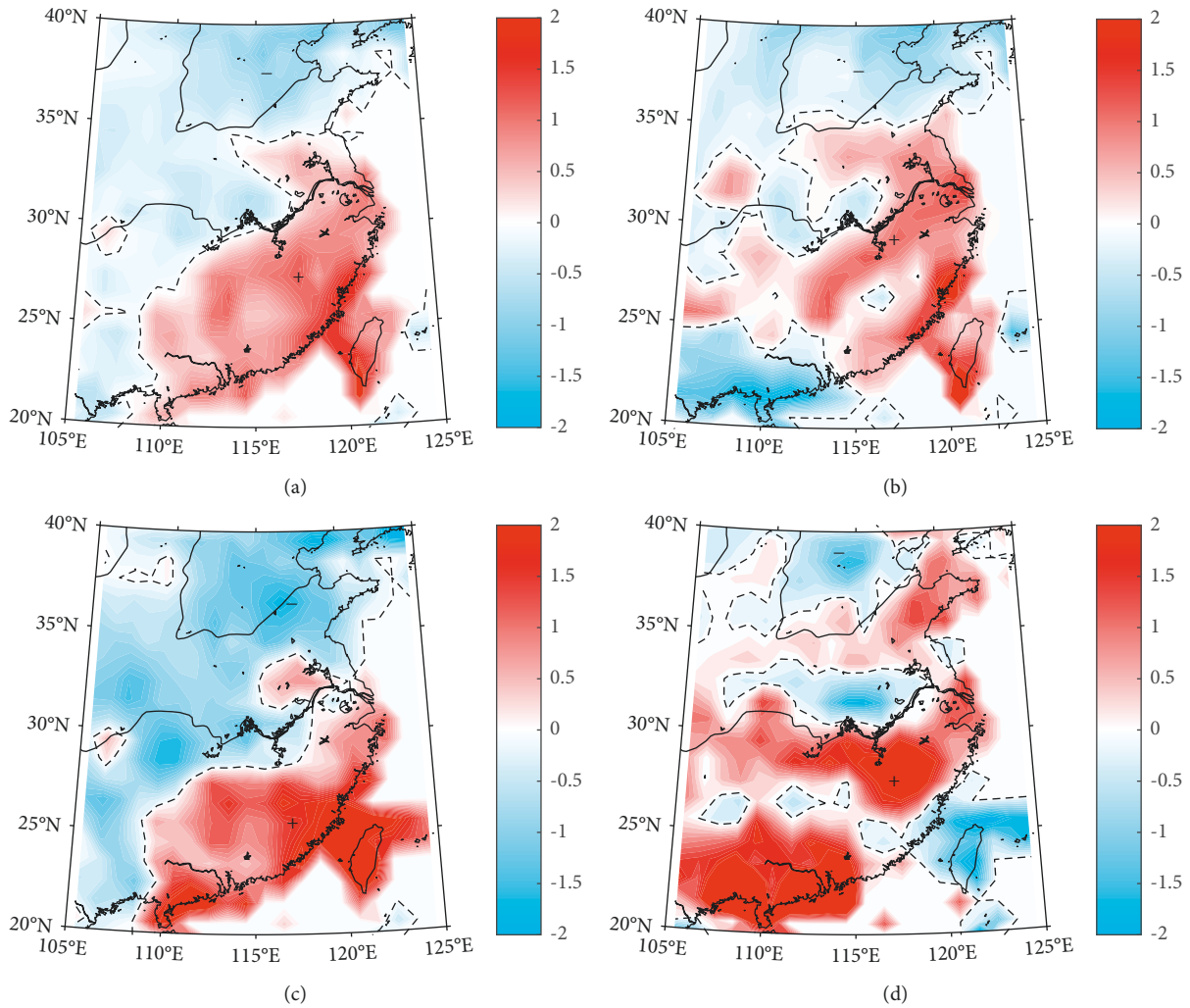
FIGURE 4: Distribution of summer daily precipitation anomalies in the Spring Kuroshio SST anomaly years. (a) Strong Annual Climate Mean; (b) Type A Anomaly. (c) Type B Anomaly; (d) Type D Anomaly.

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

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 multiyear 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.