

## Article

# Modeling of Precipitation Prediction Based on Causal Analysis and Machine Learning

Hongchen Li<sup>1</sup> and Ming Li<sup>2,\*</sup>

<sup>1</sup> College of Meteorology and Oceanography, National University of Defense Technology, Changsha 410000, China; lihongchen17@nudt.edu.cn

<sup>2</sup> College of Frontier Interdisciplinary, National University of Defense Technology, Changsha 410000, China

\* Correspondence: liming17c@nudt.edu.cn

**Abstract:** The factors influencing precipitation in western China are quite complex, which increases the difficulty in determining accurate predictors. Hence, this paper models the monthly measured precipitation data from 240 meteorological stations in mainland China and the precipitation data from the European Centre for Medium-Range Weather Forecasts and the National Climate Centre and employs 88 atmospheric circulation indices to develop a precipitation prediction scheme. Specifically, a high-quality grid-point field is created by fusing and revising the precipitation data from multiple sources. This field is combined with the Empirical Orthogonal Function decomposition and the causal information flow. Next, the best predictors are screened through Empirical Orthogonal Function decomposition and causal information flow, and a data-driven precipitation prediction model is established using a Back Propagation Neural Network and a Random Forest algorithm to conduct the 1-month, 3-month, and 6-month precipitation predictions. The results show that: The machine learning-based precipitation prediction model has high accuracy and is generally able to predict the precipitation trend in the western region better. The Random Forest algorithm significantly outperforms the Back Propagation Neural Network algorithm in the prediction of the three starting times, and the prediction ability of both models gradually decreases as the starting time increases. Compared with the 2022 flood season prediction scores of the Institute of Atmospheric Sciences of the Chinese Academy of Sciences, the model improves the prediction of 1-month and 3-month precipitation in the western region and provides a new idea for the short-term climate prediction of precipitation in western China.

**Keywords:** multi-source precipitation data; causal analysis; Random Forest; Back Propagation Neural Network; Empirical Orthogonal Function decomposition



**Citation:** Li, H.; Li, M. Modeling of Precipitation Prediction Based on Causal Analysis and Machine Learning. *Atmosphere* **2023**, *14*, 1396. <https://doi.org/10.3390/atmos14091396>

Academic Editors: Stephan Havemann and Stefano Federico

Received: 1 August 2023

Revised: 28 August 2023

Accepted: 31 August 2023

Published: 4 September 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The western region of China, the main part of which is located west of 110°E, is far from the sea and has unique natural geographic conditions, characterized by interlocking plateaus and mountains, coexisting basins, deserts, and lakes, and nurturing glaciers, permafrost, and lakes within its territory, and it is the source of the largest rivers in Asia and Europe [1]. The western region has complex climate types with obvious regional characteristics, within which the northwestern region suffers from perennial water shortages and frequent drought and desertification disasters; the southwestern region has many mountainous areas, which are prone to flash floods and mudslides during extensive rainfall in the summer [2]. In recent years, with global warming and the enhancement of human activities, the natural environment in the west has undergone significant changes, such as the rise of the snow line in the plateau region, the retreat of glaciers, the decrease in the area of permafrost, the reduction in groundwater resources, and the expansion of land desertification, which have triggered a series of ecological and environmental problems [3]. In particular, water resource security and ecological environment security problems caused by

climate change in western China will constrain China's economic and social development. Therefore, it is of great practical significance to further improve the level of precipitation prediction in western China.

At present, the forecasting of the precipitation seasonal trends is mainly based on methods that combine statistics, dynamics, and power statistics. Statistical methods predict the precipitation trend in future seasons by analyzing historical contemporaneous data and establishing mathematical models. Commonly used statistical methods include regression and time series models. Tang et al. [4] relied on wavelet regression to build a statistical model, tested it in China, and successfully obtained the general trend of seasonal changes in precipitation. Mai et al. [5] explored the probabilistic forecasting method of short-term heavy precipitation in Chengdu city based on the idea of the "ingredient method" combined with regression analysis. With the development of numerical models, the dynamical models have gradually become the main tool for climate prediction. Such models simulate the atmospheric circulation system through the atmospheric circulation model and then project the possible precipitation changes. In recent years, the ability of such seasonal prediction models to predict atmospheric circulation, ENSO phenomena, and Asian summer winds has been significantly improved. However, their ability to predict precipitation in East Asia remains limited [6]. Therefore, many experts and scholars have developed a combined dynamical and statistical prediction model based on a dynamical model, which further improves the precipitation prediction accuracy by considering the historical precipitation characteristics (e.g., SST anomalies and ENSO), the changes in climatic indices (e.g., Southern Oscillation Index and North Atlantic Oscillation Index), and the atmospheric-oceanic physical mechanisms of a particular region. These characteristics are combined with real-time monitoring data and expert empirical judgment to output precipitation prediction. For instance, Sun et al. [7] investigated a seasonal precipitation prediction method based on wavelet analysis and Kriging interpolation. Bukhari et al. [8] investigated a seasonal precipitation prediction method based on regression analysis and a Dynamical-statistical model. Alireza et al. [9] provide a novel methodology for modeling multivariate dependence structures of meteorological drought characteristics (severity, duration, peak, and interarrival time) based on the combination of four-dimensional Vine Copulas and Data Mining algorithms. Pan et al. [10] established a revised method for ECMWF precipitation forecasting based on Kalman dynamic frequency. Chen et al. [11] explored the effect of different convective-scale ensemble forecast memberships on precipitation forecasts. All of these studies have improved precipitation prediction techniques to some extent.

With the development of big data and artificial intelligence technology, Machine Learning (ML) shows great application potential in weather forecasting. Distinguishing itself from traditional statistical methods, machine learning excels in dealing with nonlinear problems and can be utilized to discover and extract new interrelated signals from the Earth system. AI-based forecasting of weather elements is typical of data-driven modeling, which emphasizes learning rules from historical data and reasoning and forecasting about new data. Commonly used ML algorithms in numerical weather prediction are the Back Propagation Neural Network (BPNN) [12,13], Support Vector Machine (SVM) [14,15], Random Forest (RF) [15,16], and Bayesian Network (BN) [17,18]. In recent years, with the development of deep learning theory, deep learning methods represented by Convolutional Neural Networks (CNN) [19,20] and Long Short Term Memory (LSTM) [13,21] have been applied in the field of the climate. These machine learning algorithms provide new ideas for the development of the field of weather forecasting, such as weather forecast accuracy improvement, weather phenomenon identification, causality mining, and extreme weather detection.

The most important aspect of machine learning modeling is the feature selection, which involves selecting the forecast factors in the precipitation prediction. Using irrelevant variables as the input variables in the model will not only increase the computational effort of the prediction system, but also reduce the accuracy of the forecast [22]. The

most commonly used method in previous studies relating to forecast factor selection was correlation analysis, which selects the best factor by analyzing the correlation between the output variables and the forecast factors. Although correlation analysis is widely used, it does not provide an in-depth understanding of the causal mechanisms behind the system dynamics, which is extremely important in meteorology and oceanography. Causal analysis can uncover hidden relationships in the system, thus overcoming some of the key drawbacks of correlation analysis [23]. For example, through causal inference, Marlene et al. [24] found that sea ice concentration in the Barents-Kara Sea is an important driver of the mid-latitude circulation affecting the Arctic Oscillation in winter. McGraw et al. [25] examined the causal relationship between Arctic sea ice and atmospheric circulation using Granger causal analysis. Li et al. [26] developed a new data-driven prediction technique for sea ice density by screening the key environmental factors affecting the sea ice density through causal analysis based on the Granger causality test. Liang et al. [27] demonstrated that the South China Sea significantly influences the Pacific-North America remote sensing model, using the causal inference of information flow. Docquier et al. [28] used causal information flow to verify that air and sea surface temperatures and ocean heat transport are the main drivers of the recent and future changes in the Arctic sea ice. These studies demonstrate that causal analysis has a stronger ability to uncover hidden relationships than correlation analysis and that it is a promising method for screening forecast factors. At present, the study of applying causal analysis to the screening of precipitation predictors in the western region of China has not been reported.

A lot of research has been conducted on improving precipitation prediction techniques; however, the existing research on precipitation prediction in the western region is still very limited, and most of the studies are focused on a certain region in the western region. Although there is a lack of research on precipitation prediction in the whole western region, we can still gain a lot of insights from the previous studies. Guo et al. [1] applied the Empirical Orthogonal Function (EOF) and the Rotated Empirical Orthogonal Function (REOF) to classify the western region of China into nine precipitation types in order to study the spatial and temporal variability characteristics of precipitation in the western region. This study shows that the EOF can decompose the meteorological element field, which changes with time, into a spatial function part and a temporal function part, and condense the change information of the original element field into the first few principal components and their corresponding spatial functions, which provides powerful help in studying the complex climate types in the western region. Luo et al. [29] found that the westerly trough is an important influence on summer precipitation in southwestern China, and Zhu et al. [30] found that the characteristics of the westerly rapids, the structure of the North Pacific Sea Level Pressure, the characteristics of the North Atlantic Oscillation, and the strength of the Mascarene High Pressure are the four contemporaneous key influences affecting precipitation in northwestern China. Both of these studies revealed that the atmospheric circulation index is an important factor influencing precipitation in the western region.

The spatial and temporal distribution of precipitation in western China is uneven and the influencing factors are complex, and there is a big gap between the level of numerical model forecasting and the operational needs. Therefore, it is necessary to utilize the advantages of the EOF, causal analysis, and machine learning to further improve the precipitation forecasting capability in western China. Combined with the previous studies, this paper condenses the spatial and temporal variation information of the multi-source precipitation field on the first few principal component variables through EOF decomposition. In addition, in order to solve the problem of the precipitation influencing factors in the western region being complicated and the best forecast factor being difficult to select, the best forecast factor with a significant causal relationship with each principal component variable is screened from the set of 88 atmospheric circulation indices based on the causal information flow. Finally, considering that some years of the precipitation observation in the western region are not suitable for deep learning methods, this paper

uses RF and BPNN modeling, combined with the actual precipitation data, to establish a precipitation prediction method suitable for the western region.

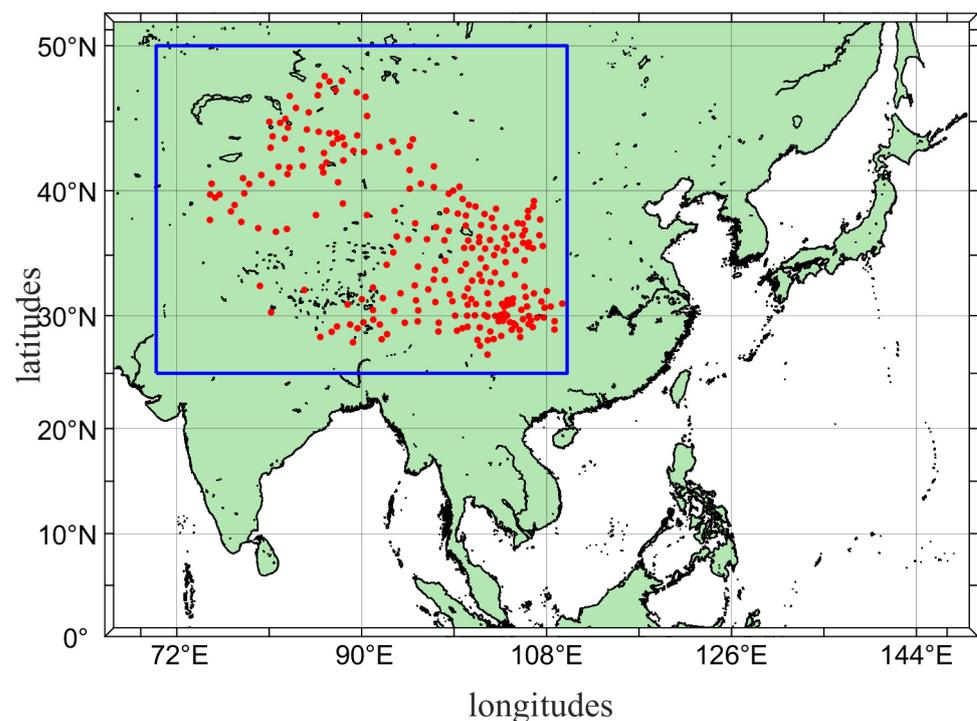
## 2. Information and Methodology

This section specifically describes the sources of data and spatial and temporal resolution, the metrics (Root Mean Square Error and Anomaly Correlation Coefficients) used to assess the effectiveness of precipitation forecasts, the causal analysis methods (causal information flow) used to filter the best forecast factors, and the ML methods (BPNN and RF) used for modeling precipitation forecasts.

### 2.1. Data Sources

The forecast factor information is derived from the monthly average of 88 atmospheric circulation indices in the Climate System Monitoring Index Set provided by the National Climate Centre (NCC), which mainly include atmospheric circulation indices such as the sub-height, the East Asian trough, the polar vortex, the Eurasian circulation type, the remote correlation, and the Pacific trade winds. The time scale is January 1951–February 2023, and in the case of missing measurements, the moving median of the same month in each of the 15 years before and after the missing month is used to supplement it.

This paper uses the climate model prediction data from the European Centre for Medium-Range Weather Forecasts (ECMWF) and the National Climate Center (NCC), both of which are from the MODES system, with a spatial range of 25° N–50° N, 70° E–140° E and a spatial resolution of 1° × 1°. The ECMWF model data are from February 1993 to September 2022, and the NCC model data are from February 1991 to December 2022. Both sets of model data have a time step of 1 month. Meanwhile, we also employ the monthly total precipitation data from 240 stations in the western region (see Figure 1) for January 1985 to September 2021.



**Figure 1.** Map of 240 meteorological stations in the western region, where the blue-boxed area (25° N–50° N, 70° E–140° E) is the scope of the western region studied in this paper, and the red dots represent weather stations.

As the three data and the atmospheric circulation index set have different start and cutoff times, in order to ensure the same length of time for each data, only the data from

February 1993–September 2022 are selected in this paper to carry out the precipitation prediction experiments in western China.

### 2.2. Assessment Methodology

The Root Mean Square Error (RMSE) and Anomalous Correlation Coefficient (ACC) are used for the precipitation evaluation metrics in the western region. RMSE is a commonly used statistical indicator for assessing the difference between forecast models or estimates and observations. It measures the size of the average deviation of the model predictions from the observations and is sensitive to outliers. The RMSE formula is as follows:

$$RMSE = \left[ \frac{1}{n} \sum_{i=1}^N (\hat{y}_i - y_i)^2 \right]^{\frac{1}{2}} \tag{1}$$

where  $\hat{y}_i$  is the predicted temperature,  $y_i$  is the true temperature (the output variable of the test set),  $\bar{y}_i$  is the mean value of the real temperature, and  $n$  represents the total number of samples.

ACC is a statistical measure of the similarity between two anomaly fields, which is commonly used to assess the ability of numerical weather prediction models and climate model simulations to quantify whether a model or forecast can capture and reproduce the observed anomalies. The ACC has a subjective range of  $-1$  to  $1$ , where  $1$  indicates a perfect match between the model and the observations,  $0$  means no correlation, and  $-1$  means a perfect negative correlation. Higher ACC values indicate better agreement between the model and observations regarding anomaly pattern and magnitude. The formula for calculating ACC is as follows:

$$A_{cc} = \frac{\sum_{i=1}^n (y_i - \bar{y})(o_i - \bar{o})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (o_i - \bar{o})^2}} \tag{2}$$

where  $n$  is the total sample size,  $y_i$  and  $o_i$  denote the predicted and observed values, and  $\bar{y}$  and  $\bar{o}$  denote the mean of the predicted and observed values.

### 2.3. Basic Theory of Information Flow

Information flow is a physical quantity independently proposed by Professor Xiangsan, which analyzes the transfer entropy and utilizes the Granger causality test to quantitatively characterize the causal relationship between variables (or events) and measure the causality through the time rate of information from one sequence of variables to the other. This causality is unidirectional. The amount of information exchanged between the two variables indicates the magnitude of the causal relationship and the direction [31]. Measuring the causality between variables through information flow or information transfer enables causal analysis to be formulated and quantified.

Consider a two-dimensional stochastic system where the original expression for the information flow can be written as:

$$dX = F(X, t)dt + B(X, t)dW \tag{3}$$

where  $F = (F_1, F_2)^T$  is the vector of drift coefficients (micro-vectorizable field),  $B = (b_{ij})$  is the diffusion coefficient matrix, and  $W$  is the standard Wiener process vector. Let  $g_{ij} = \sum_k b_{ik}b_{jk}$ . Then, the information flow from variable  $X_2$  to  $X_1$  is:

$$T_{2 \rightarrow 1} = -E \left( \frac{1}{\rho_1} \frac{\partial F_1 \rho_1}{\partial x_1} \right) + \frac{1}{2} E \left( \frac{1}{\rho_1} \frac{\partial^2 g_{11} \rho_1}{\partial x_1^2} \right) \tag{4}$$

where  $\rho_1$  is the marginal density function of  $X_1$  and  $E$  is the mathematical expectation. Formally, the information flow from variable  $X_2$  to  $X_1$  equals the difference between the

marginal entropy change rate of  $X_1$  and the marginal entropy change rate of the system, excluding  $X_2$ .

On this basis, Xiangsan proved that the maximum likelihood estimation of the information flow in Equation (3) is very tight and involves only common statistics, i.e., sample covariance. Thus, for a sequence of variables  $X_1$  and  $X_2$ , the information flow rate (in units of transfer per unit of time) from  $X_2$  to  $X_1$  is:

$$T_{2 \rightarrow 1} = \frac{C_{11}C_{12}C_{2,d1} - C_{12}^2C_{2,d1}}{C_{11}^2C_{22} - C_{11}C_{12}^2} \tag{5}$$

where  $C_{ij}$  is the sample covariance between  $X_i$  and  $X_j$ , and  $C_{i,dj}$  is the covariance between  $X_i$  and  $\dot{X}_j$ .  $\dot{X}_j$  is a differential approximation of  $dX_j/dt$  in the Eulerian forward format  $\dot{X}_j(n) = [X_j(n+k) - X_j(n)]/k\Delta t$ . Typically,  $k$  equals 1, but for highly chaotic and densely sampled sequences  $k = 2$  should be chosen to avoid a falsely large  $\dot{X}_j$ . The information flow in the opposite direction can be obtained by exchanging the positions of subscripts 1 and 2.

To compare the causality strength, Xiangsan normalized the information flow formula [32]:

$$\begin{cases} Z_{2 \rightarrow 1} \equiv |T_{2 \rightarrow 1}| + \left| \frac{dH_1^*}{dt} \right| + \left| \frac{dH_1^{noise}}{dt} \right| \\ \frac{dH_1^*}{dt} = E \left( \frac{\partial F_1}{\partial x_1} \right) \\ \frac{dH_1^{noise}}{dt} = -\frac{1}{2}E \left( g_{11} \frac{\partial^2 \log \rho_1}{\partial x_1^2} \right) - \frac{1}{2}E \left( \frac{1}{\rho_1} \frac{\partial^2 g_{11} \rho_1}{\partial x_1^2} \right) \end{cases} \tag{6}$$

where  $H_1^*$  denotes the phase space expansion along the  $X_1$  direction, and  $H_1^{noise}$  denotes the random effect. Normalizing  $Z$  provides the normalized information flow equation:

$$\tau_{2 \rightarrow 1} = \frac{T_{2 \rightarrow 1}}{Z_{2 \rightarrow 1}} \tag{7}$$

The normalized information flow  $\tau_{2 \rightarrow 1}$  computed through Equation (7) can be either zero or non-zero. If  $\tau_{2 \rightarrow 1} = 0$ ,  $X_2$  does not cause  $X_1$ , i.e., there is no causal relationship between the two. If  $\tau_{2 \rightarrow 1} \neq 0$ , a causal relationship between the two exists. At this point, we have two cases according to the sign:  $\tau_{2 \rightarrow 1} > 0$ , indicating that  $X_2$  causes  $X_1$  to tend to uncertainty, i.e.,  $X_2$  is the cause of  $X_1$ , and  $\tau_{2 \rightarrow 1} < 0$ , indicating that  $X_2$  causes  $X_1$  to tend to stability, i.e.,  $X_2$  can not be the cause of  $X_1$ . In particular, at a significance level of 0.1,  $\tau_{2 \rightarrow 1} > 1\%$  determines that the causal relationship is significant.

#### 2.4. Machine Learning Methods

In order to reduce the impact of the randomness of machine learning algorithms in the modeling process, this paper uses two different machine learning algorithms for modeling, namely Back Propagation Neural Network and Random Forests. These methods belong to machine learning classification and regression methods, with some differences in identifying and fitting the data.

##### (1) Back Propagation Neural Network

In 1986, Rumelhart and McClelland introduced multilayer networks in Parallel Distributed Processing and first proposed the Error Back Propagation (BP) algorithm. BP is the most widely used algorithm for learning neural networks, typically employed for multilayer feedforward neural grids and other types of neural grids, such as training recurrent neural grids. The BP neural network usually refers to a multilayer feedforward neural network trained with the BP algorithm, which contains one input layer, one or more

hidden layers, and one output layer. Each layer can contain multiple neurons and has good adaptive, error-resistant, and associative memory functions. The BP model is expressed as:

$$P_k = g_2 \left[ \sum_{j=1}^m w_{kj} g_1 \left( \sum_{i=1}^n w_{ji} x_i + w_{j0} \right) + w_{k0} \right] \quad (8)$$

where  $x_i$  is the input value of node  $i$ ,  $P_k$  is the output value of node  $k$ ,  $g_1$  is the implicit layer activation function, and  $g_2$  is the output layer activation function.  $m$  and  $n$  are the number of neurons in the input layer and output layer, respectively,  $w_{j0}$  is the deviation of the  $j$ th neuron in the hidden layer, and  $w_{k0}$  is the deviation of the  $k$ th neuron in the output layer.  $w_{kj}$  is the weight of the output node  $k$  with implicit node  $j$ , and  $w_{ji}$  is the weight of input node  $i$  with implied node  $j$ .

## (2) Random Forest

Random Forest uses Decision Tree (DT) as the base learning machine, which is a classifier that uses multiple decision trees for integration and trains and predicts the samples. Thus, the predicted value of the RF is obtained by averaging the prediction of all decision trees. Additionally, RF efficiently handles input samples with high-dimensional feature space and complex data structures with strong robustness.

## 3. Experiments on Precipitation Prediction Based on Causal Analysis and Machine Learning Methods

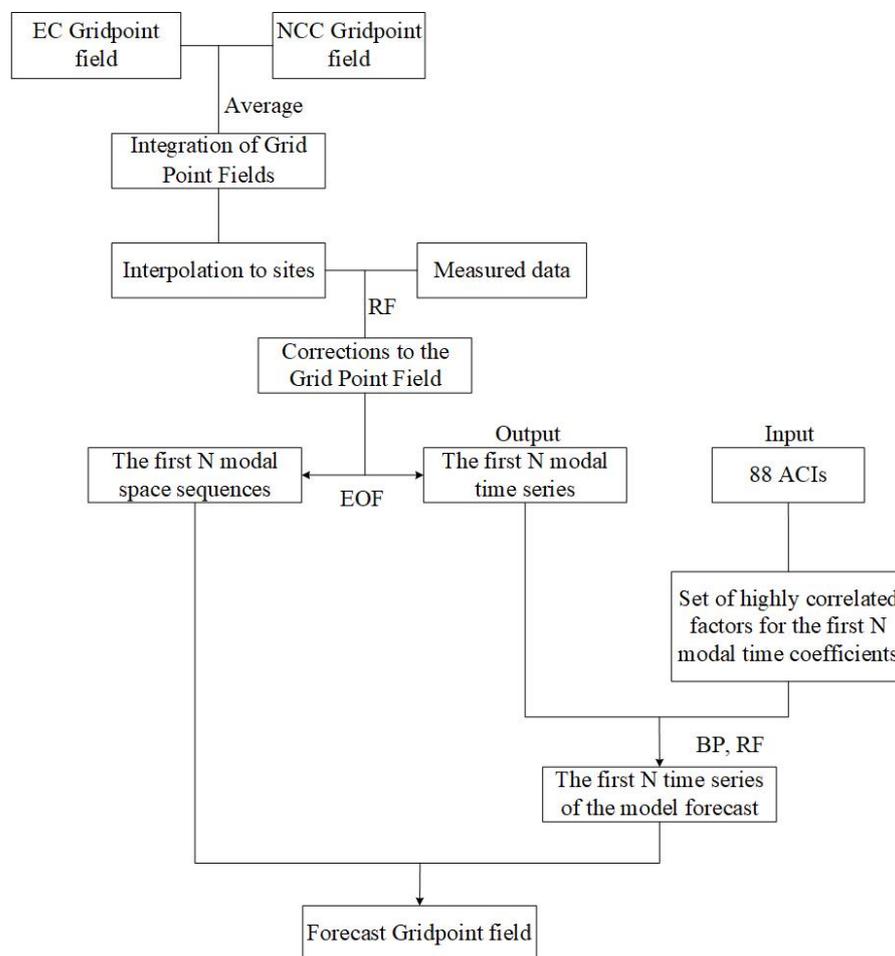
Based on the specific methodology presented in the previous section, this section begins with a data fusion revision of the multi-source precipitation data to obtain a high-quality grid point field. Subsequently, the best predictors are screened by combining EOF decomposition and causal information flow, and the precipitation prediction model is established using the BPNN and RF algorithms to carry out the experiments of precipitation prediction for 1 month, 3 months, and 6 months, respectively.

### 3.1. Technical Process for Modeling Precipitation Forecasts

This section elaborates on the data-driven prediction model based on causal information flow and machine learning algorithms. Figure 2 illustrates the model's architecture, which is divided into four aspects: lattice field revisions, factor selection, model building, and model prediction and effect testing.

- (1) Grid point field revision: This part takes the mean value of the EC and NCC grid point field to form a new grid point field, which is used to interpolate the grid point data into station data. After interpolation, the precipitation data of 240 stations are used as inputs, the measured data of 240 stations are used as outputs, and the RF algorithm is used to train the revised model. Based on the above model, the whole grid-point field is input to revise the original fusion grid-point field.
- (2) Factor selection: The revised grid point field is stored in a  $26 \times 41 \times 344$  standard format, where the first two dimensions are spatial, representing the latitude and longitude of the data, and the third dimension is temporal. Then, the revised grid point field is decomposed by the EOF into time series and spatial series, and the corresponding modes are selected according to the variance contribution of each mode. Each selected precipitation mode is subjected to information flow-based causal analysis using the set of atmospheric circulation indices to obtain a set of highly correlated factors for the time coefficients of the first  $N$  modes of precipitation.
- (3) Model building: This part sets the initial parameters according to the characteristics of THE BPNN and RF and combines the screened key factors with the decomposed time series for modeling.
- (4) Model prediction and effect test: The time coefficients of the model prediction and the spatial coefficients decomposed by the EOF are reduced to the prediction grid field.

The prediction time limits are 1, 3, and 6 months, and the RMSE and ACC indicators evaluate the precipitation prediction ability of each station.



**Figure 2.** Flow chart for precipitation prediction.

### 3.2. Precipitation Prediction Experiments

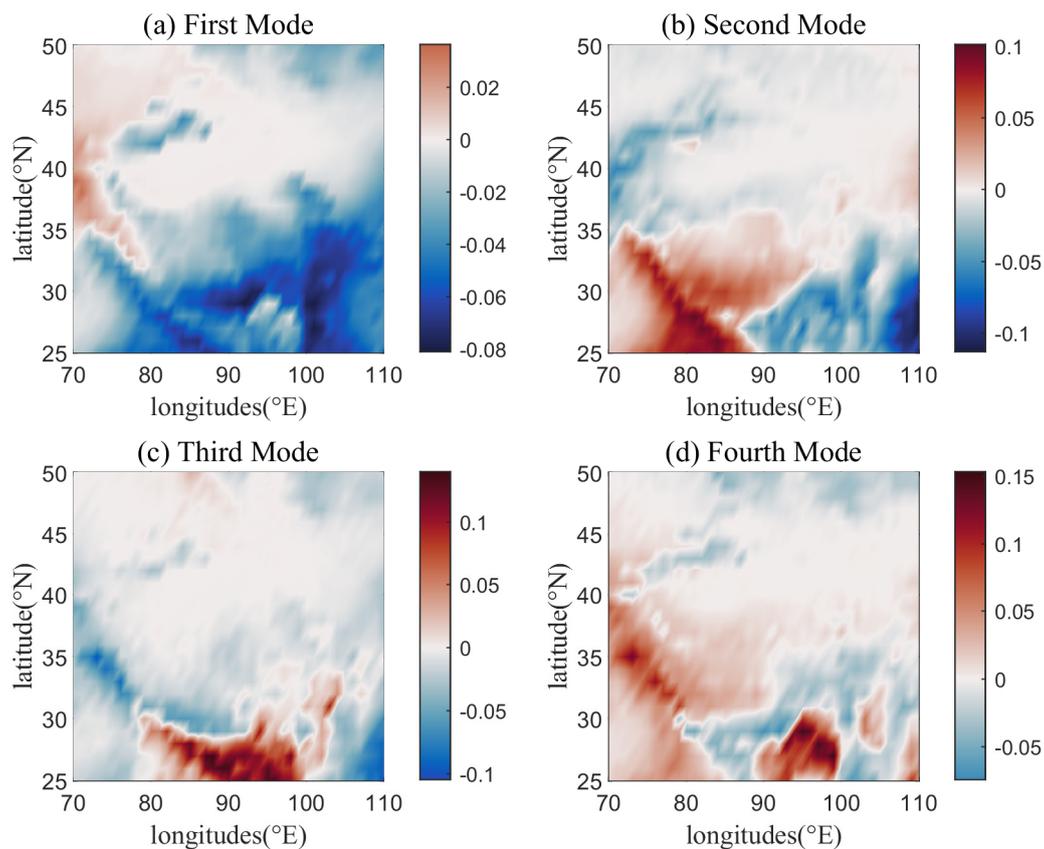
This subsection shows, in detail, two of the most important processes in precipitation prediction experiments, i.e., determining the optimal forecast factor and the setting of the machine learning model parameters.

#### 3.2.1. Selection of the Forecasting Factors

The revised gridded field was subjected to EOF decomposition to obtain each modality’s time series and spatial series. Several modes that have the greatest impact on precipitation in the Western theatre are filtered out based on the variance contribution of each mode. Table 1 reports the variance contribution rates of the first eight precipitation modes, revealing that the first four EOF modes have a larger variance contribution rate, with a cumulative variance contribution rate of 74%. Therefore, this paper conducts precipitation prediction using the time series of the first four modes. Figure 3 depicts the distribution of the time series with space.

**Table 1.** Variance contribution of the first eight EOF modal precipitation time series.

Model	1	2	3	4	5	6	7	8
Expvar	56.7	10.6	4.1	2.6	1.1	0.9	0.6	0.5



**Figure 3.** Distribution of the first four EOF modal precipitation time series over space (the latitude and longitude range of the study area is 25° N–50° N, 70° E–140° E, i.e., the blue-boxed area of Figure 1), (a–d) are the first to fourth EOF modal precipitation time series, respectively.

The 88 atmospheric circulation indices are subjected to information flow-based causal analysis with each of the first four EOF modal time coefficients. Table 2 reports the standardized information flow between the precipitation's first EOF modal time coefficients and the general circulation indices, where columns 1 and 4 represent the ordinal numbers of the 88 atmospheric circulation indices (see Table A1 for the specific names of the indices), columns 2–3 represent the standardized information flow corresponding to the atmospheric circulation indices with the ordinal numbers from 1 to 44, and columns 5–6 represent the standardized information flow corresponding to the atmospheric circulation indices with the ordinal numbers from 45 to 88, with the bolded ordinal numbers in the table representing the best forecasting factors that are finally screened out. We can see standardized information flow is asymmetric and directional, so the direction of causality can be identified. The causal analysis shows that 66 atmospheric circulation indices have significant causal relationships with the monthly precipitation in the western region, and the standardized information flow of the 9th, 20th, 50th, 54th, 71th, and 78th indices indicate that six circulation indices are unidirectional causes of precipitation variations, resulting in 72 of the best forecast factors screened. Similarly, the best forecast factors for the second to fourth EOF modes of precipitation are 69, 41, and 42 items, respectively.

**Table 2.** Standardized information flow between precipitation in the first EOF mode and forecast factors, where highlighted in the table are the best forecast factors screened, and the arrows represent the direction.

Index	Rain→	→Rain	Index	Rain→	→Rain
1	0.499	0.316	45	0.069	0.000
2	0.539	0.378	46	0.049	0.348
3	0.534	0.382	47	0.144	0.155
4	0.013	0.010	48	0.141	0.326
5	0.063	0.002	49	0.061	0.189
6	0.094	0.242	50	0.009	0.360
7	0.279	0.339	51	0.551	0.419
8	0.411	0.378	52	0.366	0.389
9	0.005	0.017	53	0.220	0.360
10	0.448	0.368	54	0.003	0.317
11	0.013	0.142	55	0.552	0.412
12	0.318	0.303	56	0.001	0.011
13	0.547	0.361	57	0.088	0.020
14	0.418	0.352	58	0.537	0.405
15	0.015	0.014	59	0.045	0.234
16	0.068	0.024	60	0.044	0.019
17	0.047	0.205	61	0.034	0.186
18	0.112	0.312	62	0.041	0.030
19	0.135	0.331	63	0.001	0.003
20	0.001	0.012	64	0.515	0.428
21	0.158	0.329	65	0.313	0.100
22	0.016	0.104	66	0.099	0.323
23	0.370	0.523	67	0.045	0.134
24	0.057	0.332	68	0.005	0.007
25	0.317	0.137	69	0.004	0.008
26	0.254	0.253	70	0.002	0.005
27	0.403	0.520	71	0.008	0.014
28	0.309	0.002	72	0.020	0.006
29	0.315	0.204	73	0.010	0.013
30	0.186	0.048	74	0.020	0.027
31	0.352	0.527	75	0.012	0.014
32	0.296	0.255	76	0.000	0.000
33	0.396	0.518	77	0.007	0.012
34	0.286	0.063	78	0.001	0.025
35	0.132	0.190	79	0.001	0.002
36	0.187	0.202	80	0.055	0.039
37	0.067	0.017	81	0.000	0.000
38	0.336	0.528	82	0.257	0.206
39	0.075	0.000	83	0.039	0.086
40	0.189	0.154	84	0.053	0.058
41	0.130	0.014	85	0.029	0.006
42	0.176	0.306	86	0.141	0.115
43	0.200	0.096	87	0.132	0.128
44	0.338	0.459	88	0.002	0.000

### 3.2.2. Machine Learning Model Parameter Settings

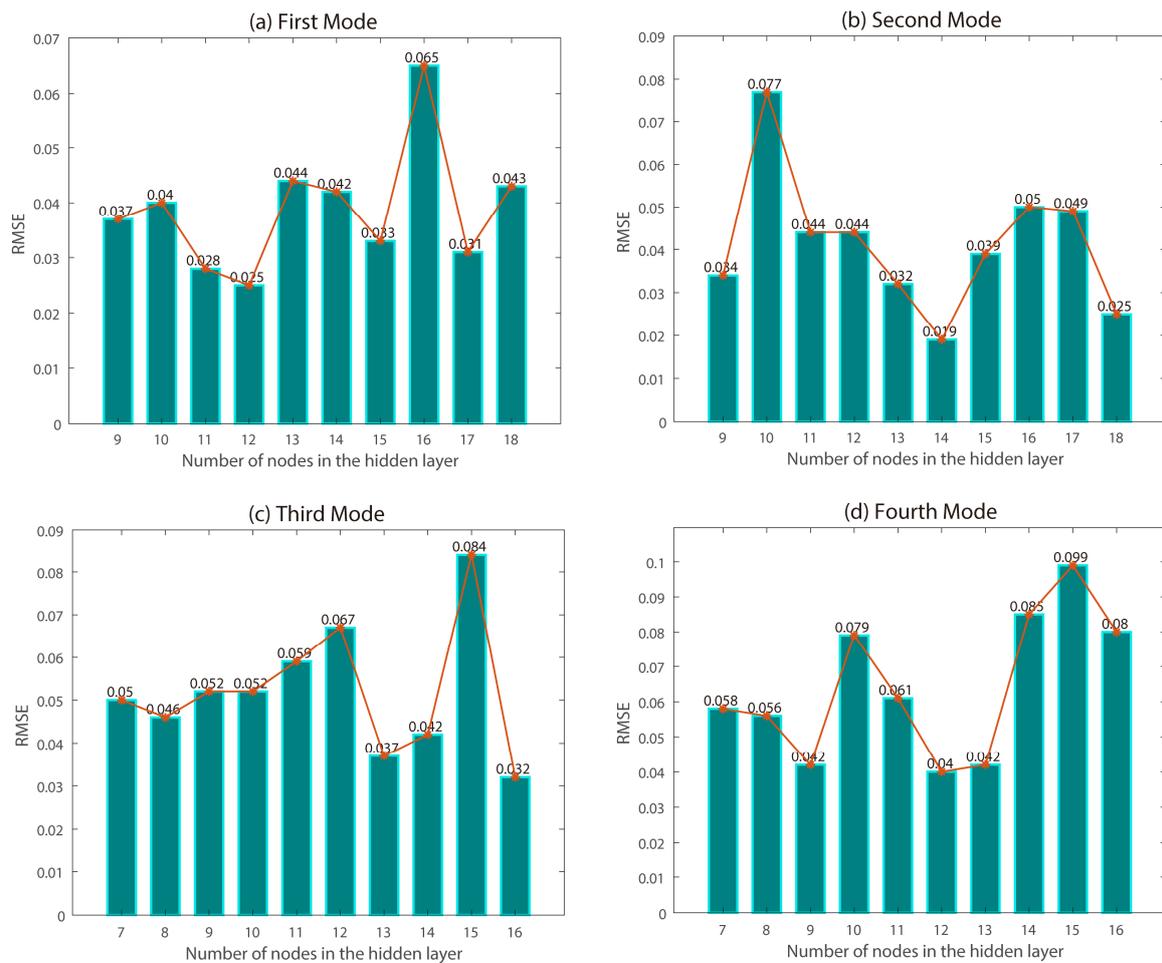
The parameter ranges of the two algorithms of the BPNN and RF are listed in Table 3. In order to avoid overfitting, the parameter settings are as simple as possible to reduce the model complexity, and all of the data are normalized. The number of nodes in the input layer of the BPNN is the number of forecast factors corresponding to the four EOF modes 72, 69, 41, and 42, respectively, and the number of nodes in the output layer is 1. The number of hidden layer nodes is determined by the empirical formula  $k = \sqrt{m + n} + a$ , where k is the number of hidden layer nodes, m and n are the number of nodes in the input layer and the output layer, respectively, and a is an integer between 1~10. Different parameter combinations within the parameter range are modeled separately using the

training set for modeling. The number of decision trees in a random forest is uniformly 500, and the minimum number of leaves is 5.

**Table 3.** Model parameter settings.

ML Algorithms	Parameterization
BPNN	Number of input layer nodes: 72, 69, 41 and 42
	Number of implicit layer nodes:
	First mode: {9, 10, 11, ..., 18}
	Second mode: {9, 10, 11, ..., 18}
	Third mode: {7, 8, 9, ..., 16}
Fourth mode: {7, 8, 9, ..., 16}	
Number of output layer nodes: 1	
RF	Number of decision trees: 500
	Minimum leaves number: 5

Figure 4 demonstrates the RMSE for four EOF modes with a starting time of 3 months and the BPNN taking different numbers of hidden layer nodes. It can be seen that the optimal number of hidden layer nodes for the first to fourth EOF modes are 12, 14, 16, and 12, respectively, corresponding to RMSEs of 0.025, 0.019, 0.032, and 0.040.



**Figure 4.** First to fourth EOF modes, starting time 3 months, RMSE when BPNN takes a different number of hidden layer nodes.

#### 4. Analysis of Precipitation Prediction Results

Based on the revised precipitation grid field, this paper adopts the BPNN and RF algorithms to establish the precipitation prediction model and conduct the 1 m, 3 m, and 6 m precipitation forecasts. The predicted grid point fields output from the model are interpolated to each station, compared, and analyzed with the measured values at the station. Then, the model’s precipitation forecasting effect at each weather station is evaluated using the ACC and RMSE indicators.

Tables 4 and 5 present the ACC and RMSE metrics distribution for the two machine learning algorithms at the three starting times, respectively. For the BPNN algorithm, the number of stations with an ACC greater than 0.4 at each starting time is 168, 152, and 145, respectively, and the stations with a negative correlation between the forecast and measured precipitation gradually appear as the starting time increases. The number of stations with an RMSE index of less than 10 at each starting time is 90, 84, and 81, respectively, and the number of stations with an RMSE index of more than 30 is 44, 54, and 102, respectively. Combining the two indexes, the predicted precipitation level of the model decreases with the increase in the starting time. For the RF algorithm, the number of stations with an ACC greater than 0.4 at each reporting time is 176, 159, and 153, respectively, and there are no stations with no correlation or even a negative correlation between the forecast and measured precipitation. The number of stations with an RMSE of less than 10 at each starting time is 104, 94, and 102, and the number of stations with an RMSE of more than 30 is 12, 3, and 40, respectively. The same conclusion can be obtained by combining the two indexes, revealing that the RF algorithm is better than the BPNN algorithm considering the forecasting effect.

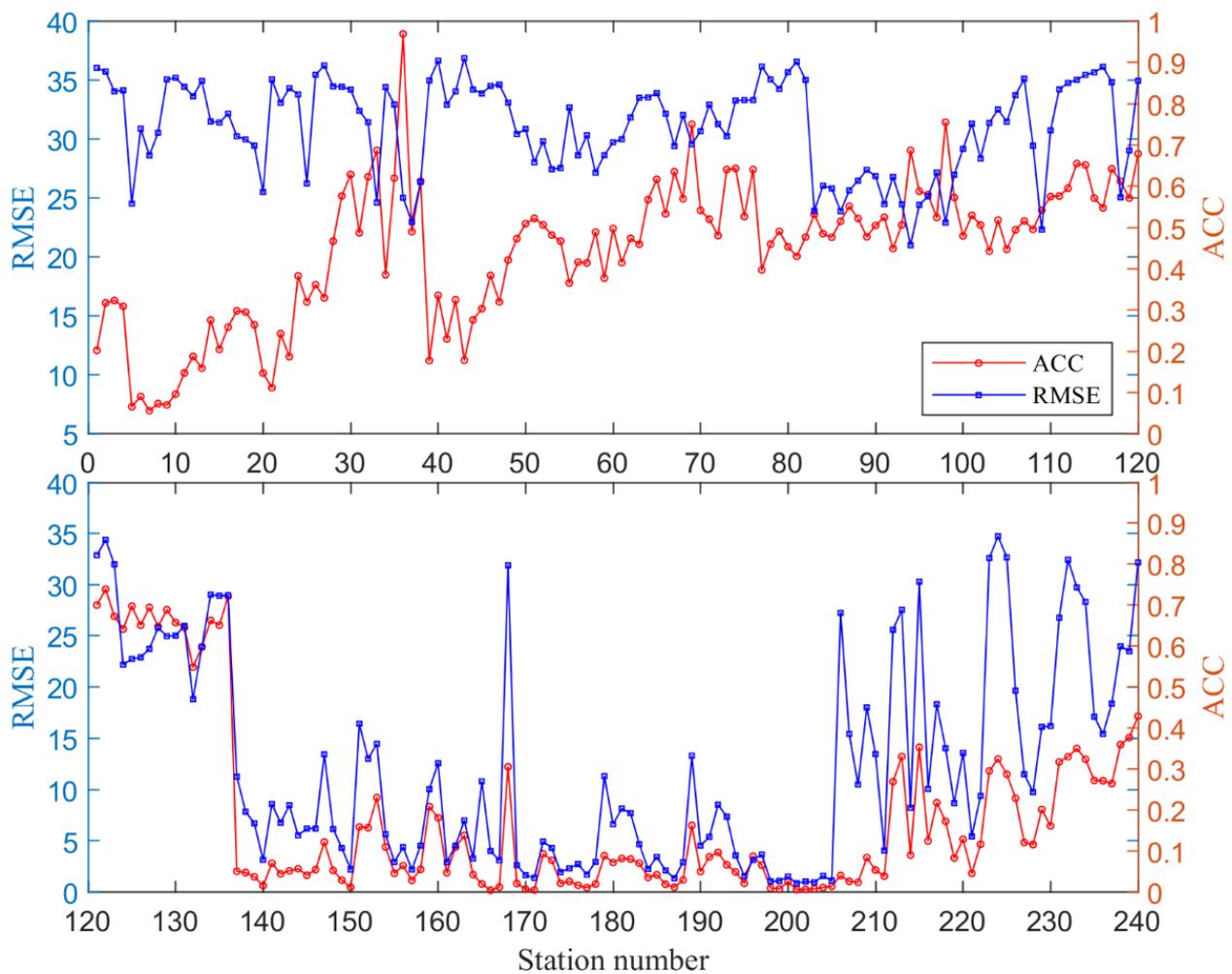
**Table 4.** Distribution of ACC metrics for different ML algorithms with different starting times.

ACC	BPNN			RF		
	1 Month	3 Month	6 Month	1 Month	3 Month	6 Month
$A_{cc} < 0$	0	22	18	0	0	0
$0 < A_{cc} \leq 0.2$	47	37	57	42	56	56
$0.2 < A_{cc} \leq 0.4$	25	29	20	22	25	31
$0.4 < A_{cc} \leq 0.6$	22	35	18	25	33	42
$0.6 < A_{cc} \leq 0.8$	72	97	126	63	69	86
$0.8 < A_{cc} < 1$	74	20	1	88	57	25

**Table 5.** Distribution of RMSE metrics for different ML algorithms with different starting times.

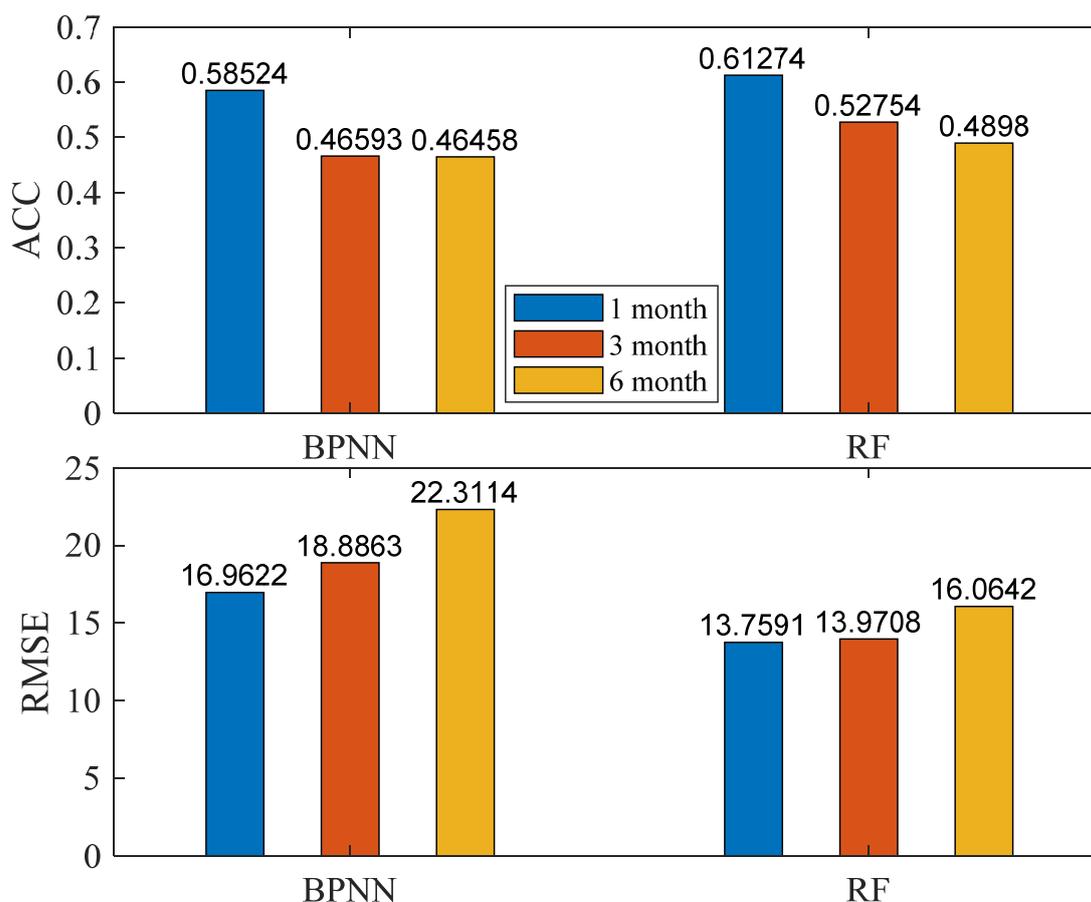
RMSE	BPNN			RF		
	1 Month	3 Month	6 Month	1 Month	3 Month	6 Month
$0 < RMSE \leq 1$	10	9	17	11	23	10
$1 < RMSE \leq 10$	80	75	64	93	71	92
$10 < RMSE \leq 20$	56	48	24	68	54	56
$20 < RMSE \leq 30$	50	54	33	56	89	42
$30 < RMSE \leq 40$	34	36	65	10	3	30
$40 < RMSE$	10	18	37	2	0	10

Figure 5 depicts the correspondence between the RMSE and ACC indicators for 240 stations using the RF algorithm and a starting time of 3 m. The ACC indicator is the same as the RMSE indicator for all stations. By combining all of the stations, it can be obtained that the ACC indicator and RMSE indicator have a good correspondence, and the station with a smaller ACC indicator corresponds to a larger value of the RMSE indicator, i.e., the model’s precipitation prediction level at this station is poorer.



**Figure 5.** RMSE and ACC indicator values for 240 stations with RF algorithms and a 3-month reporting starting time. The blue line is the RMSE metric value for 240 stations, and the red line is the ACC metric value for 240 stations.

Figure 6 illustrates the average RMSE and ACC metrics for the two ML algorithms at three starting times as the overall precipitation prediction level of the model at 240 stations. It can be obtained that the precipitation prediction level of the model gradually decreases with the increase in the starting time, and for the BPNN algorithm, the average ACC metrics of the model at the three starting times are 0.59, 0.47, and 0.46, respectively, corresponding to the average RMSE metrics of 16.96, 18.89, and 22.31. For the RF algorithm, the average ACC metrics of the model at the three starting times are 0.61, 0.53, and 0.49, and the corresponding average RMSE metrics are 13.76, 13.97, and 16.06, respectively. There is an obvious correspondence between the RMSE metrics and the ACC metrics. Thus, comparing the two metrics affords a better evaluation of the model’s precipitation prediction level and, at the same time, corroborates with the conclusions obtained in Figure 5. The RF algorithm is significantly better than the BPNN algorithm at different starting times, with the ACC metrics improved by 4.7%, 13.2%, and 5.4%, and the RMSE metrics reduced by 18.9%, 26%, and 28%, respectively.



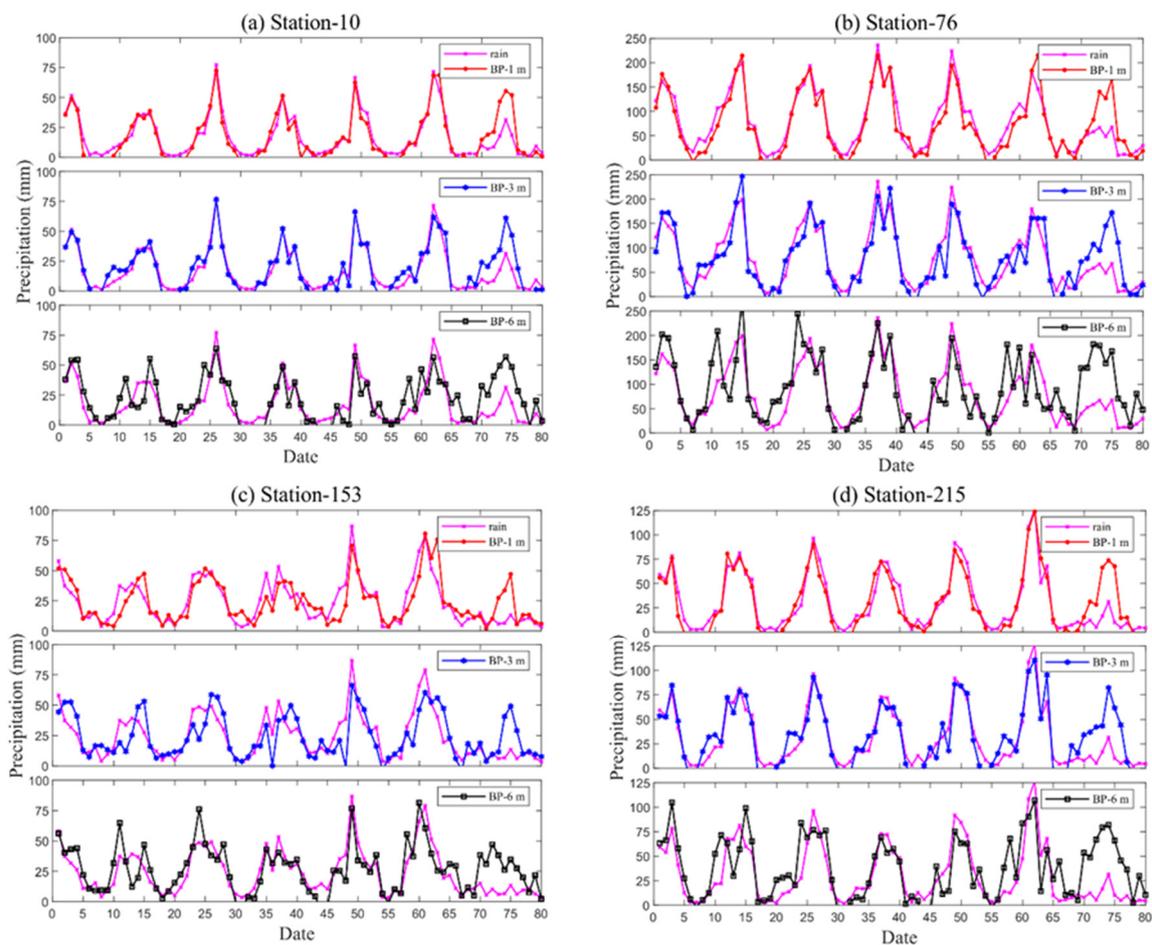
**Figure 6.** Mean RMSE and ACC metrics values for 240 stations with starting report times of 1, 3, and 6 months using BPNN and RF algorithms.

In order to demonstrate the forecasting effect of the model more intuitively, four stations are randomly selected in the Northwest, Southwest, and Plateau zones to demonstrate the effect of the precipitation prediction. Figures 7 and 8 compare the predicted and measured precipitation using the BPNN algorithm and the RF algorithm at the 10th (Northwest), 76th (Southwest), 153th (Northwest), and 215th (Plateau) stations under different starting times. It can be seen that the deviation of the predicted precipitation from the measured precipitation gradually increases with the increase in the starting time and that the RF algorithm is better than the BPNN algorithm in predicting precipitation. The conclusions obtained above are better reflected in the two graphs.

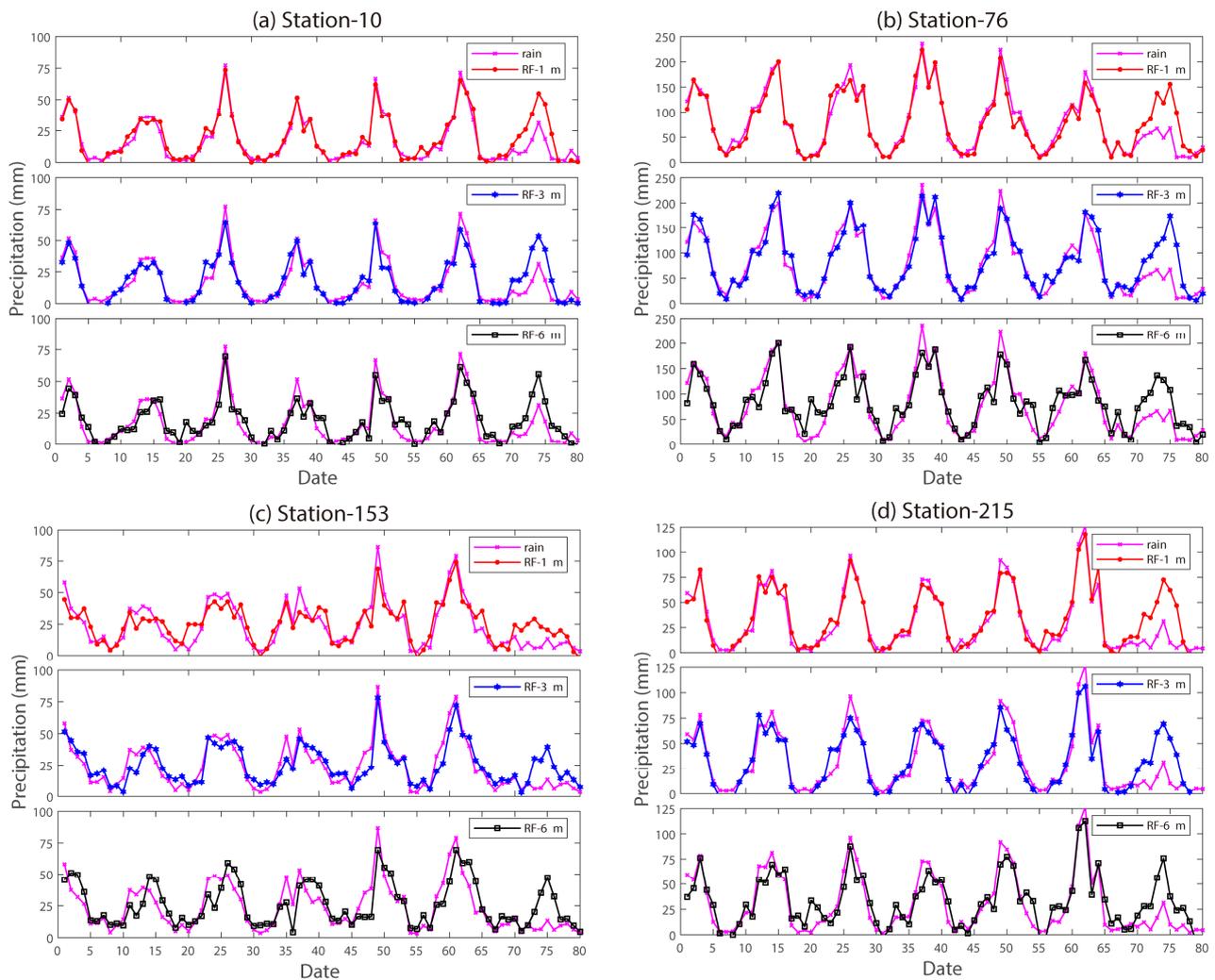
In summary, the model's precipitation prediction level gradually decreases with the increase in the starting time. For the BPNN algorithm, the average ACC metrics of the model at the three starting times are 0.59, 0.47, and 0.46, respectively, and the corresponding average RMSE metrics are 16.96, 18.89, and 22.31. For the RF algorithm, the average ACC metrics of the model at the three starting times are 0.61, 0.53, and 0.49, respectively, and the corresponding average RMSE metrics are 13.76, 13.97, and 16.06. During the starting time of this paper, the RF algorithm significantly outperforms the BPNN algorithm in terms of prediction, with the ACC metrics improving by 4.7%, 13.2%, and 5.4%, and the RMSE metrics decrease by 18.9%, 26%, and 28%, respectively. There is an obvious correspondence between the RMSE and the ACC indicator, which is contrastive, allowing a better evaluation of the model's precipitation prediction level. Indeed, the stations with a poorer predicted precipitation level have smaller ACC values, corresponding to larger RMSE values.

In March 2022, the Climate Prediction Group of the Institute of Atmospheric Physics of the Chinese Academy of Sciences provided the China Meteorological Administration with predictions of precipitation forms during the flood season (June–August) in China. The

testing reality shows that the Institute's forecast opinion captures the national precipitation form well, with a predicted ACC score of 0.53 [33]. Although the study presented in this paper only focuses on the western region, the machine learning-based precipitation prediction model in this paper has high accuracy and is generally able to predict the precipitation trend in the western region better when compared with the benchmark of 0.53, and the model has, to a certain extent, improved the prediction of precipitation at 1 and 3 months, and there is still room to improve the prediction of precipitation at 6 months.



**Figure 7.** For the 10th, 76th, 153th, and 153th stations, the BPNN algorithm is used, and the predicted with measured precipitation are compared under the 1, 3, and 6 months starting time, respectively. (a–d) represent the predicted precipitation versus actual precipitation under the 1, 3, and 6 months starting time for the 10th, 76th, 153th, and 215th stations, respectively. The purple lines represent measured precipitation, the red lines represent 1-month forecasted precipitation, the blue lines represent 3-month forecasted precipitation, and the black lines represent 6-month forecasted precipitation. Horizontal coordinates 0–80 indicate 80 months between July 2011 and September 2021.



**Figure 8.** For the 10th, 76th, 153th, and 153th stations, the RF algorithm is used, and the predicted with measured precipitation are compared under the 1, 3, and 6 months starting time, respectively. (a–d) represent the predicted precipitation versus actual precipitation under the 1, 3, and 6 months starting time for the 10th, 76th, 153th, and 215th stations, respectively. The purple lines represent measured precipitation, the red lines represent 1-month forecasted precipitation, the blue lines represent 3-month forecasted precipitation, and the black lines represent 6-month forecasted precipitation. Horizontal coordinates 0~80 indicate 80 months between July 2011 and September 2021.

## 5. Conclusions

This paper overcomes the problems of complex precipitation influencing factors and the difficulty in determining the predictors by using the measured month-based precipitation data from 240 meteorological stations in the western region of China, the European Centre for Medium-Range Weather Forecast, and the National Climate Centre model precipitation data, and a set of 88 atmospheric circulation indices. The best predictors are screened by combining the EOF decomposition and the causal information flow to develop the western region precipitation prediction model based on the Back Propagation Neural Network and Random Forest algorithm. The investigated precipitation prediction involves a starting time of 1, 3, and 6 months. From our work, the following conclusions are obtained:

- (1) The RF algorithm is significantly better than the BPNN algorithm in terms of prediction, and the predictive ability of both models gradually decreases with the increase in the starting time. The highest average ACC metrics for the model at the three start times are 0.61, 0.53, and 0.49, respectively, corresponding to average RMSE metrics of 13.76, 13.97, and 16.06, respectively. Taking the ACC score (0.53) of the 2022 flood season prediction by the Institute of Atmospheric Sciences of the Chinese Academy of Sciences as a criterion, the precipitation prediction model based on machine learning has high accuracy and is generally able to predict the precipitation trend in western region of China better, and the model has, to a certain extent, improved the prediction ability of 1-month and 3-month precipitation, which can provide a new idea for the short-term climate prediction of precipitation in western China.
- (2) Shortcomings: First, although the precipitation prediction model based on machine learning built in this paper can better predict the precipitation in the western region, there is still a gap in the precipitation prediction effect at individual stations. Second, this paper only considers the atmospheric circulation index when screening the best forecasting factors, and does not consider the effect of the sea temperature index and other related indices (e.g., the sunspot index, the number of cold air counts, the Southern Oscillation index, etc.) on precipitation in the western region.
- (3) Improvement direction: First, take the sea temperature index into consideration when considering the forecasting factors. Second, try more machine learning algorithms and explore the effect of different machine learning algorithms on precipitation prediction in the western region. Finally, when evaluating the effect of precipitation prediction, more evaluation indexes should be introduced to establish a more sound evaluation system for precipitation prediction.

**Author Contributions:** H.L.: Data curation, Formal analysis, Writing—original draft, Methodology, Software; M.L.: Conceptualization, Writing—review and editing, Resources, Software. All authors have read and agreed to the published version of the manuscript.

**Funding:** National Natural Science Foundation, Project approval number: 62073332.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

**Table A1.** Names of Atmospheric Circulation Indices.

Serial Number	Atmospheric Circulation Index Name
1	Northern Hemisphere Subtropical High Area Index
2	North African Subtropical High Area Index
3	North African-North Atlantic-North American Subtropical High Area Index
4	Indian Subtropical High Area Index
5	Western Pacific Subtropical High Area Index
6	Eastern Pacific Subtropical High Area Index
7	North American Subtropical High Area Index
8	Atlantic Subtropical High Area Index
9	South China Sea Subtropical High Area Index
10	North American-Atlantic Subtropical High Area Index

Table A1. Cont.

Serial Number	Atmospheric Circulation Index Name
11	Pacific Subtropical High Area Index
12	Northern Hemisphere Subtropical High Intensity Index
13	North African Subtropical High Intensity Index
14	North African-North Atlantic-North American Subtropical High Intensity Index
15	Indian Subtropical High Intensity Index
16	Western Pacific Subtropical High Intensity Index
17	Eastern Pacific Subtropical High Intensity Index
18	North American Subtropical High Intensity Index
19	North Atlantic Subtropical High Intensity Index
20	South China Sea Subtropical High Intensity Index
21	North American-North Atlantic Subtropical High Intensity Index
22	Pacific Subtropical High Intensity Index
23	Northern Hemisphere Subtropical High Ridge Position Index
24	North African Subtropical High Ridge Position Index
25	North African-North Atlantic-North American Subtropical High Ridge Position Index
26	Indian Subtropical High Ridge Position Index
27	Western Pacific Subtropical High Ridge Position Index
28	Eastern Pacific Subtropical High Ridge Position Index
29	North American Subtropical High Ridge Position Index
30	Atlantic Sub Tropical High Ridge Position Index
31	South China Sea Subtropical High Ridge Position Index
32	North American-North Atlantic Subtropical High Ridge Position Index
33	Pacific Subtropical High Ridge Position Index
34	Northern Hemisphere Subtropical High Northern Boundary Position Index
35	North African Subtropical High Northern Boundary Position Index
36	North African-North Atlantic-North American Subtropical High Northern Boundary Position Index
37	Indian Subtropical High Northern Boundary Position Index
38	Western Pacific Subtropical High Northern Boundary Position Index
39	Eastern Pacific Subtropical High Northern Boundary Position Index
40	North American Subtropical High Northern Boundary Position Index
41	Atlantic Subtropical High Northern Boundary Position Index
42	South China Sea Subtropical High Northern Boundary Position Index
43	North American-Atlantic Subtropical High Northern Boundary Position Index
44	Pacific Subtropical High Northern Boundary Position Index
45	Western Pacific Sub Tropical High Western Ridge Point Index
46	Asia Polar Vortex Area Index
47	Pacific Polar Vortex Area Index
48	North American Polar Vortex Area Index
49	Atlantic-European Polar Vortex Area Index
50	Northern Hemisphere Polar Vortex Area Index
51	Asia Polar Vortex Intensity Index
52	Pacific Polar Vortex Intensity Index
53	North American Polar Vortex Intensity Index
54	Atlantic-European Polar Vortex Intensity Index
55	Northern Hemisphere Polar Vortex Intensity Index
56	Northern Hemisphere Polar Vortex Central Longitude Index
57	Northern Hemisphere Polar Vortex Central Latitude Index
58	Northern Hemisphere Polar Vortex Central Intensity Index
59	Eurasian Zonal Circulation Index
60	Eurasian Meridional Circulation Index
61	Asian Zonal Circulation Index
62	Asian Meridional Circulation Index
63	East Asian Trough Position Index
64	East Asian Trough Intensity Index
65	Tibet Plateau Region-1 Index
66	Tibet Plateau Region-2 Index
67	India-Burma Trough Intensity Index

**Table A1.** *Cont.*

Serial Number	Atmospheric Circulation Index Name
68	Arctic Oscillation
69	Antarctic Oscillation
70	North Atlantic Oscillation
71	Pacific/ North American Pattern
72	East Atlantic Pattern
73	West Pacific Pattern
74	North Pacific Pattern
75	East Atlantic-West Russia Pattern
76	Tropical-Northern Hemisphere Pattern
77	Polar-Eurasia Pattern
78	Scandinavia Pattern
79	Pacific Transition Pattern
80	30 hPa zonal wind Index
81	50 hPa zonal wind Index
82	Mid-Eastern Pacific 200 mb Zonal Wind Index
83	West Pacific 850 mb Trade Wind Index
84	Central Pacific 850 mb Trade Wind Index
85	East Pacific 850 mb Trade Wind Index
86	Atlantic-European Circulation W Pattern Index
87	Atlantic-European Circulation C Pattern Index
88	Atlantic-European Circulation E Pattern Index

**Table A2.** Full names and acronyms of related terms in the paper.

Full Name	Acronym
Anomalous Correlation Coefficient	ACC
Atmospheric Circulation Indices	ACIs
Back Propagation Neural Network	BPNN
Bayesian Network	BN
Convolutional Neural Network	CNN
Decision Tree	DT
European Centre for Medium-Range Weather Forecasts	ECMWF
Empirical Orthogonal Function	EOF
Long Short Term Memory	LSTM
Machine Learning	ML
National Climate Centre	NCC
Rotated Empirical Orthogonal Function	REOF
Root Mean Square Error	RMSE
Support Vector Machine	SVM

## References

- Guo, H.; Li, D.L.; Lin, S.; Dong, Y.X.; Sun, L.D.; Huang, L.N.; Lin, J.J. Temporal and spatial variation of precipitation over western China during 1954–2006. *J. Glaciol. Geocryol.* **2013**, *35*, 1165–1175.
- Chen, Z.K.; Zhang, S.Y.; Luo, J.L.; Li, Z. Climatic analysis of precipitation anomalies in northwest China. *J. Desert Res.* **2013**, *33*, 1874–1883.
- Nan, Z.T.; Gao, Z.S.; Li, S.X.; Wu, T.H. Permafrost changes in the northern limit of permafrost on the Qinghai–Tibet Plateau in the last 30 years. *Acta Geogr. Sin.* **2003**, *58*, 817–823.
- Wu, G.C.; Qin, S.; Hang, C.C.; Ma, Z.S.; Shi, C.M. Seasonal precipitation variability in mainland China based on entropy theory. *Int. J. Climatol.* **2021**, *41*, 5264–5276.
- Mai, Z.N.; Xu, D.B.; Xiao, T.G.; Yan, X.J.; Lu, S. Preliminary Study on Probability Forecasting of Flash-Heavy-Rain in Chengdu. *Plateau Mt. Meteorol. Res.* **2022**, *42*, 127–134.
- Wang, Y.; Li, H.X.; Wang, H.J.; Sun, B.; Chen, H.P. Evaluation of the simulation capability of CMIP6 global climate model for extreme precipitation in China and its comparison with CMIP5. *Acta Meteorol. Sin.* **2021**, *79*, 369–386.
- Sun, H.; Gao, Z. Prediction of seasonal precipitation based on wavelet analysis and Kriging interpolation. *Comput. Electron. Agric.* **2020**, *177*, 105644.

8. Anaraki, M.V.; Farzin, S.; Mousavi, S.F. Dynamical-statistical hybrid model for seasonal precipitation prediction: A case study of Indus basin, Pakistan. *J. Hydrol.* **2020**, *588*, 125010.
9. Alireza, F.; Saeed, F.; Sayed-Farhad, M. Meteorological drought analysis in response to climate change conditions, based on combined four-dimensional vine copulas and data mining (VC-DM). *J. Hydrol.* **2021**, *603*, 127135.
10. Pan, L.J.; Xue, C.F.; Zhang, H.F.; Gao, X.X.; Liu, M.; Liu, J.H.M. ECMWF precipitation calibration based on the Kalman dynamic frequency matching method. *Meteorol. Mon.* **2022**, *48*, 73–83.
11. Chen, L.L.; Xia, Y. The Influence of Ensemble Size on Precipitation Forecast in a Convective Scale Ensemble Forecast System. *J. Appl. Meteorol. Sci.* **2023**, *34*, 142–153.
12. Wang, Y.T. Precipitation forecast of the Wujiang River Basin based on artificial bee colony algorithm and backpropagation neural network. *Alex. Eng. J.* **2020**, *59*, 1473–1483. [[CrossRef](#)]
13. Chen, S.; Sun, Y.N.; Sa, R.N. Research on precipitation prediction based on LSTM and BP neural network. *GANSU Water Resour. Hydropower Technol.* **2023**, *59*, 7–11.
14. Anaraki, M.V.; Farzin, S.; Mousavi, S.-F.; Karami, H. Uncertainty Analysis of Climate Change Impacts on Flood Frequency by Using Hybrid Machine Learning Methods. *Water Resour. Manag.* **2021**, *35*, 199–223. [[CrossRef](#)]
15. Nakhaei, M.; Mohebbi Tafreshi, A.; Saadi, T. An evaluation of satellite precipitation downscaling models using machine learning algorithms in Hashtgerd Plain, Iran. *Model. Earth Syst. Environ.* **2023**, *9*, 2829–2843. [[CrossRef](#)]
16. Huang, C.; Li, Q.P.; Xie, Y.J.; Peng, J.D. Prediction of summer precipitation in Hunan based on machine learning. *Trans. Atmos. Sci.* **2021**, *45*, 191–202.
17. Ashutosh, S.; Manish, K.G. Bayesian network for monthly rainfall forecast: A comparison of K2 and MCMC algorithm. *Int. J. Comput. Appl.* **2016**, *38*, 199–206.
18. Xing, W.; Han, W.Q.; Zhang, L. Improving the prediction of western North Pacific summer precipitation using a Bayesian dynamic linear model. *Clim. Dyn.* **2020**, *55*, 831–842. [[CrossRef](#)]
19. Chen, J.P.; Feng, Y.R.; Meng, W.G.; Meng, W.G. A correction method of hourly precipitation forecast based on convolutional neural network. *Meteor. Mon.* **2021**, *47*, 60–70.
20. Wang, Y.C.; Wei, J.H.; Li, Q.; Qiao, Z.; Yang, K.; Zhu, X.; Bao, S.P.; Wang, Z.J. Radar echo-based study on convolutional recurrent neural network model for precipitation nowcast. *Water Resour. Hydropower Eng.* **2023**, *54*, 24–41.
21. Zhu, K.; Yang, Q.; Zhang, S.; Jiang, S.; Wang, T.; Liu, J.; Ye, Y. Long lead-time radar rainfall nowcasting method incorporating atmospheric conditions using long short-term memory networks. *Front. Environ. Sci.* **2023**, *10*, 1054235. [[CrossRef](#)]
22. Li, M.; Liu, K.F. Probabilistic prediction of significant wave height using dynamic bayesian network and information flow. *Water* **2020**, *12*, 2075. [[CrossRef](#)]
23. Runge, J.; Bathiany, S.; Bollt, E.; Camps-Valls, G.; Coumou, D.; Deyle, E.; Glymour, C.; Kretschmer, M.; Mahecha, M.D.; Muñoz-Mari, J.; et al. Inferring causation from time series in Earth system sciences. *Nat. Commun.* **2019**, *10*, 2553. [[CrossRef](#)]
24. Marlene, K.; Dim, C.; Donges, J.F.; Runge, J. Using Causal Effect Networks to Analyze Different Arctic Drivers of Midlatitude Winter Circulation. *J. Clim.* **2016**, *29*, 4069–4081.
25. McGraw, M.C.; Barnes, E.A. New Insights on Subseasonal Arctic–Midlatitude causal connections from a regularized regression model. *J. Clim.* **2020**, *33*, 213–228. [[CrossRef](#)]
26. Li, M.; Zhang, R.; Liu, K.F. Machine learning incorporated with causal analysis for short-term prediction of sea ice. *Front. Mar. Sci.* **2021**, *8*, 649378. [[CrossRef](#)]
27. Zhang, Y.; Liang, X.S. The causal role of South China Sea on the Pacific–North American teleconnection pattern. *Clim. Dyn.* **2022**, *59*, 1815–1832. [[CrossRef](#)]
28. Docquier, D.; Vannitsem, S.; Ragone, F.; Wyser, K.; Liang, X.S. Causal links between Arctic Sea ice and its potential drivers based on the rate of information transfer. *Geophys. Res. Lett.* **2022**, *49*, e2021GL095892. [[CrossRef](#)]
29. Luo, W.; Yin, H.; Yang, S.; Zhou, Y.; Ran, L.; Jiao, B.; Lai, Z. Response of a westerly-trough rainfall episode to multi-scale topographic control in southwestern China. *Atmos. Ocean. Sci. Lett.* **2022**, *15*, 100148. [[CrossRef](#)]
30. Zhu, J.T.; Yang, Q.Y.; Li, X.; Li, Y. Characteristics and East Downscaling Forecast Model of Summer Precipitation in Northwest China. *Plateau Meteorol.* **2023**, *42*, 646–656.
31. Liang, X.S. Unraveling the cause-effect relation between time series. *Phys. Rev. E* **2014**, *90*, 052150. [[CrossRef](#)] [[PubMed](#)]
32. Liang, X.S. Normalizing the causality between time series. *Phys. Rev. E* **2015**, *92*, 022126. [[CrossRef](#)] [[PubMed](#)]
33. Peng, J.B.; Zheng, F.; Fan, F.X.; Chen, H.; Lang, X.M.; Zhan, Y.L.; Lin, C.H.; Zhang, Q.Y.; Lin, R.P.; Li, C.F.; et al. Climate Prediction and Outlook in China for the Flood Season 2022. *Clim. Environ. Res.* **2022**, *27*, 547–558.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.