sustainability

MDPI

Article

A Statistical Prediction Model for Summer Precipitation in China Based on TSD Method and EOF Modes' Time Coefficients

Zihuang Xie [1], Yimin Zhu [1,2,*], Yijia Hu [1], Yao Ha [1] and Zhong Zhong [1,2]

[1] College of Meteorology and Oceanography, National University of Defense Technology, Changsha 410003, China
[2] Jiangsu Collaborative Innovation Center for Climate Change, School of Atmospheric Sciences, Nanjing University, Nanjing 210023, China
* Correspondence: zhuym@nudt.edu.cn or zhuym@21cn.com

Abstract: It is a challenge to improve the skill of seasonal precipitation prediction, because there are many factors affecting summer precipitation in China, which are found on different time scales and have complex interactions with each other. For these reasons, we establish a prediction model with the time-scale decomposition (TSD) method to investigate whether the TSD has an improving effect on the prediction skill of summer precipitation in China. Using this statistical model, the predictors and predictands will be separated into interannual and interdecadal time scales, after which Empirical Orthogonal Function (EOF) decomposition is performed on these two components, and their time coefficients are predicted, respectively. The hindcast cross-validation results show that the model without TSD has prediction skills only in some regions of East China and South China. Compared with the model without TSD, surprisingly, the model with TSD can significantly improve the prediction performance in more regions in China, such as Xinjiang Province and Northeast China. The anomaly correlation coefficients (ACC) between hindcast precipitation with TSD and observation are higher in most years than that without TSD. The results of the independent sample test show that the forecast model with TSD has a stable and gratifying prediction skill, and the averaged ACC is increased by more than 0.1.

Keywords: time-scale decomposition; Empirical Orthogonal Function; time coefficients; statistical prediction model; summer precipitation

Citation: Xie, Z.; Zhu, Y.; Hu, Y.; Ha, Y.; Zhong, Z. A Statistical Prediction Model for Summer Precipitation in China Based on TSD Method and EOF Modes' Time Coefficients. *Sustainability* **2023**, *15*, 10928. https://doi.org/10.3390/su151410928

Academic Editor: Jan Hopmans

Received: 19 May 2023
Revised: 6 July 2023
Accepted: 11 July 2023
Published: 12 July 2023

1. Introduction

China is located in the climate region of the East Asian monsoon, and its summer precipitation is affected by many factors, such as the East Asian summer monsoon (EASM) [1–4], West Pacific Subtropical High (WPSH) [5–8], El Niño-Southern Oscillation (ENSO) [9–12], Pacific decadal oscillation (PDO) [13–15], etc. The summer precipitation in China exhibits obvious characteristics of not only interannual and interdecadal variability but also the uneven spatial distribution of drought and flood. With the intensification of global warming, extreme precipitation events also become more and more frequent in China. For example, during the Meiyu period in 2020, the middle and lower reaches of the Yangtze River Valley experienced the strongest rainfall since 1960, lasting up to 62 days, while South China suffered severe drought. In July 2021, Henan Province, located in central China, experienced a rare local rainstorm which then triggered floods, causing serious casualties and property losses. Therefore, summer precipitation has been the key to short-term climate forecast research in China.

The interannual variability is one of the most significant characteristics of summer precipitation in China, which is mainly affected by EASM and is closely related to anomalies of WPSH [5–8], sea surface temperature (SST) [9–11,16], sea ice [17–19] and other forcing

signals. Wu and Zhou [5] believed that the strong interannual variability of summer precipitation in East China is caused by the significant interannual variability of WPSH. Some studies have paid attention to the impact factors of EASM and WPSH, including ENSO in the tropical Pacific and SST anomaly in the tropical Indian Ocean. Huang and Wu [9] proposed that in the development phase of ENSO, SST in western tropical Pacific is colder in summer, and the convection over South China Sea is weak, which makes the WPSH move southward and leads to increased precipitation in the Yangtze–Huaihe River Valley; the study of Feng et al. [11] showed that in the following summer of EP El Niño (with the maximum SST anomaly typically occurring in the eastern equatorial Pacific), precipitation increases in the south of the Yangtze River and decreases in the Yangtze–Huaihe River Valley, while in the following summer of CP El Niño (with the maximum SST anomaly typically occurring in the central equatorial Pacific), precipitation increases in the Yangtze River and Huang–Huai River Valley and decreases in the south of the Yangtze River. Zhou et al. [16] pointed out that after the 2019 Indian Ocean dipole (IOD) event disappeared in winter, the Indian Ocean basin warming (IOBW) was maintained by a weak El Niño event until the 2020 summer, which triggered the abnormal anticyclone over western Pacific and enhanced the westerly jet over the Yangtze River Valley, resulting in more precipitation in China. For the continental part of China, studies have also shown that SST anomaly has an impact on their summer precipitation [20,21]. In addition, it has been revealed that decreased spring sea ice concentration (SIC) in the Arctic Ocean and the Greenland Sea will lead to increased summer rainfall in Northeast China and Central China between the Yangtze River and the Yellow River, while summer rainfall decreased in South China [17].

Moreover, on the interdecadal time scale, Ding et al. [2] studied the interdecadal relationship between summer precipitation of China and EASM, and pointed out that the summer precipitation in East China has two interdecadal adjustments in 1978 and 1992, respectively, with the main rain belt moving southward gradually; Feng et al. [13] studied the modulation of PDO on the EASM in El Niño decay year, that is, when El Niño and PDO are in phase (El Niño event/high PDO), abnormal anticyclones and cyclones appear near the Philippines and Japan, respectively, and then the precipitation anomaly presents a meridional triple structure in China, while when El Niño and PDO are out of phase (El Niño event/low PDO), the abnormal anticyclone is in the Northwest Pacific and the precipitation anomaly presents a meridional dipole structure; the study of Yang et al. [22] showed that during the positive phase of interdecadal Pacific Oscillation (IPO), the abnormal cyclone over North China leads a further northerly rain belt and East China will present a pattern of wet–north and dry–south, and when the phase of IPO is negative, the above situations are opposite. Zhang et al. [23] examined the joint influence of PDO, Atlantic multidecadal oscillation (AMO), and IOBW, and revealed the physical mechanism of interdecadal variability of summer precipitation in East Asia and China over the past century through the combination of different phases among the three.

As mentioned above, many factors have different effects on summer precipitation in China, and they also have complex interactions with each other, which makes the forecast a great challenge. Previous research has revealed one of the most important reasons is that the prediction signal on the interannual time scale is inconsistent with that on the interdecadal time scale [24]. Therefore, it is necessary to separate the time scale of the predictands into different time scales, investigate the influencing factors and prediction models of each time scale separately, and then overlay their results as the overall prediction of the objective physical quantity. This method is called the time-scale decomposition (TSD) prediction method [25–28]. Some scholars have researched this problem before; Hu et al. [25] decomposed the rainfall of flood season in the middle and lower reaches of the Yangtze River Valley into interannual components of less than 8 years, interdecadal components between 8 and 30 years, and interdecadal components longer than 30 years, then selected the predictors, respectively and established a statistical forecast model through multiple linear regression method; Liu and Li [26] adopted a statistical downscaling scheme based

on the TSD to forecast summer rainfall in the Yangtze–Huaihe River Valley; Guo et al. [27] conducted a stepwise regression on the interannual and interdecadal components of summer rainfall in North China, which is based on cross-validation, to select the best predictors and establish a downscaling model; Song et al. [28] proposed a TSD threshold regression downscaling approach for forecasting early summer rainfall in South China. All the conclusions mentioned above show that the application of the TSD method can effectively improve the performance of summer precipitation prediction models.

EOF (Empirical Orthogonal Function) decomposition method was first proposed by Pearson [29], and then Lorenz [30] applied this method to the research of meteorology. It decomposes the variable field that changes over time into the part of spatial patterns that do not change over time and the part of time coefficients that only depend on time. The EOF decomposition method has the advantage of a fast convergence rate, and the eigenvectors are orthogonal to each other. The spatial patterns can reflect the main spatial distribution characteristics of the variable fields, and the corresponding time coefficients can reflect the weight of each EOF mode. Therefore, this method has been widely adopted in the field of climate analysis and forecast, especially for precipitation [31,32]. Xing et al. [33] proposed an EOF–PLS regression-based method to make a long-lead seasonal forecast of summer precipitation in China, in which the previous SST and 2 m temperature are used as predictors. Ma and Sun [34] combined EOF decomposition with multiple linear regression to study the relationships between the SST of preceding winter and EOF leading modes of precipitation in Northeast China from June to August, respectively, and then constructed a statistical prediction scheme. Huang et al. [35] used EOF analysis and the interannual increment approach to propose an effective prediction method for summer precipitation over eastern China. Liu and Zhu [36] took the regression coefficients of EOF leading modes as predictands and proposed an approach to search for potential prediction skills of predictors by ordinary least squares and cross-validation methods, namely Potential Skill Map (PSM). They also developed an automatic selector of predictors based on this approach to establish a statistical prediction model for summer precipitation anomaly in China.

To sum up, due to the complex impact factors and the interaction between them, there is still a great challenge for the forecast of summer precipitation in China. As mentioned earlier, the application of the TSD method can significantly improve the precipitation prediction models, while currently, predictions using the EOF decomposition method usually consider signals on an interannual time scale only.

Therefore, we raise the question here, as mentioned earlier, that it is necessary to divide the predictands and predictors into interannual and interdecadal time scales; if the TSD method is added to the precipitation prediction model based on EOF decomposition, can its prediction accuracy also be improved? Based on the above hypotheses, we filtered and decomposed the summer precipitation anomalies in China into interannual and interdecadal time scales, and then used their time coefficients of EOF leading modes as the predictands. With the linear relationship between each grid point in the physical quantity fields and the time series, the hindcast results were tested, and the predictors were obtained to establish a statistical prediction model based on the TSD method and the time coefficients of EOF mode. Through cross-validation and an independent sample test, it is found that the prediction after applying the TSD method is indeed improved.

2. Materials and Methods

Our analysis and prediction are based on the summer (from June to August) precipitation anomaly in China. Considering the forcing factors, as mentioned earlier, many studies have shown that the SST anomalies in the tropical Pacific and Indian Oceans are closely related to the summer precipitation in China [9–11,16]. There are also studies focusing on the role of Atlantic SST [37,38]. The impact of Arctic sea ice on summer precipitation in different regions of China has also been confirmed [19,39]. In addition, 500 hPa geopotential height can reflect the changes of the WPSH [5–8], teleconnection wave train [39,40], trough/ridge [41,42], etc. In terms of identifying upper-level jets, 200 hPa

zonal wind helps, and its convergence and divergence also provide conditions for precipitation anomaly [43,44]. These circulation anomalies can directly or indirectly affect or modulate the summer precipitation in China. So, it can be seen that SST, SIC, 500 hPa geopotential height, and 200 h Pa zonal wind are suitable as the predictands of the summer precipitation in China.

Therefore, the following data are used in this study: (1) the monthly observation precipitation data set of 160 stations in China from China Meteorological Administration, and the distribution of stations are shown in Figure 1; (2) the monthly reanalysis data set with the horizontal resolution of 2.5° × 2.5° from NCAR/NCEP (National Center for Atmospheric Research/National Centers for Environmental Prediction), including spring (from March to May) 200 hPa zonal wind (U200) and 500 hPa geopotential height (H500) on the global range [45]; (3) the monthly SST and SIC observation data set in winter (from previous December to February) with the horizontal resolution of 1.0°×1.0° from UK Met Office Hadley Centre [46], and the range of SST is (0°–360°, 60° S–60° N) and SIC is (0°–360°, 60–90° N). The time range of our study is 1961 to 2020; the period 1961–2010 is used for hindcast based on cross-validation test and 2011–2020 for independent sample test. In order to reduce the effect of global warming, the linear trend of SST and SIC is removed additionally.

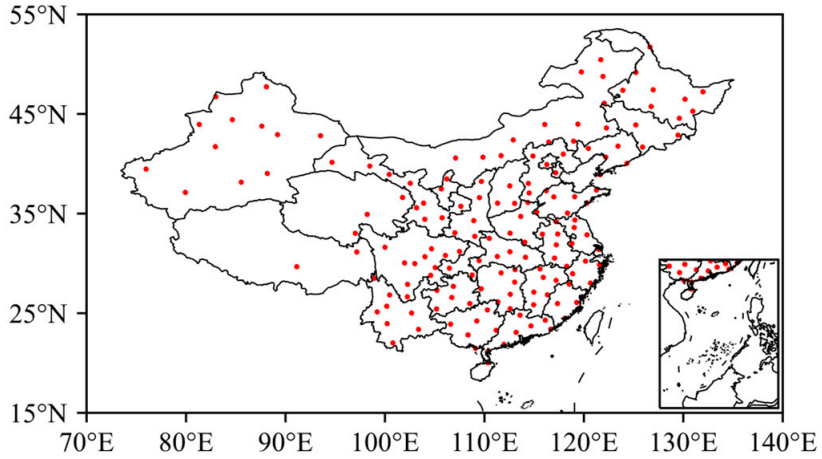

Figure 1. Spatial distribution of 160 stations in China.

Firstly, before all the analysis and prediction, the TSD method is applied. To achieve TSD of precipitation anomaly, referring to the previous works [25–27], a Butterworth filter with an 8-year cutoff is used, and it separates the variable into interannual and interdecadal signals for subsequent study. To avoid the influence of possible false boundary points, the precipitation data from 1951 to 2020 are filtered, but only the data from 1961 to 2010 are used for the selection of predictors and the establishment of the prediction model. Other variables are not filtered.

Secondly, the primary variability of summer precipitation on interannual and interdecadal time scales is extracted using the EOF method, assuming that there are n years of precipitation data Y and the predicted year is the i-th year. In order to avoid bringing the signal of the predicted year into the hindcast model through the EOF leading modes, which may affect the accuracy and credibility of the result, this study will perform EOF decomposition based on the idea of cross-validation. The number m of EOF modes used in the prediction model is determined by the comparison between the reconstructed and the observed precipitation fields. By this step, we obtain the time coefficients of suitable EOF leading modes on interannual and interdecadal time scales, i.e., $T_1(m,n)$ and $T_2(m,n)$, to be the predictands and select the corresponding predictors in the following, respectively.

Thirdly, after obtaining the time coefficients $T_1(m,n)$ and $T_2(m,n)$ of EOF modes, referring to the study of Liu and Zhu [36], we will preliminarily select the possible predictors from the previous autumn and winter variable fields $X(x,y,n)$ through the following steps.

x and y describe the range of the variable fields X in zonal and meridional directions, respectively. Here, the ordinary least squares method is used individually for fitting the linear function relationships $R_1(m,x,y)$ between $X(x,y,n)$ and $T_1(m,n)$, and $R_2(m,x,y)$ between $X(x,y,n)$ and $T_2(m,n)$. Next, substitute the variable fields $X_{p,q,i}$ $(1 \le p \le x, 1 \le q \le y)$ of the i-th year into $R_1(m,x,y)$ or $R_2(m,x,y)$ for each mode to obtain the hindcast time coefficient fields $\hat{T}_{1i}(m,x,y)$ or $\hat{T}_{2i}(m,x,y)$, and repeat the above steps for n years to obtain $\hat{T}_1(m,x,y,n)$ and $\hat{T}_2(m,x,y,n)$.

Finally, calculate the correlation coefficients between $\hat{T}_1(m,x,y,n)$ and $T_1(m,n)$ or $\hat{T}_2(m,x,y,n)$ and $T_2(m,n)$, to be the selection foundation for potential predictors. In addition, the final determination method for the predictors of the time coefficients corresponding to each EOF mode and the establishment of the prediction model will be illustrated in detail in Section 3.3.

Before and after applying the TSD method to the precipitation anomaly fields, EOF decomposition, selection of predictors, and establishment of prediction models will be conducted. The complete prediction process using the TSD method is shown in Figure 2.

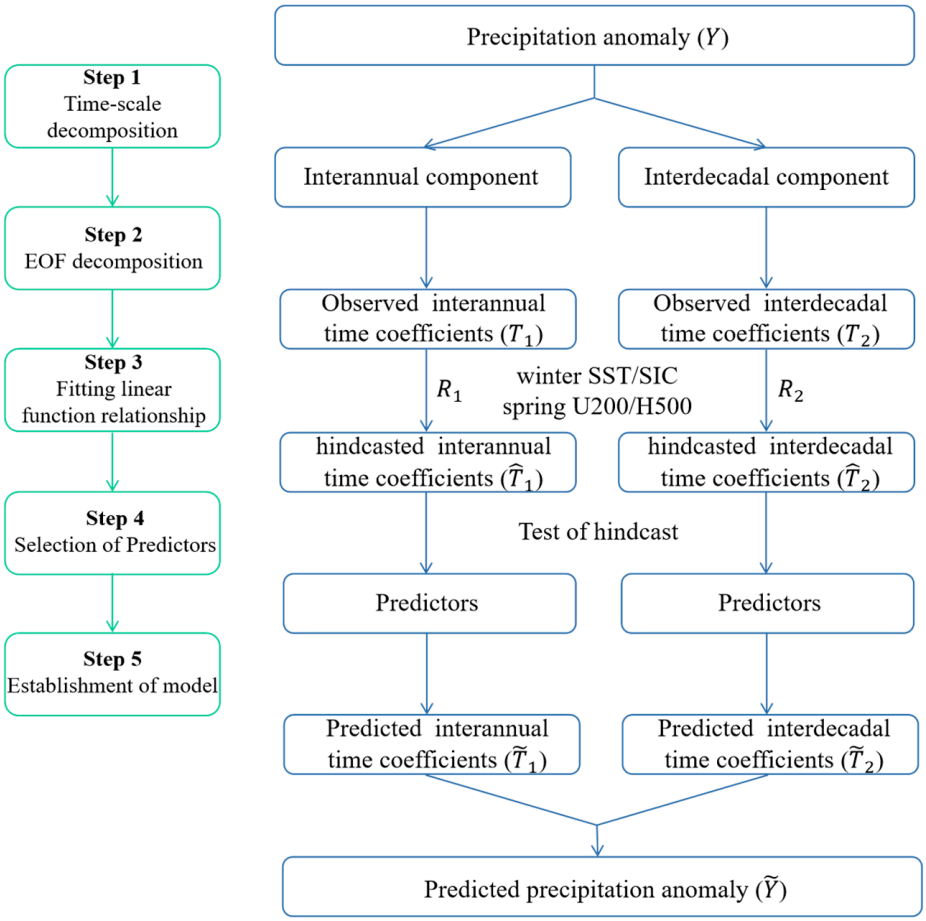

Figure 2. The complete process of the prediction model with the TSD method.

3. Results

3.1. Determination of the Number m in EOF and Comparison of Reconstructed Fields

In the second step mentioned in the previous section, considering how many EOF modes (number m) we need to reconstruct the precipitation anomaly fields is a very critical problem for the prediction. On the one hand, more EOF modes are needed to better reconstruct the abnormal variability of summer precipitation in China. On the other hand, previous studies have suggested that too many EOF modes may make the prediction model unstable [36,47,48]. Therefore, the number of EOF modes (m) is determined by the anomaly correlation coefficient (ACC) and spatial variance ratio between the observed fields and

the reconstructed fields. The reconstructed precipitation anomalies can be calculated based on the spatial patterns $V_{k,i}$ (k represents the ordinal number of the mode) and their corresponding time coefficients $T_{k,i}$ of EOF modes according to the following Formula (1)

$$Y_i = \sum_{k=1}^{m} T_{k,i} V_{k,i} \quad (1)$$

From Figure 3, it is obvious that before and after using the TSD method, the anomaly correlation coefficient (ACC) and spatial variance ratio between the observed fields and the reconstructed fields will increase as the number of EOF modes increase. It can be found in Figure 3a,b that although there are differences between reconstructed fields using the same number of modes before and after TSD, both of them increase significantly from two to eight modes, and the growth rate slows down after more than eight modes, which exhibit the nonlinear growth characteristic in general. For this result, regardless of whether TSD is performed or not, the interception of the first eight EOF modes for analysis can reduce the possible instability effect on the prediction model, and better represent the main characteristics of the observed precipitation.

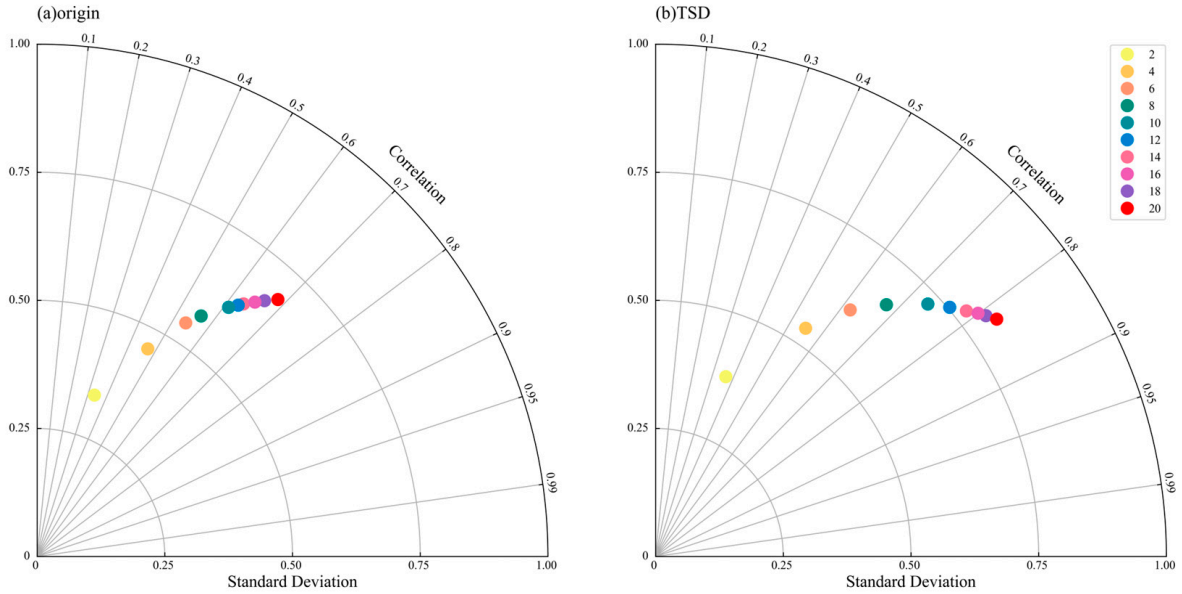

Figure 3. Taylor diagrams of the observed precipitation anomaly fields and the reconstructed precipitation anomaly fields (**a**) without and (**b**) with TSD during 1961–2010 (dots represent the different numbers of EOF modes used for reconstruction).

Based on the spatial patterns and time coefficients of the first eight modes, the precipitation anomaly fields are reconstructed, and temporal correlation coefficients (TCC) are calculated to compare the influence of the TSD method on the reappearance ability of the precipitation anomaly (Figure 4). From Figure 4a,b, the spatial distribution of TCC have great differences between the reconstructed precipitation anomaly fields with and without TSD. For the reconstructed fields of original precipitation anomaly, TCC are small in Northwest China, Northeast China, and a part of Southwest China, which fail the significance test at 95% level (Figure 4a). In contrast, TCC of the reconstructed precipitation anomaly fields with TSD have been significantly improved in these regions, although some stations in Northwest China still perform poorly (Figure 4b). Therefore, when using the first eight EOF modes for reconstruction, although the summer precipitation anomaly in East China, especially in the Yangtze–Huaihe River Valley, can be reproduced with or without TSD to a certain degree, the reconstructed fields after TSD can represent the characteristic of variation of precipitation in western and Northeast China.

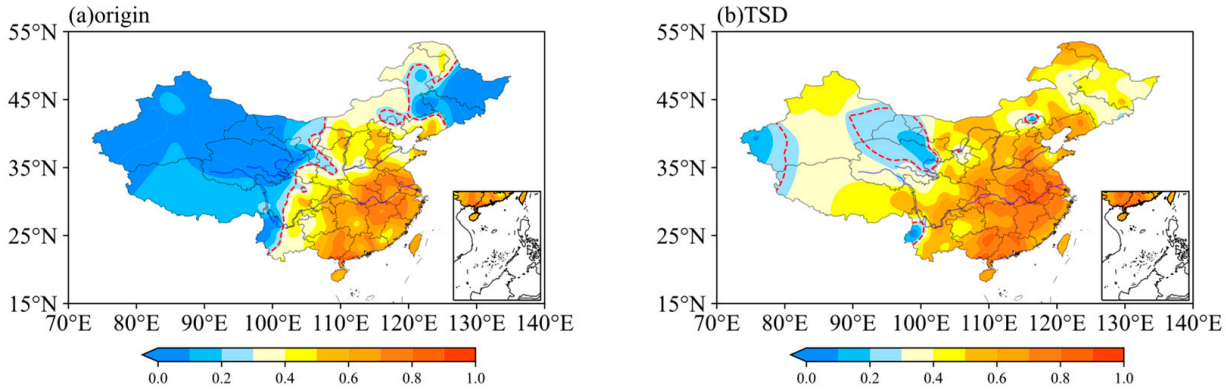

Figure 4. Spatial distribution of TCC between the reconstructed precipitation anomaly fields (**a**) without and (**b**) with TSD and the observed precipitation anomaly fields using the first eight EOF modes during 1961–2010 (the dashed line represents the 95% confidence test level).

In addition, the ACC of two reconstructed fields is calculated, respectively (Figure 5). It is not difficult to find that, except in 1961, 1982, 1989–1991, and 2010, the ACC of the reconstructed fields after TSD is higher than without TSD. The average ACC of the former between 1961 and 2010 is 0.67, while for the latter is 0.57, with the difference close to 0.1. ACC of the former range from 0.38 to 0.86, while the latter range from 0.35 to 0.82, and both the minimum and maximum value of the former are higher than the latter. Therefore, according to the comparison of TCC and ACC, it can be concluded that the reconstructed precipitation anomaly fields after TSD, which use the first eight EOF modes and time coefficients, can better restore the observed fields and provide a basis for the establishment of the prediction model.

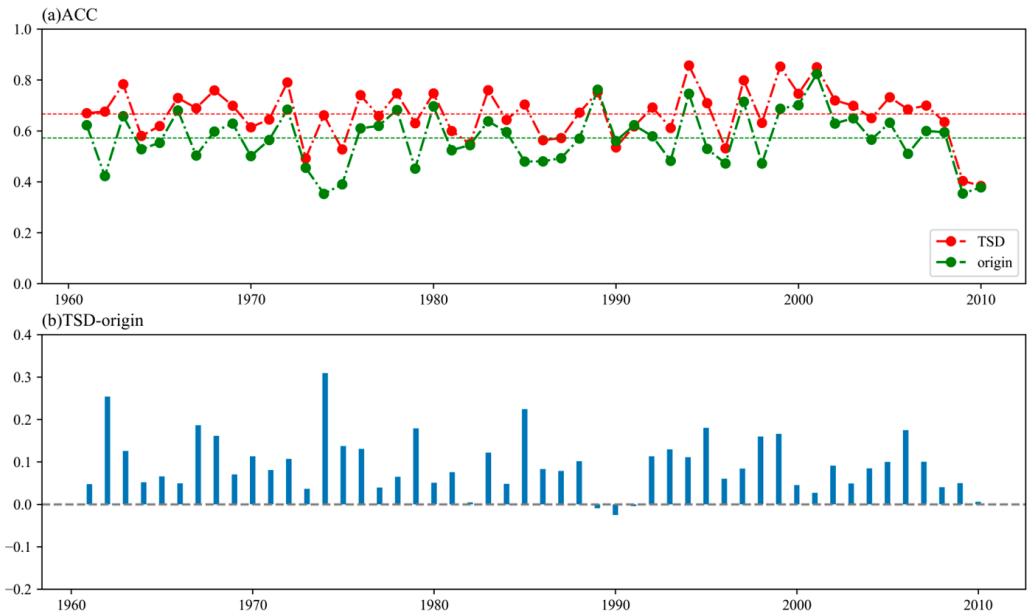

Figure 5. (**a**) Time series of ACC between the reconstructed precipitation anomaly fields without (green line) and with (red line) TSD and the observed precipitation anomaly fields using the first eight EOF modes and (**b**) their difference during 1961–2010.

Through the comparison and analysis of the reconstructed fields with and without the TSD method, in the following research, we take the time coefficients of the first eight EOF modes of summer precipitation in China as the predictands. According to the principle of EOF decomposition, the predicted fields can be represented as

$$\widetilde{Y}_i = \sum_{k=1}^{8} \widetilde{T}_{k,i} V_{k,i} \quad (2)$$

$\widetilde{Y}_i$ represents the predicted precipitation anomaly field for the i-th year, and $\widetilde{T}_{k,i} (1 \leq k \leq 8)$ represents the time coefficients of each mode for the i-th year, which will be predicted by the model. For the prediction model without TSD, the predicted field of this year is obtained by multiplying the predicted time coefficients with the spatial patterns of the first eight EOF modes. For the prediction model using the TSD method, on the interannual and interdecadal time scales, the predicted fields are obtained by multiplying the predicted time coefficients with the spatial patterns of the first eight EOF modes separately, and then they will be summed up to obtain the final predicted fields. Therefore, selecting suitable predictors for the prediction models before and after TSD, and achieving the prediction of the time coefficients, will be the key to the next section.

3.2. Preliminary Selection of Predictors

The result we need is the summer precipitation anomaly in China. Most work only considered annual signals when selecting its predictors. However, as mentioned above, many studies have shown that interdecadal signals may also play an important role in it. Therefore, when selecting predictors, it is beneficial to consider interannual and interdecadal signals at the same time to improve prediction skills. In addition, correlation analysis is easily disturbed by extreme climate events, which may affect the result of the prediction model. Liang [49] also pointed out that the correlation between two phenomena does not mean that there is a causal relationship, and two independent events affected by the same external factor often have a high correlation but no substantial causal relationship. Although the predictors obtained by correlation analysis have statistical or physical significance to some degree, they are only necessary conditions for predictand rather than sufficient conditions [36]. Therefore, we will preliminarily choose predictors based on the hindcast skill on each grid of the variable fields, rather than the simple correlation analysis between physical quantities.

First, according to the steps in Section 2, the linear function relationships between the variable fields, including the previous winter SST/SIC and spring U200/H500, and the time coefficients of the first eight EOF modes are fitted. Using these relationships, all grids on the variable fields can predict time coefficients for all modes every year. Then, the correlation coefficients between the observed and predicted time series are calculated, from which the grids with significant correlation (passing the confidence test level above 95%) will be selected as potential predictors, which is obviously different from the traditional correlation analysis calculating the correlation coefficients between the predictand and alternative variables fields or indices. The potential predictors are obtained on the basis of the test for the hindcast coefficients, and they are sufficient conditions for predictands. In addition, due to removing the data of the predicted year and adjacent two years before fitting, the possible impact of extreme climate events on the prediction model has been reduced. Therefore, the predictors selected preliminarily through the above process have a more stable statistical relationship with the time coefficients corresponding to each EOF mode.

Compared with other modes, the precursor signals of EOF 2 have a wider range and are clear to compare [36]. In addition, many previous studies have focused on the relationship between summer precipitation in China and ENSO and its associated circulation, while studies on SIC are less. Therefore, we take the time coefficients of EOF 2 and the previous winter SIC as an example to compare traditional correlation analysis with our method, focusing mainly on the differences of predictors before and after TSD.

First, we compare Figure 6a–c with Figure 6d–f generally. It can be seen that the area of our potential predictors is significantly smaller than the area with significant correlation. Before TSD, the area of SIC anomaly significantly related to the time coefficients is mainly located in the Barents–Kara Sea, and is also scattered near Baffin Bay and Greenland Sea (Figure 6a). However, the potential predictors for the time coefficients of EOF 2 only exist in the Barents Sea (Figure 6b). The differences between correlation analysis and potential

predictors are also reflected in the signals of interdecadal (Figure 6b,e) and interannual (Figure 6c,f) time scales. For the interdecadal time scale, comparing Figure 6b,e, it can be seen that the SIC anomaly near Baffin Bay, Hudson Bay, Novaya Zemlya, and the Chukchi Sea has a significantly positive correlation with the time coefficients of EOF 2, while there is significantly negative in the Greenland Sea (Figure 6b). However, there is no potential predictor for the time coefficients of EOF 2 in the Chukchi Sea, and the range of other regions is also obviously reduced (Figure 6e). For the interannual time scale, the time coefficients of EOF 2 also have a significant positive correlation with the SIC in the Barents Sea, while the correlation north of the Laptev Sea is significantly negative (Figure 6c). But the latter region failed to select potential predictors (Figure 6f). From the above analysis, for potential predictors, it can be found that there are differences before and after using the TSD method, and there are also obvious differences between the interannual and interdecadal components after TSD. This also indicates that the prediction of time coefficients on different time scales after TSD is meaningful for the improvement of prediction skills. In addition, the potential predictors selected by the test of hindcast coefficients are included in the region with a significant correlation between the time coefficients and the variable fields. The selection of potential predictors is stricter than traditional correlation analysis, and the previous SIC anomaly with significant correlation may not necessarily be potential predictors. These conclusions are the same for other modes and variables also (figure omitted).

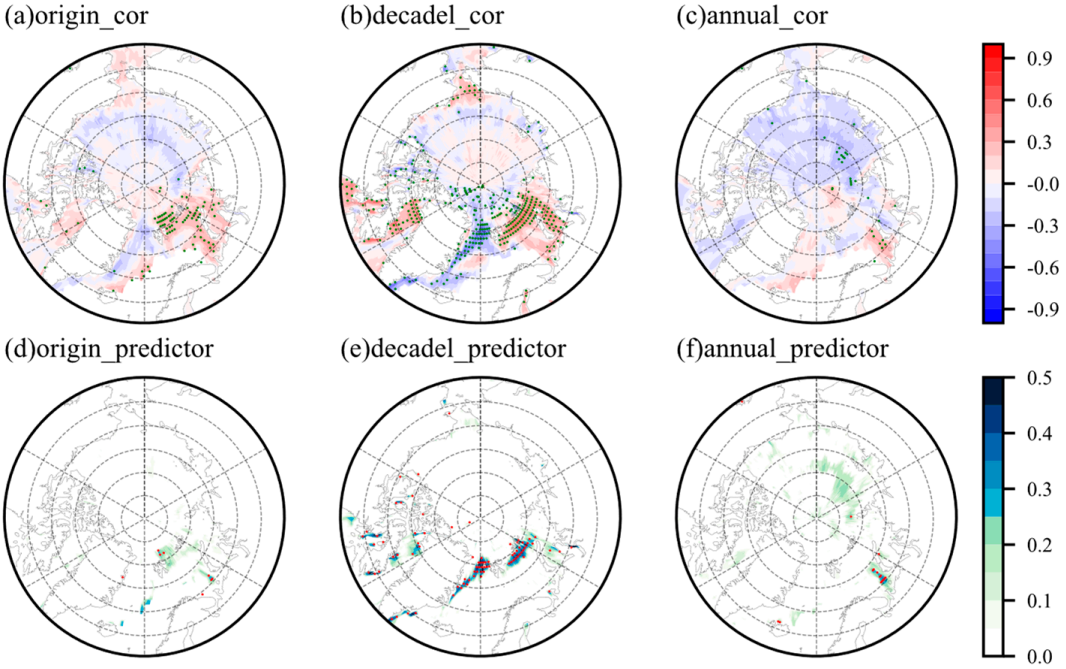

Figure 6. Spatial distribution of (**a**–**c**) TCC and (**d**–**f**) potential predictors between time coefficients of EOF 2 and the SIC anomaly in the previous winter during 1961–2010 (the dotted area represents passing the confidence test level above 95%).

3.3. Establishment of Prediction Model and Hindcast Test

After the preliminary selection of the predictors in the previous section, the precursor signals with significant correlation but poor prediction results have been removed. In this section, these potential predictors will be further tested. From Equations (1) and (2), it can be seen that the spatial distribution of EOF modes used for the reconstruction or prediction of precipitation anomaly fields are the same, and the differences lie only between their time coefficients. It is also widely known that the major characteristics of the precipitation fields can be represented by the EOF modes with high variance contributions, combining the positive or negative time coefficients. Therefore, here we first test the symbol consistency

rate between the predicted and observed time coefficients. The hindcast period is 50 years from 1961 to 2010, and the threshold of symbol consistency in this study is set to 60% by many tests, that is, 30 years. So, for the potential predictors selected in Section 3.2, if the predicted and observed time coefficients have the same symbol for 30 or even more years on the same grid, this grid will be retained, while other grids that do not meet this condition will be eliminated. Here, the predictors are selected based on the actual hindcast result once again. Sorting the hindcast time coefficients given by the above grids for the same year, the median of this series will be the final value predicted for this year, and the corresponding physical quantity will be the final predictor, which also reduces the subjectivity and avoids the overfitting problem that may exist in multiple regression prediction models. These steps are all performed on time coefficients of EOF modes for the interannual and interdecadal components after TSD, respectively.

Through the evaluation and selection of the potential predictors, we have structured prediction models of the time coefficients corresponding to the EOF leading modes of summer precipitation in China before and after TSD. Figure 7 shows the hindcast results of time coefficients for the first three EOF modes in 1961–2010. Since the predictand is the annual summer precipitation in China, here we mainly compare the hindcast results before TSD (Figure 7a,d,g) and interannual components of TSD (Figure 7b,e,h). For the prediction model without TSD, the correlation coefficients between the observed and predicted time series of the first three EOF modes are 0.39, 0.52, and 0.63, respectively. For the prediction model of the interannual component after TSD, the correlation coefficients between the observed and predicted time series are 0.53, 0.53, and 0.52, respectively. Therefore, the prediction accuracy of time coefficients has been improved after TSD for the first two modes, especially for the first mode with the largest variance contribution; its correlation coefficient has increased by nearly 0.15. Although the correlation coefficient of the third mode has decreased, the predicted precipitation anomaly field we hope to receive is obtained by combining two components of different time scales for the first eight EOF modes, which will be analyzed and tested later.

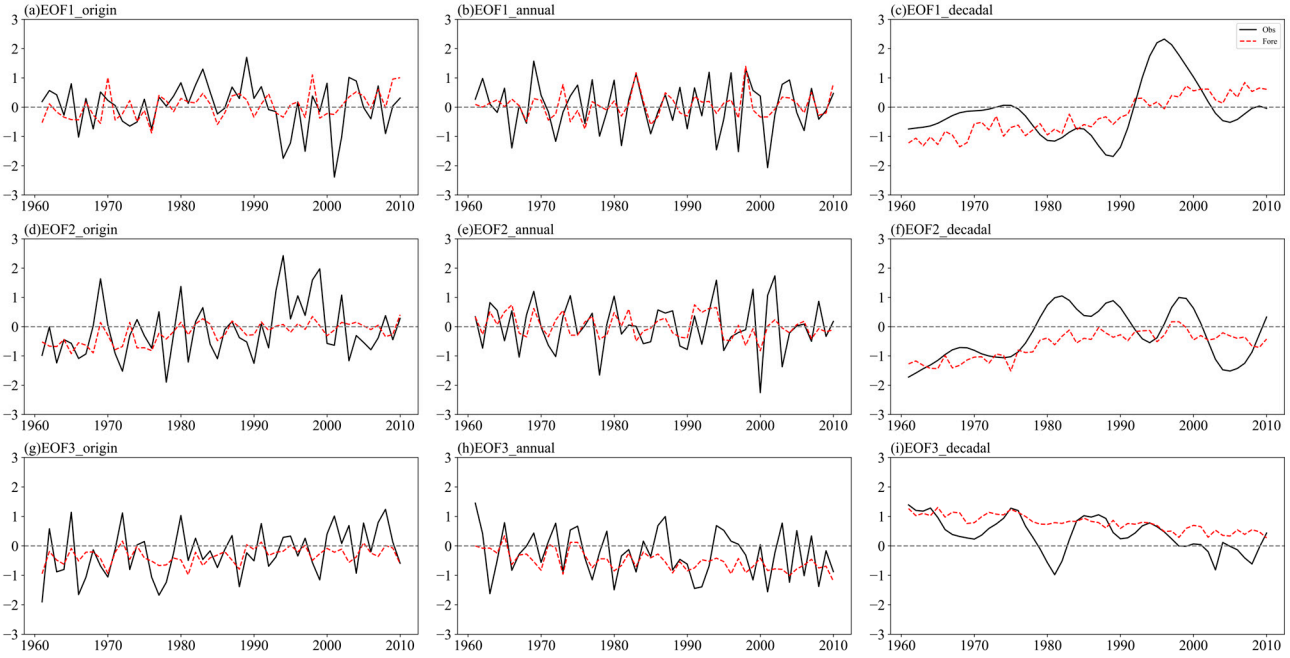

Figure 7. Observed (black solid line) and predicted (red dotted line) time coefficients of the first three EOF modes (the left column represents non-TSD, the middle column represents the interannual component of TSD, and the right column represents the interdecadal component of TSD).

So far, we have completed the prediction of time coefficients. The ultimate goal we want to achieve is to predict the summer precipitation anomaly fields in China. Therefore,

according to Equations (1) and (2), we need to multiply the predicted time coefficients of the first eight EOF modes with their spatial patterns to obtain the final result, and compare the prediction skill before and after TSD. It can be seen from Figure 8a,b that there are also significant differences in TCC's spatial distribution with and without TSD for the predicted fields. The hindcast fields without TSD only have high prediction skills in some areas of Central China, East China, and South China, while the rest regions all fail to pass the significance test at the 95% level. For the predicted fields after TSD, the region with high TCC is more expanded than the former, and a small number of stations in Xinjiang Province and Northeast China also represent high TCC. Therefore, after TSD, the predicted fields have higher prediction skills in TCC test.

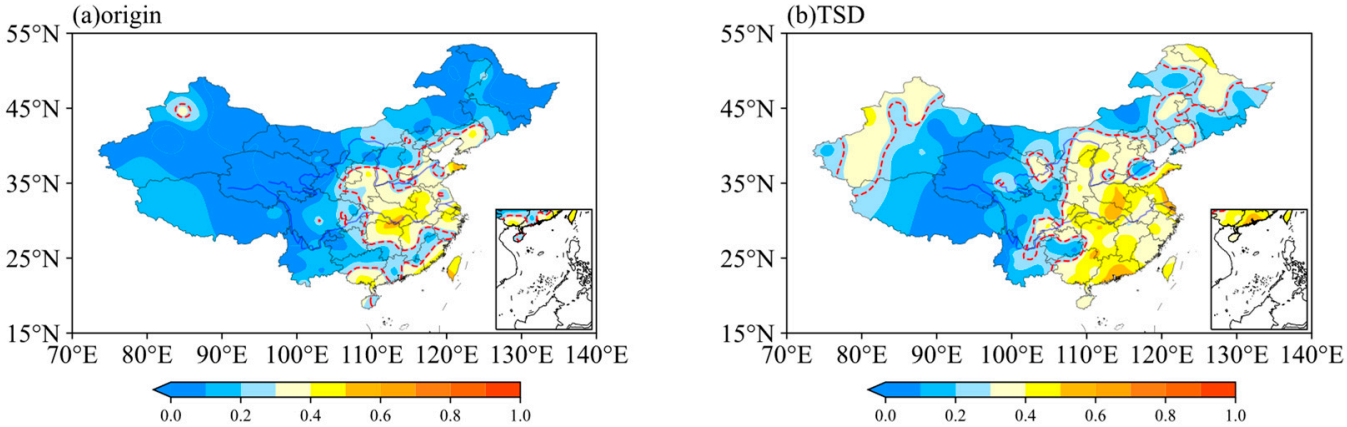

Figure 8. Same as Figure 4, but for the predicted precipitation anomaly fields in the hindcast test (the dashed line represents the 95% confidence test level).

We also calculated the ACC of predicted fields before and after TSD (Figure 9). During the hindcast period, the ACC of the predicted fields without TSD ranged from −0.087 to 0.61, with an average of 0.30, while the ACC of the predicted fields after TSD ranged from −0.086 to 0.67, with an average of 0.37. It can be seen from the difference between them in 50 years, the ACC of the predicted fields with TSD is higher than that without TSD for 30 years. Conversely, the ACC of the predicted fields before TSD is higher than after for only 13 years (the differences are so small in other years). Through the test of ACC, it can also be concluded that the prediction model performed better after TSD for summer precipitation anomaly in China.

3.4. Result of Independent Sample Test

In addition, we also tested the predicted precipitation anomaly fields from 2011 to 2020 using the prediction model for an independent sample test (Figure 10). It is seen that although the ACC of the forecast model without TSD is higher than that with TSD in several years, its prediction skill in 10 years is unstable, ranging from −0.38 to 0.37, and the average is only −0.002 (i.e., negative skill). Especially in 2011, 2015, and 2020, ACC was very low. In contrast, the ACC of the forecast model after TSD is positive in almost all years except three years close to 0. The overall performance is stable, and the average is 0.10, more than 0.1 higher than the forecast model without TSD. Therefore, it is not difficult to find that the TSD method can effectively improve the prediction skill of the original forecast model without TSD in an independent sample test.

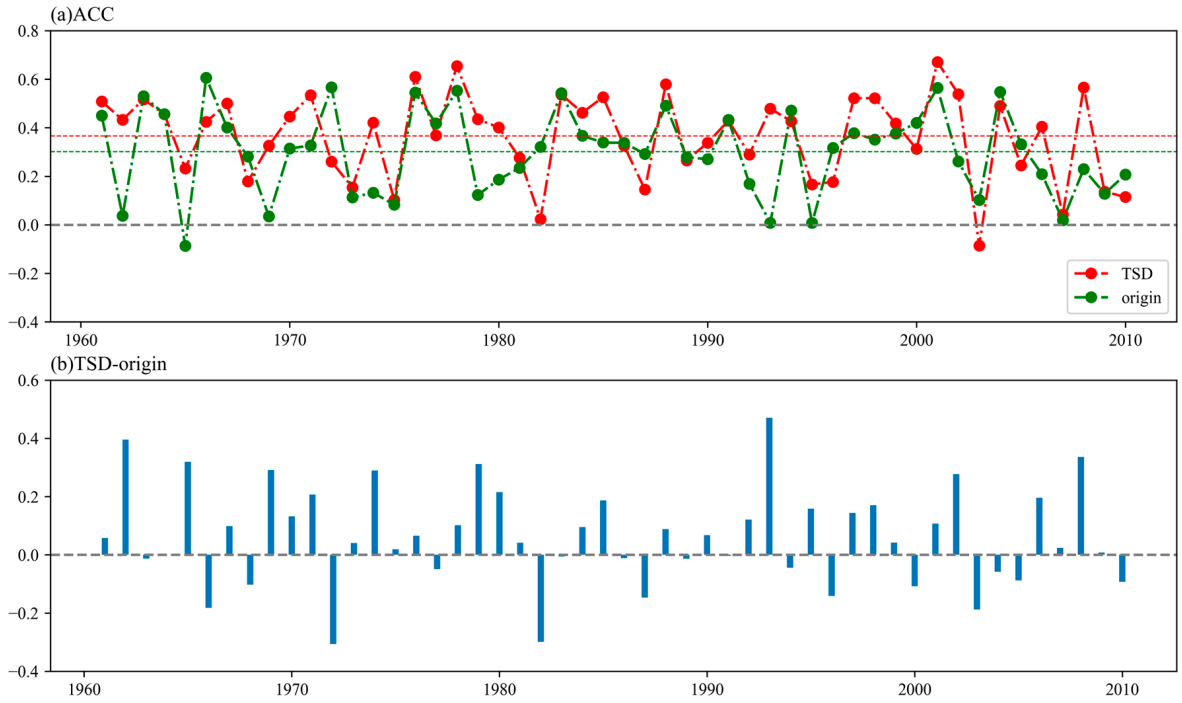

Figure 9. Same as Figure 5, but for the predicted precipitation anomaly fields in the hindcast test.

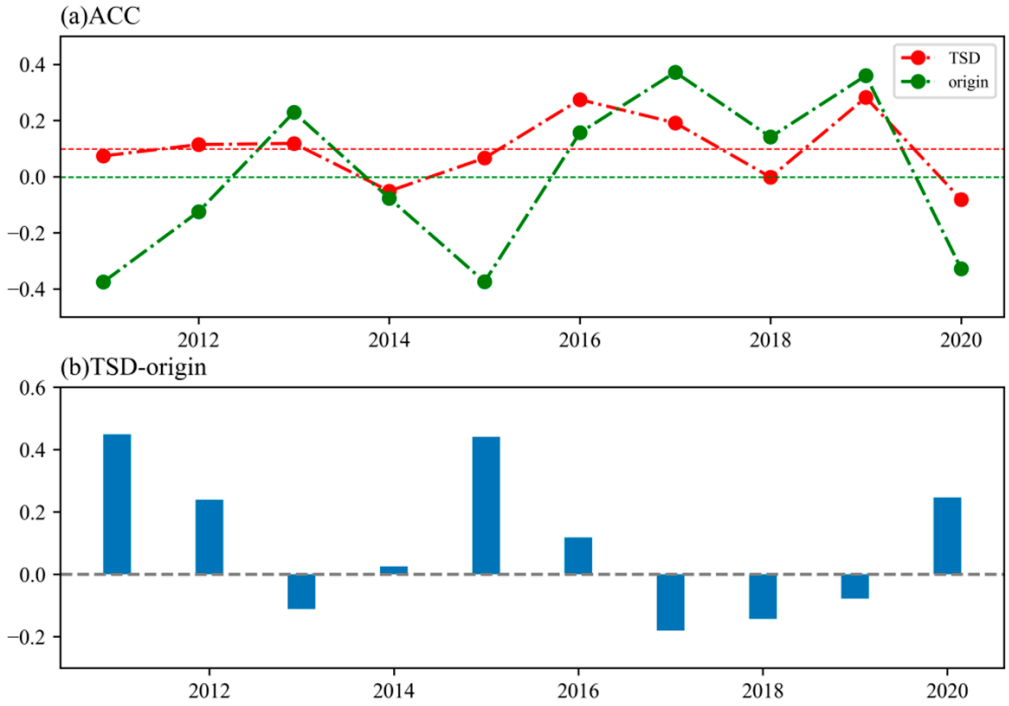

Figure 10. Same as Figures 5 and 9, but for the precipitation anomaly fields in the independent sample.

4. Discussion and Conclusions

The summer precipitation in China is influenced by forcing signals on multiple time scales and exhibits interannual and interdecadal variation characteristics, which make its prediction difficult. Therefore, it is necessary to divide the precipitation into interannual and interdecadal time scales and predict them separately. This is the time-scale decomposition (TSD) method used in our study. In addition, the time coefficients corresponding to the EOF modes are taken as the predictands, and the predictors are selected from variable fields according to the hindcast result of every grid. Then, statistical prediction models

are established for the summer precipitation anomalies in China before and after using the TSD method, and their prediction abilities are compared through the hindcast test and independent sample test. The main conclusions are as follows:

(1) According to the nonlinear growth characteristics of the ACC and spatial variance ratio of precipitation anomaly fields reconstructed by different numbers of EOF modes, the first eight EOF modes and their time coefficients are selected to reduce the possible instability impact on the prediction models. Through the test of TCC, it can be found that the precipitation anomaly in East China can be represented to a certain degree when the first eight modes are used for reconstruction, whether or not TSD is carried out. But the reconstructed fields after TSD have a better reappearance ability in West and Northeast China. ACC shows that the minimum and maximum values of the reconstructed fields after TSD are higher than those without TSD, and the difference between their average is close to 0.1. This comparison shows that the reconstructed fields with TSD can better represent the precipitation anomaly;

(2) The potential predictors are preliminarily selected for the time coefficients of the first eight EOF modes from the variable fields. It is based on the test result of hindcast time coefficients. Before TSD, the SIC anomaly significantly correlated with the time coefficients of EOF 2 are mainly located in the Barents–Kara Sea, while the potential predictors only existed in the Barents Sea. On the interdecadal time scale, the SIC anomaly in Baffin Bay, Hudson Bay, near New Earth Island, and the Chukchi Sea are significantly positively correlated with the time coefficients of EOF 2, but there is no potential predictor in the Chukchi Sea. On the interannual time scale, the SIC anomaly is significantly negatively correlated with the time coefficients of EOF 2 in the north of the Laptev Sea, but no potential predictor has been found here. Similar to SIC, it can be seen in H500, U200, and SST that there are differences in potential predictors not only before and after TSD, but also for interannual and interdecadal components after TSD. Through the hindcast of the time coefficients by these physical variable fields, we clearly see the effect of the TSD method in distinguishing predictors on different time scales;

(3) By constraining the threshold of symbol consistency between the predicted and observed time coefficients, the potential predictors are further selected, and then the final predicted time coefficients are defined as the median of the predicted values. The hindcast results of the time coefficients for the first three EOF modes show that the prediction accuracy has been improved after TSD. Based on the predicted time coefficients, the predicted fields are reconstructed and tested. In the hindcast test, the predicted fields without TSD only have high TCC in some areas of Central China, East China, and South China, while they have a wider range after TSD. And ACC of the predicted fields after TSD is higher than that without TSD in most years. In the independent sample test, the prediction skill after TSD is basically positive and more stable, and the average ACC has increased by more than 0.1. Therefore, TSD can improve the prediction ability of the original prediction model.

It needs to be noticed that China has a vast territory, multiple underlying surface conditions, and topographies such as Tibet Plateau, North China Plain, and Sichuan Basin, so the climate presents complex regional characteristics, especially the precipitation exhibits obvious locality. Chen et al. [50] used monthly precipitation station data set to decompose China into 13 regions using the Rotated Empirical Orthogonal Function method. This result can not only identify the climate characteristics of eastern China, but also central and western China in detail, which is conducive to searching for predictors affecting regional climate. On the basis of this division, Gu et al. [51] used the downscaling method to establish a precipitation prediction model, which is applicable to most regions of China. The predictors selected are independent and have physical significance, thereby improving the prediction skill of regional precipitation anomaly. Considering the results of our study, although precipitation prediction skill has significantly improved in some regions after TSD, it can be seen that there is still poor prediction skill in the other regions like Qinghai and Inner Mongolia Province, which may be due to the significant difference in climate characteristics

of precipitation between them and other regions in China. Therefore, selecting different predictors and establishing prediction models for precipitation in different regions can help improve their predictability.

In addition, we select the predictors from the reanalysis data set of the preceding winter and spring and establish statistical prediction models, and if the variable fields of the same summer can be used, it will help to improve the prediction skill, too. At present, dynamic climate models are also developing rapidly, and they already have high prediction ability for large-scale circulation fields. However, due to problems in model resolution and parameterization of physical processes, the prediction of regional precipitation by dynamic models still cannot satisfy the actual demand [52,53]. Therefore, statistical downscaling methods are proposed and widely applied [54–56]. The statistical relationships between large-scale circulation affecting regional climate and predictands are established by years of observation data. And then, the large-scale variables output by the dynamic model will be substituted into these relationships to predict climate elements. So how to utilize the large-scale circulation fields of the same summer from the dynamic model, based on the prediction model proposed in this study that combines EOF decomposition and TSD method, to improve the prediction ability for summer precipitation anomaly in different regions of China, will become the next problem to solve.

Author Contributions: Conceptualization, Y.Z.; methodology, Y.Z. and Y.H. (Yijia Hu); validation, Y.H. (Yao Ha) and Z.Z.; formal analysis, Z.X.; investigation, Z.X.; data curation, Y.H. (Yijia Hu); writing—original draft preparation, Z.X.; writing—review and editing, Y.Z. and Y.H. (Yijia Hu); supervision, Z.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported jointly by the National Key R&D Program of China, grant number 2022YFF0801702, and the Jiangsu Collaborative Innovation Center for Climate Change.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: The authors would like to extend their sincere appreciation to the National Key R&D Program of China and the Jiangsu Collaborative Innovation Center for Climate Change.

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

References

1. Huang, R.; Sun, F. Impacts of the Tropical Western Pacific on the East Asian Summer Monsoon. *J. Meteorol. Soc. Jpn. Ser. II* **1992**, *70*, 243–256. [CrossRef]
2. Ding, Y.; Wang, Z.; Sun, Y. Inter-decadal Variation of the Summer Precipitation in East China and Its Association with Decreasing Asian Summer Monsoon. Part I: Observed Evidences. *Int. J. Climatol.* **2008**, *28*, 1139–1161. [CrossRef]
3. Wang, L.; Qian, Y.; Zhang, Y.; Zhao, C.; Leung, L.R.; Huang, A.; Xiao, C. Observed Variability of Summer Precipitation Pattern and Extreme Events in East China Associated with Variations of the East Asian Summer Monsoon. *Int. J. Climatol.* **2016**, *36*, 2942–2957. [CrossRef]
4. Chen, F.; Chen, J.; Huang, W. Weakened East Asian Summer Monsoon Triggers Increased Precipitation in Northwest China. *Sci. China Earth Sci.* **2021**, *64*, 835–837. [CrossRef]
5. Wu, B.; Zhou, T. Oceanic Origin of the Interannual and Interdecadal Variability of the Summertime Western Pacific Subtropical High. *Geophys. Res. Lett.* **2008**, *35*, L13701. [CrossRef]
6. Ye, T.-S.; Zhi, R.; Zhao, J.-H.; Gong, Z.-Q. The Two Annual Northward Jumps of the West Pacific Subtropical High and Their Relationship with Summer Rainfall in Eastern China under Global Warming. *Chin. Phys. B* **2014**, *23*, 069203. [CrossRef]
7. Zhang, Q.; Zheng, Y.; Singh, V.P.; Luo, M.; Xie, Z. Summer Extreme Precipitation in Eastern China: Mechanisms and Impacts. *J. Geophys. Res. Atmos.* **2017**, *122*, 2766–2778. [CrossRef]
8. Yu, R.; Zhai, P. Changes in Summer Persistent Precipitation over the Middle-Lower Reaches of the Yangtze River and Associated Atmospheric Circulation Patterns. *J. Meteorol. Res.* **2021**, *35*, 393–401. [CrossRef]
9. Huang, R.; Wu, Y. The Influence of ENSO on the Summer Climate Change in China and Its Mechanism. *Adv. Atmos. Sci.* **1989**, *6*, 21–32.

10. Zhang, R.; Sumi, A.; Kimoto, M. A Diagnostic Study of the Impact of El Niño on the Precipitation in China. *Adv. Atmos. Sci.* **1999**, *16*, 229–241. [CrossRef]
11. Feng, J.; Chen, W.; Tam, C.-Y.; Zhou, W. Different Impacts of El Niño and El Niño Modoki on China Rainfall in the Decaying Phases. *Int. J. Climatol.* **2011**, *31*, 2091–2101. [CrossRef]
12. Han, T.; Wang, H.; Sun, J. Strengthened Relationship between Eastern ENSO and Summer Precipitation over Northeastern China. *J. Clim.* **2017**, *30*, 4497–4512. [CrossRef]
13. Feng, J.; Wang, L.; Chen, W. How Does the East Asian Summer Monsoon Behave in the Decaying Phase of El Niño during Different PDO Phases? *J. Clim.* **2014**, *27*, 2682–2698. [CrossRef]
14. Song, W.-J.; Wang, Q.-G. CP El Niño and PDO Variability Affect Summer Precipitation over East China. *Adv. Meteorol.* **2020**, *2020*, 2747194. [CrossRef]
15. Qin, M.; Dai, A.; Li, D.; Hua, W. Understanding the Inter-decadal Variability of Autumn Precipitation over North Central China Using Model Simulations. *Int. J. Climatol.* **2020**, *40*, 874–886. [CrossRef]
16. Zhou, Z.-Q.; Xie, S.-P.; Zhang, R. Historic Yangtze Flooding of 2020 Tied to Extreme Indian Ocean Conditions. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2022255118. [CrossRef] [PubMed]
17. Wu, B.; Zhang, R.; Wang, B.; D'Arrigo, R. On the Association between Spring Arctic Sea Ice Concentration and Chinese Summer Rainfall. *Geophys. Res. Lett.* **2009**, *36*, 666–678. [CrossRef]
18. Liu, Y.; Zhu, Y.; Wang, H.; Gao, Y.; Sun, J.; Wang, T.; Ma, J.; Yurova, A.; Li, F. Role of Autumn Arctic Sea Ice in the Subsequent Summer Precipitation Variability over East Asia. *Int. J. Climatol.* **2020**, *40*, 706–722. [CrossRef]
19. Tian, Y.; Gao, Y.; Guo, D. The Relationship between Melt Season Sea Ice over the Bering Sea and Summer Precipitation over Mid-Latitude East Asia. *Adv. Atmos. Sci.* **2021**, *38*, 918–930. [CrossRef]
20. Lu, B.; Li, H.; Wu, J.; Zhang, T.; Liu, J.; Liu, B.; Chen, Y.; Baishan, J. Impact of El Niño and Southern Oscillation on the Summer Precipitation over Northwest China. *Atmos. Sci. Lett.* **2019**, *20*, e928. [CrossRef]
21. Ma, Q.; Li, Z.; Lei, H.; Chen, Z.; Liu, J.; Wang, S.; Su, T.; Feng, G. Interannual Variability of Extreme Precipitation during the Boreal Summer over Northwest China. *Remote Sens.* **2023**, *15*, 785. [CrossRef]
22. Yang, Q.; Ma, Z.; Fan, X.; Yang, Z.-L.; Xu, Z.; Wu, P. Decadal Modulation of Precipitation Patterns over Eastern China by Sea Surface Temperature Anomalies. *J. Clim.* **2017**, *30*, 7017–7033. [CrossRef]
23. Zhang, Z.; Sun, X.; Yang, X.-Q. Understanding the Interdecadal Variability of East Asian Summer Monsoon Precipitation: Joint Influence of Three Oceanic Signals. *J. Clim.* **2018**, *31*, 5485–5506. [CrossRef]
24. Fan, K.; Wang, H.; Choi, Y.-J. A Physically-Based Statistical Forecast Model for the Middle-Lower Reaches of the Yangtze River Valley Summer Rainfall. *Chin. Sci. Bull.* **2008**, *53*, 602–609. [CrossRef]
25. Hu, Y.; Zhong, Z.; Zhu, Y.; Ha, Y. A Statistical Forecast Model Using the Time-Scale Decomposition Technique to Predict Rainfall during Flood Period over the Middle and Lower Reaches of the Yangtze River Valley. *Theor. Appl. Climatol.* **2018**, *132*, 479–489. [CrossRef]
26. Liu, N.; Li, S. Predicting Summer Rainfall over the Yangtze–Huai Region Based on Time-Scale Decomposition Statistical Downscaling. *Weather Forecast* **2014**, *29*, 162–176. [CrossRef]
27. Guo, Y.; Li, J.; Li, Y. A Time-Scale Decomposition Approach to Statistically Downscale Summer Rainfall over North China. *J. Clim.* **2012**, *25*, 572–591. [CrossRef]
28. Song, L.; Duan, W.; Li, Y.; Mao, J. A Timescale Decomposed Threshold Regression Downscaling Approach to Forecasting South China Early Summer Rainfall. *Adv. Atmos. Sci.* **2016**, *33*, 1071–1084. [CrossRef]
29. Pearson, K. On Lines and Planes of Closest Fit to Systems of Points in Space. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **1901**, *2*, 559–572. [CrossRef]
30. Lorenz, E. Empirical orthogonal functions and statistical weather prediction. In *Statistical Forecasting Project*; Science Report 1; MIT (NTIS AD 110268); Department of Meteorology: Cambridge, MA, USA, 1956; p. 49.
31. Nguyen, H.; Mehrotra, R.; Sharma, A. Correcting for Systematic Biases in GCM Simulations in the Frequency Domain. *J. Hydrol.* **2016**, *538*, 117–126. [CrossRef]
32. Aryal, Y.; Zhu, J. Multimodel Ensemble Projection of Meteorological Drought Scenarios and Connection with Climate Based on Spectral Analysis. *Int. J. Climatol.* **2020**, *40*, 3360–3379. [CrossRef]
33. Xing, W.; Wang, B.; Yim, S.-Y. Long-Lead Seasonal Prediction of China Summer Rainfall Using an EOF–PLS Regression-Based Methodology. *J. Clim.* **2016**, *29*, 1783–1796. [CrossRef]
34. Ma, J.; Sun, J. New Statistical Prediction Scheme for Monthly Precipitation Variability in the Rainy Season over Northeastern China. *Int. J. Climatol.* **2021**, *41*, 5805–5819. [CrossRef]
35. Huang, Y.; Wang, H.; Zhang, P. A Skillful Method for Precipitation Prediction over Eastern China. *Atmos. Ocean. Sci. Lett.* **2022**, *15*, 29–36. [CrossRef]
36. Liu, B.; Zhu, C. Potential Skill Map of Predictors Applied to the Seasonal Forecast of Summer Rainfall in China. *J. Appl. Meteorol. Sci.* **2020**, *31*, 570–582.
37. Nie, Y.; Sun, J. Causes of Interannual Variability of Summer Precipitation Intraseasonal Oscillation Intensity over Southwest China. *J. Clim.* **2022**, *35*, 3705–3723. [CrossRef]
38. Sun, B.; Wang, H. Enhanced Connections between Summer Precipitation over the Three-River-Source Region of China and the Global Climate System. *Clim. Dyn.* **2019**, *52*, 3471–3488. [CrossRef]

39. Xu, H.; Chen, H.; Wang, H. Interannual Variation in Summer Extreme Precipitation over Southwestern China and the Possible Associated Mechanisms. *Int. J. Climatol.* **2021**, *41*, 3425–3438. [CrossRef]
40. Tang, Y.; Duan, A.; Hu, D. Influence of the Antarctic Oscillation on Summer Precipitation over East Asia. *Atmos. Res.* **2023**, *292*, 106847. [CrossRef]
41. Zhang, W.; Hari, V.; Wang, S.-Y.; LaPlante, M.D.; Garfin, G.; Affram, G.; Kumar, R. Fewer Troughs, Not More Ridges, Have Led to a Drying Trend in the Western United States. *Geophys. Res. Lett.* **2022**, *49*, e2021GL097089. [CrossRef]
42. Zhang, W.; Villarini, G. On the Weather Types That Shape the Precipitation Patterns across the U.S. Midwest. *Clim. Dyn.* **2019**, *53*, 4217–4232. [CrossRef]
43. Bibi, A.; Ullah, K.; Yushu, Z.; Wang, Z.; Gao, S. Role of Westerly Jet in Torrential Rainfall during Monsoon over Northern Pakistan. *Earth Space Sci.* **2020**, *7*, e2019EA001022. [CrossRef]
44. Hong, X.; Lu, R.; Li, S. Interannual Relationship between the West Asian and East Asian Jet Meridional Displacements in Summer. *J. Clim.* **2021**, *34*, 621–633. [CrossRef]
45. Kalnay, E.; Kanamitsu, M.; Kistler, R.; Collins, W.; Deaven, D.; Gandin, L.; Iredell, M.; Saha, S.; White, G.; Woollen, J.; et al. The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Am. Meteorol. Soc.* **1996**, *77*, 437–471. [CrossRef]
46. Rayner, N.A. Global Analyses of Sea Surface Temperature, Sea Ice, and Night Marine Air Temperature since the Late Nineteenth Century. *J. Geophys. Res.* **2003**, *108*, 4407. [CrossRef]
47. Zhang, J.; Sun, Z.; Lan, G. Weather statistical analysis of Asian 500 mb mean circulations. *J. Nanjing Inst. Meteorol.* **1979**, *1*, 33–41.
48. Zhang, J.; Sun, Z.; Chen, S. On the stability of empirical orthogonal function (EOF). *Acta. Meteorol. Sin.* **1981**, *39*, 82–88.
49. Liang, X.S. Unraveling the Cause-Effect Relation between Time Series. *Phys. Rev. E* **2014**, *90*, 052150. [CrossRef]
50. Chen, L.J.; Chen, D.L.; Wang, H.J.; Yan, J.H. Regionalization of Precipitation Regimes in China. *Atmos. Ocean. Sci. Lett.* **2009**, *5*, 301–307.
51. Gu, W.Z.; Chen, L.J.; Li, W.J.; Chen, D.L. Development of a Downscaling Method in China Regional Summer Precipitation Prediction. *Acta. Meteorol. Sin.* **2011**, *3*, 303–315. [CrossRef]
52. Wang, B.; Lee, J.-Y.; Kang, I.-S.; Shukla, J.; Park, C.-K.; Kumar, A.; Schemm, J.; Cocke, S.; Kug, J.-S.; Luo, J.-J.; et al. Advance and Prospectus of Seasonal Prediction: Assessment of the APCC/CliPAS 14-Model Ensemble Retrospective Seasonal Prediction (1980–2004). *Clim. Dyn.* **2009**, *33*, 93–117. [CrossRef]
53. Paul, S.; Liu, C.M.; Chen, J.M.; Lin, S.H. Development of a Statistical Downscaling Model for Projecting Monthly Rainfall over East Asia from a General Circulation Model Output. *J. Geophys. Res.* **2008**, *113*, D15117. [CrossRef]
54. Liu, Y.; Ren, H.-L. A Hybrid Statistical Downscaling Model for Prediction of Winter Precipitation in China. *Int. J. Climatol.* **2015**, *35*, 1309–1321. [CrossRef]
55. Lu, Z.; Guo, Y.; Zhu, J.; Kang, N. Seasonal Forecast of Early Summer Rainfall at Stations in South China Using a Statistical Downscaling Model. *Weather Forecast* **2020**, *35*, 1633–1643. [CrossRef]
56. Wang, N.; Ren, H.-L.; Liu, Y.; Deng, Y.; Meng, X.; Wu, J.; Zhou, F. Multi-Predictor Ensembles Improving Seasonal Prediction of Summer Rainfall over the Bohai Sea Rim Based on Statistical Downscaling of BCC_CSM1.1 m. *Atmos. Res.* **2022**, *275*, 106221. [CrossRef]