Article

Not peer-reviewed version

An Enhanced Neural Network Forecasting System for the July Precipitation over the Middle-Lower Reaches of the Yangtze River

Wenyan Liu and Xiangjun Shi *

Posted Date: 27 January 2025

doi: 10.20944/preprints202501.1982.v1

Keywords: climate forecasting system; neural network; improving techniques; stability and interpretability

doi:10.20944/preprints202501.1982.v1

Article

An Enhanced Neural Network Forecasting System for the July Precipitation over the Middle-Lower Reaches of the Yangtze River

Wenyan Liu and Xiangjun Shi *

School of Atmospheric Sciences, Nanjing University of Information Science and Technology, Nanjing 210044, China
* Correspondence: shixj@nuist.edu.com

Abstract: Forecasting July precipitation using prophase winter sea surface temperature through a non-linear machine learning model remains challenging. Given the scarcity of observed samples and more attention should be paid to anomalous precipitation events, the shallow neural network (NN) and several improving techniques are employed to establish the statistical forecasting system. To enhance the stability of predicted precipitation, the final output precipitation is an ensemble of multiple NN models with optimal initial seeds. The precipitation data from anomalous years are amplified to focus on anomalous events than normal events. Some artificial samples are created based on the relevant background theory to mitigate the problem of insufficient sample size for model training. Sensitivity experiments indicate that the above techniques could improve the stability and interpretability of the forecasting system. Rolling forecasts further indicate that the forecasting system is robust and half of the anomalous events can be successfully predicted. These improving techniques used in this study can be applied not only to the precipitation over the middle-lower reaches of the Yangtze River but also to other climate events.

Keywords: climate forecasting system; neural network; improving techniques; stability and interpretability

1. Introduction

The middle-lower reaches of the Yangtze River (MLRYR) are densely populated and economically developed regions of China. Unfortunately, these areas are also characterized by significant summer rainfall variability and frequent occurrences of hydrometeorological extremes [1–3]. Therefore, enhancing the predictive capability for anomalous summer precipitation over MLRYR is essential for reducing the adverse impacts of droughts and floods.

Previous studies have provided some theoretical knowledge and useful experiences such as follows: the preceding winter sea surface temperature (SST) is the most important indicator for the forecasting of summer precipitation [4–8], subdividing the summer season into distinct time segments could improve precipitation forecasts [9–11], and statistical methods still play a pivotal role in climate forecasts at lead times longer than one season [12–14]. It is noteworthy that previous statistical forecasting systems usually employ linear regression models to capture the relations between precipitation and SST [15–17], while the inherent complexity and nonlinear relations pose challenges for linear regression [18–23]. In recent years, machine-learning methods, which can capture complex nonlinear relationships, have been widely used in climate forecasts [24–28]. Hence, this study chooses one common-used machine-learning model to forecast the July precipitation over MLRYR based on the preceding winter SST. Meanwhile, some other techniques are also employed to improve the forecasting system.

Different from other machine-learning application fields, climate forecasts have the following unique characteristics. Firstly, considering service requirements, more attention should be paid to

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 January 2025 doi:10.20944/preprints202501.1982.v1

anomalous events rather than normal events [29–31]. Secondly, the sample size is considerably constrained (usually less than 100, far different from big data) because the observation period is relatively short (usually less than 100 years). Therefore, there are not sufficient observed samples for effective model training, thereby inducing instability of the trained model [32,33]. Lastly, there has been a great deal of background knowledge, particularly with regard to anomalous events [34–37]. For instance, an El Niño event might trigger a western Pacific anticyclone. This then strengthens the subtropical high-pressure system and boosts moisture transport from the South China Sea to MLRYR, increasing the chance of heavy rainfall [38–40]. This characteristic could be used for developing forecasting models, such as checking the model's interpretability. In this study, the above unique characteristics of climate forecasts would be considered and utilized.

The organization of this article is as follows. Section 2 introduces the forecasting system developed by this study. Section 3 analyzes the experiment results from this forecasting system. Section 4 discusses the experiences got from this study, and Section 5 provides the conclusions.

2. Data and Methods

Here, the July precipitation is predicted solely based on prophase winter SST. Considering the small observed sample size, the machine-learning framework in the forecasting system should be simple (i.e., the number of model's parameters should be small). Hence, a shallow neural network (NN) framework, comprising dozens of parameters, is employed to establish the forecasting system. As a consequence, the input information for the NN framework can only encompass a few variables extracted from the global SST field.

2.1. Input Data

The July precipitation over MLRYR is calculated from the 160 stations monthly precipitation data provided by the National Climate Center of the China Meteorological Administration. The area of MLRYR spans from 28° to 33.5°N and 113° to 122°E [41,42], including 19 stations. The precipitation anomaly percentage which exceeds 25% or falls below −25% is taken as an anomalous event. From 1951 to 2023, there are 19 high anomalous years and 23 low anomalous years. To enhance the effect of these anomalous years in the model training process, the precipitation anomaly percentages of these years are amplified by 25%, as shown in Figure 1. The amplified percentages of 1954 and 1969, which exceed 100%, are capped within 100% to avoid excessive influence from these two years. At last, it is noteworthy that the precipitation anomaly percentage is calculated based on the 21-point (annual) moving average. In other words, there is no decadal-scale trend in this data. Hereafter, the July precipitation anomaly percentage is referred to as Precipitation, which is convenient for expression. Similarly, the scaled Precipitation, which would be used for establishing forecast model, is referred to as Rain.

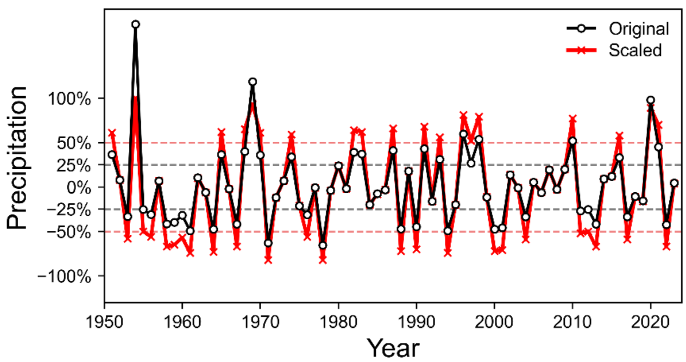


Figure 1. The original precipitation anomaly percentage (black line) alongside the scaled precipitation anomaly percentage (red line). Dashed lines indicate the threshold values for anomalous events.

 doi:10.20944/preprints202501.1982.v1

The IRI/LDEO Climate Data Library provides a dataset of monthly SST anomaly dataset with a spatial resolution of 5°×5° from 1856 to 2023 [43]. This study calculates the winter SST anomalies based on this dataset. Moreover, the decadal-scale trend, which is not conducive to climate forecasts, is removed through a 21-point (annual) moving average approach [44,45]. Regarding the precipitation forecast, the global SST fields include not only useful indicative signals but also useless signals. Figure 2 shows the correlations between the detrended prophase winter SST anomalies (hereafter referred to as SST) and the Rain. It is clear that useful indicative signals mostly come from the Pacific and Indian Oceans (30°E to 90°W, 40°S to 60°N). Thus, only the SST over these regions is used for precipitation forecasts. To let as few variables as possible describe the SST field, the gridded SST data are decomposed into 20 principal components via Empirical Orthogonal Function (EOF) orthogonal decomposition. The 20 corresponding time series serve as potential predictors.

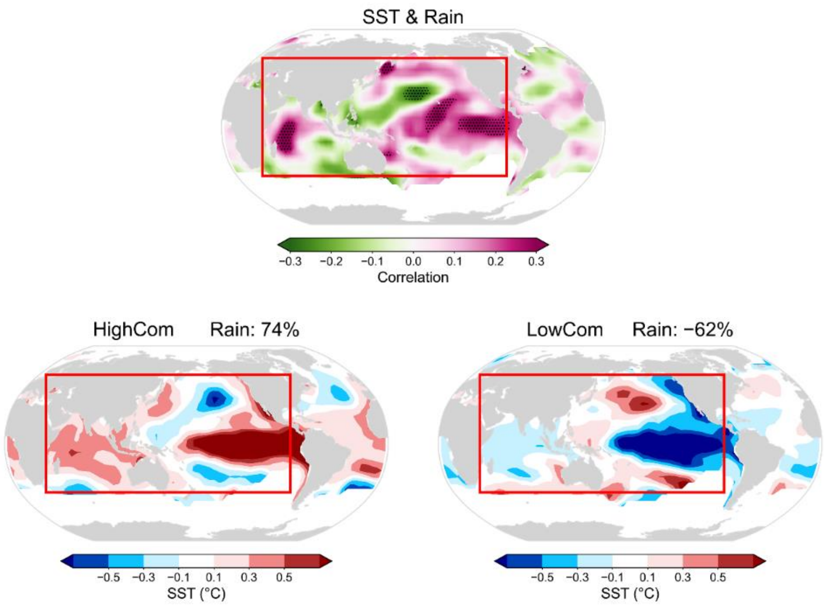


Figure 2. The linear correlation coefficients between SST and Rain (upper panel, stippling indicates the t-test confidence level greater than 95%) and the composite analyses (lower panel) in high (left) and low (right) precipitation anomalous years. The values in top right corner are the corresponding Rain. The red-boxed indicates the SST area used for forecasting system.

Inspired by strategies for enlarging the sample size for model training (e.g., data enhancement and synthetic data generation [46,47]), we also produce some synthetic samples. There have been many theoretical studies about the anomalous precipitation caused by SST. The composite analysis of anomalous years shown in Figure 2 can be explained, to some extent, by physical mechanisms [48–50]. In other words, these two composites (low and high) can be regarded as theoretical samples embodying the important characteristics that the forecast model should capture. Moreover, it is reasonable to assume that the Rain (output predictand) would increase or decrease with enhanced or weakened SST signal (input predictors). Based on this idea, some synthetic samples used for model training can be created through multiplying the composite analysis data (SST and corresponding Rain) by values around 1.0. Here, both high and low composite analyses provide 20 synthetic samples (40 in total). For clear display, only six synthetic samples are shown in Figure 3. To ensure all synthetic samples belong to anomalous events (with Rain exceeding 50% or falling below −50%) while simultaneously avoiding rare strong extreme events, the multiplied coefficient should be kept within an appropriate range, neither too small nor too large. Furthermore, to avoid that the role of synthetic samples in model training is too weak or too strong, the number of synthetic samples should also be kept within an appropriate range. The experiences from previous sensitivity experiments suggest that 40 is a reasonable number.

 doi:10.20944/preprints202501.1982.v1

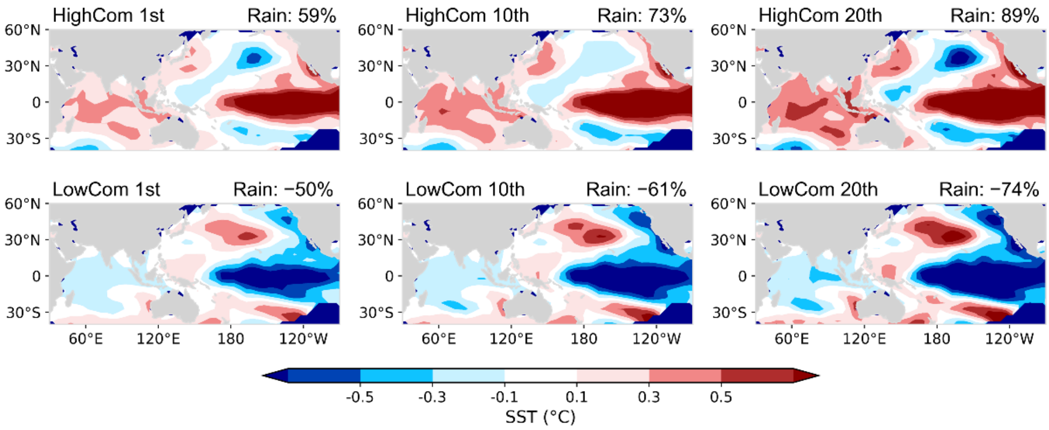
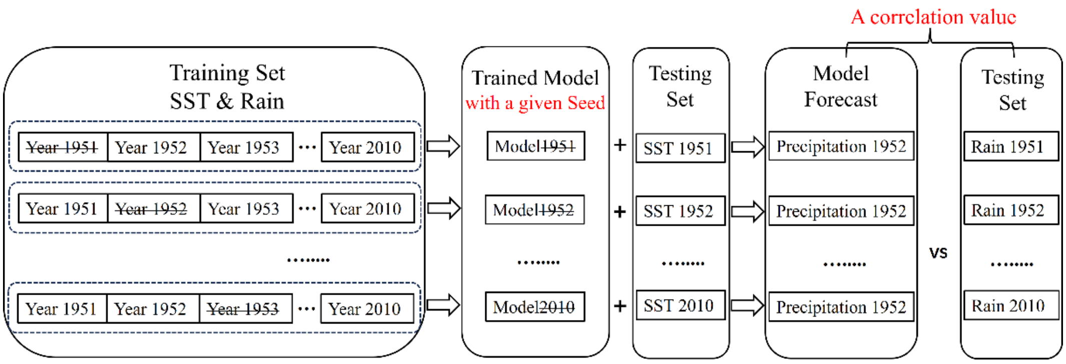


Figure 3. The SST (colored, in unit of °C) and corresponding Rain (top right corner) from the six synthetic samples created by the high (upper panel) or low (lower panel) composite analysis data. These synthetic samples have been sorted by strength (order number in the top left corner).

2.2. Forecasting System

The shallow NN (i.e., NN with one or two hidden layers) framework is employed in the forecasting system due to the small sample size. The NN model is trained with all available observed samples (i.e., samples before the predicted year) and synthetic samples, and the training process will stop when the loss is less than a pre-defined threshold. All hyperparameters of this NN framework (i.e., training threshold and the neuron numbers of every layer) are adjustable. To ensure the optimal performance of the NN framework, a set of optimal hyperparameters is first determined based on extensive preliminary experiments (refer to Appendix A for details).

Owing to the utilization of certain random variables (e.g., initial weight and bias) during the model training process, the performance of the trained NN model might be sensitive to the random seeds employed if training samples are insufficient [51–53]. To enhance the stability of forecast results, this forecasting system firstly undertakes the task of selecting 10 good seeds from a pool of 100 initialization seeds. Subsequently,10 NN models are trained with these 10 good seeds and provide 10 predictands. Eventually, the final predictand is calculated based on the average of these 10 predictands (i.e., ensemble mean). For example, supposing the current time is March 2011, the available observed samples are from 1951 to 2010. Firstly, a pool of 100 initialization seeds is set up. Secondly, the performance of each seed is analyzed via the leave-one-out cross-validation approach [54–56]. Each seed gets a forecast skill score (i.e., a correlation value). The specific processes are illustrated in Figure 4. Thirdly, 10 good seeds are selected from a seed pool based on corresponding scores. Fourthly, 10 NN models with these 10 good seeds are trained using all available samples (i.e., from 1951 to 2010). Fifthly, 10 Rains are predicted with these 10 NN models. Finally, the forecasting system output the predicted Precipitation, which is a scaled value based on the average of these 10 Rains. It is worth noting that, in the forecasting system, when adding the new observed sample from the latest year, the 10 NN forecast models will be established again.



Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 January 2025 doi:10.20944/preprints202501.1982.v1

Figure 4. Flow-process diagram for calculating the score (a correlation value) of a given seed.

2.3. Estimation Method

To mitigate the potential overfitting problems caused by small sample size, one forecasting system is often estimated via rolling forecasting experiments [57–59]. To let all sample years be predicted by the rolling experiment, one whole rolling forecasting experiment consists of two parts: forward experiment (from 1987 to 2023) and backward experiment (from 1986 to 1951). The backward experiment assumes sample years are in reverse chronological order (from 2023 to the predicted year). This assumption is acceptable because samples are independent of each other.

Besides the common-used linear correlation coefficient (Corr), the comparison between observations and predictands (i.e., forecast skill) is also quantified by another two metrics, which focus on anomalous events. Similar to the metric POD (Probability of Detection) and FAR (False Alarm Ratio) [60] used for weather events, the proportions of accurately predicted (Succ) and conflictingly predicted (Wrong) precipitation anomalies are defined here. Succ is the number of years with successfully predicted anomalous events, divided by the total number of observed anomalous years. Wrong is the number of years with contradictory predicted anomalous events, divided by the total number of predicted anomalous years. In short, the forecast skill is quantified by three scores (i.e., Corr, Succ, and Wrong).

In this study, the stability and interpretability of the forecasting system are also analyzed. The identical experiment is carried out five times (specifically, Round 1, Round 2, Round 3, Round 4, and Round 5), with only the random seeds being different. The stability is directly reflected by the differences in the experiment results of these five rounds. The interpretability of forecasting models is explored by examining the relationship between the output predictand (Precipitation) and the input predictors (SST). This input-output relationship is depicted via the sensitivity map. The sensitivity map represents the alterations in the predicted Precipitation that are induced by a 1°C increase in SST at each grid point.

2.4. Experiment Setup

Table 1 lists all the experiments analyzed in this study. During each rolling forecasting experiment, the neuron numbers of one input layer (N0) and two hidden layers (N1 and N2) remain fixed. Preliminary experiments (Appendix A) show that one optimal set of architecture hyperparameters is N0 = 6, N1 = 3, and N2 = 0 (i.e., only one hidden layer exists). The optimal experiment (OPT) is conducted with this NN framework and all improving techniques. Besides the one-hidden-layer NN framework, an extra experiment (FMK) is carried out utilizing a two-hidden-layer NN framework. The optimal combination of architecture hyperparameters (N0 = 6, N1 = 6, and N2 = 3) is applied in the FMK experiment. The comparison between the OPT and FMK experiments shows the impact of the number of hidden layers.

Table 1. All rolling forecasting experiments conducted with the forecasting system.

Experiment	Description
OPT	Optimal experiment with synthetic samples, selecting random seeds and ensemble-mean technique. The NN framework only use one hidden layer, N0=6 and N1=3. Precipitations from anomalous years are amplified.
Compare different NN frameworks	
FMK	Similar to OPT, but using two hidden layers. N0=6, N1=6, and N2=3.
Sensitivity experiments concerning the techniques of initialization seeds	
NoSD	Similar to OPT, but without selecting seeds process.
NoEM	Similar to NoSD, but even without ensemble-mean technique.
Sensitive experiments regarding the techniques of sample enhancement	
NoSS	Similar to OPT, but without synthetic samples.

 doi:10.20944/preprints202501.1982.v1

NoAF	Similar to NoSS, but without scaling anomalous precipitations.

There are two sensitivity experiments (NoSD and NoEM) concerning the techniques of initialization seeds. In the NoSD experiment, the task of selecting 10 optimal seeds from a pool of 100 initialization seeds is eliminated by simply using only 10 seeds in the pool. The NoEM experiment goes a step further to dispense with the ensemble-mean technique by merely deploying a single seed in the pool. Moreover, another two sensitivity analysis experiments (NoSS and NoAF) are conducted to explore the effects of sample enhancement techniques. In the NoSS experiment, the synthetic samples are not used for model training. In contrast to the NoSS experiment, the amplification for anomalous samples is canceled in the NoAF experiment.

3. Results and Analysis

This section first estimates the forecasting system based on the rolling forecasts from both the OPT and FMK experiments. After that, the stability and interpretability of the forecasting system are analyzed based on sensitivity experiments. Meanwhile, the effects of improving techniques are also illustrated.

3.1. Forecast Skill

Figure 5 shows the rolling forecasts from the OPT and FMK experiments. The linear correlations between observations and forecasts are both around 0.5. In the OPT experiment, half of observed anomalous years (21/42) are successfully predicted. One year (1951) is contradictory predicted among the 34 predicted anomalous years (1/34). The observed and predicted Precipitation of 1951 are 36.4% and −26.4%, respectively. This can be explained by that the SST of 1951 (not shown) is more similar to the composite SST of low Precipitation years whereas the year has a high Precipitation. In other words, the sample of year 1951 is different from the common characteristics derived from most samples. It is difficult for statistical models to provide a successful forecast in this kind of uncommon years. As compare to the OPT experiment, the rolling forecasts from the FMK experiment have a little bit strong amplitude. Correspondingly, the number of predicted anomalous years from the FMK (35) experiment is one more than that from the OPT experiment (34). Unfortunately, three anomalous years are contradictory predicted. Taken together, the rolling forecasts from the OPT experiment are a little better than that from the FMK experiment. This suggests that the number of hidden layers is not the more the better.

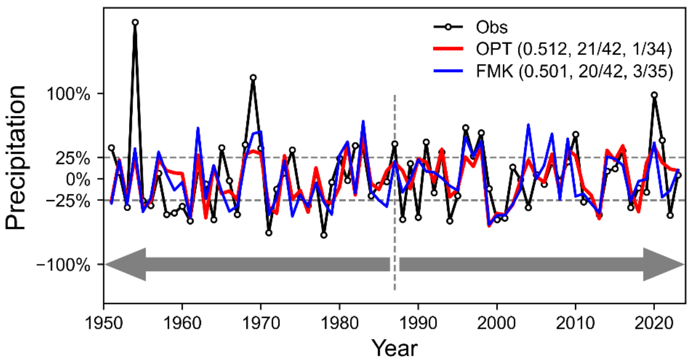


Figure 5. Rolling forecasts from the OPT (red line) and FMK (blue line) experiments. The observed Precipitation is denoted by black line. Gray dashed lines indicate the threshold values for anomalous events. Forecast skills (Corr, Succ, and Wrong) are presented adjacent to the experiment names.

3.2. Stability

The forecasting stability is also an important consideration in forecasting system estimation. The results of the five-round NoEM experiment directly illustrate the stability of the rolling forecasts triggered by the random seeds (Figure 6). The Corrs (a metric of forecast skills) of these five rounds

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 January 2025 doi:10.20944/preprints202501.1982.v1

range between 0.436 and 0.502. In terms of Succ and Wrong, the differences among these five rounds are also obvious. These pronounced differences indicate that the random seeds employed in the training process of the NN model exert a substantial influence on the model's output predictand, thereby resulting in a non-negligible degree of instability. Ensemble mean is a common-used technique for enhancing stability [61,62]. As expected, the five-round rolling forecasts from the NoSD experiment (employing the ensemble-mean technique) exhibit better stability compared to those from the NoEM experiment (Figure 6). The disparity in the five-round Corrs from the NoSD experiment, ranging from 0.473 to 0.496, is relatively small. The Corrs from the NoSD experiment generally fall within the range of those from the NoEM experiment (without the ensemble mean), because the final predicted Precipitation from the NoSD experiment is the ensemble mean of 10 predictands (i.e., 10 rounds of the NoEM experiment). In comparison with the NoSD experiment, the ensemble-mean in the OPT experiment is further based on 10 optimal seeds selected from a pool of 100 random seeds. The five- round Corrs of the OPT experiment (ranging from 0.504 to 0.517) are obviously enhanced, and the disparity among them is further reduced. In terms of Succ and Wrong, the OPT experiment also demonstrates better forecast skill compared to the NoSD experiment. To sum up, the two improving techniques involved here (i.e., optimal seeds and ensemble-mean) do enhance the forecasting system.

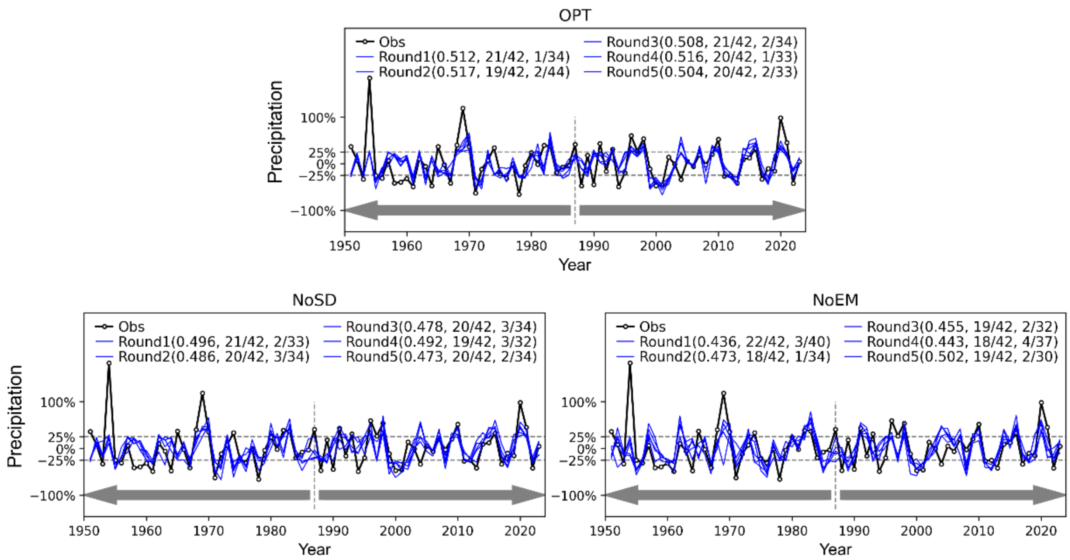


Figure 6. Forecasting stability from the OPT, NoSD, and NoEM experiments. The observed and predicted Precipitations are denoted by black and blue solid line, respectively. Gray dashed lines indicate the threshold values for anomalous events. Forecast skills (Corr, Succ, and Wrong) are presented adjacent to the experiment names.

3.3. Interpretability

This paragraph analyzes the interpretability of forecasting models based on sensitivity maps. Simultaneously, the differences in forecast skill among the OPT, NoSS, and NoAF experiments are also explained through these analyses. The sensitivity maps across different years are different (Figure 7). One reason is that the established NN models vary annually during the rolling forecasting experiment. Another reason lies in the fact that the sensitivity map is contingent not only upon the NN model itself but also on the input SST. Since the 40 synthetic samples used for model training remain identical each year during the rolling forecasting experiment, the three sensitivity maps derived from the OPT experiment are generally alike. Generally speaking, all these sensitivity maps are consistent with the background knowledge deduced from the observations (Figure 2). For example, the fact that a warmer equatorial eastern Pacific SST is conducive to heavier precipitation serves as an illustration. This implies that the interpretability of the forecasting system is deemed acceptable. In the absence of synthetic samples, both the NoSS and NoAF experiments demonstrate

 doi:10.20944/preprints202501.1982.v1

that there are pronounced disparities in the sensitivity maps across different years. Furthermore, the sensitivities from both the NoSS and NoAF experiments are generally significantly weaker than those from the OPT experiment. This could be accounted by the fact that all synthetic samples are anomalous events. Based on the definition of the sensitivity map (Section 2.3), the output predicted Precipitation can be approximated by calculating the inner product between the SST field and the sensitivity map. Therefore, the rolling forecasts from both the NoSS and NoAF experiments display a somewhat weaker amplitude and a smaller number of anomalous events (Figure 7). The sensitivity maps from the NoAF experiment are somewhat weaker than those from the NoSS experiment. Correspondingly, the number of anomalous events in the NoAF experiment (26) is fewer than that in the NoSS experiment (29). The aforesaid differences between the NoSS and NoAF experiments suggest that amplifying anomalous precipitations also contributes to the forecasting system. To summarize, the two sample-related techniques used in the forecasting system (synthetic samples and amplified anomalous events) indeed enhance the interpretability, forecast skill, and stability.

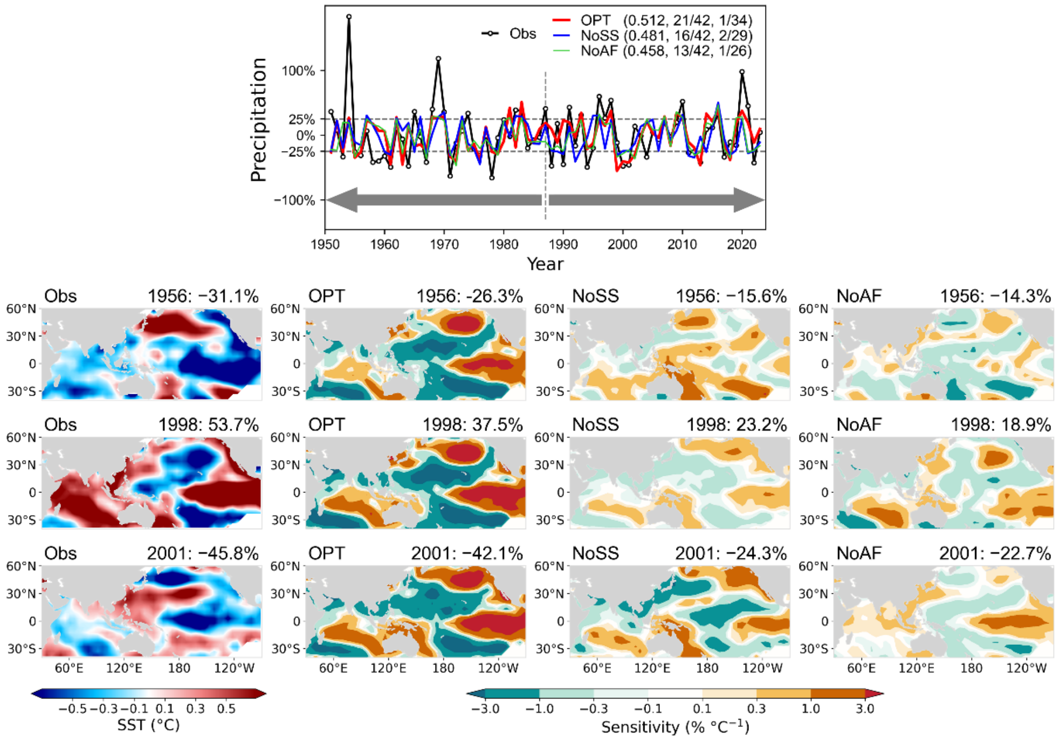


Figure 7. Rolling forecasts from the OPT, NoSS, and NoAF experiments (upper panel) and interpretability analyses in 1956, 1998 and 2001 (lower panel). The SST fields and the corresponding sensitivity maps are presented in the 1st column and the 2nd - 4th columns, respectively.

4. Discussion

Theoretical understandings of climate anomalous events can be harnessed to create artificial samples, aiming to mitigate the shortage of observed samples for model training. For instance, some previous studies suggest that the simulation results from climate models, regarded as artificial samples, can be used for model training [24,63]. This approach is exclusively applicable to those climate events that can be predicted to a certain degree by climate models, such as the El Niño or La Niña events. Nevertheless, due to the unacceptable forecast skill of climate models for predicting the anomalous precipitation events in July over MLRYR with a lead time of more than one season [64,65], the climate simulation results are not employed in this study.

5. Conclusions

This study aims to establish a robust statistical forecasting system for predicting the July precipitation over MLRYR based on the prophase winter SST. Here, we focus on discussing the

 doi:10.20944/preprints202501.1982.v1

problems encountered during the process of establishing the forecasting system with a common-used machine learning method (i.e., NN framework) and corresponding solutions mostly derived from the characteristics of climate issues.

Different from other machine-learning application fields, climate forecasts pay more attention to anomalous events rather than normal events. Additionally, in the model training process, there are not sufficient observed samples even for a very simple NN framework. But on the other side, there has been a great deal of background theoretical knowledge, particularly about anomalous events. Considering the above unique advantages and disadvantages of climate issues, we explored some techniques to enhance the stability and interpretability of the forecasting system, as well as pay more attention to anomalous events.

The five-round experiment results demonstrate that the performance of the trained NN model is unstable caused by the random seeds used in the model training process. This instability can be mitigated by ensemble-mean technique. Moreover, selecting good seeds for ensemble-mean, not only can stability be further improved, but forecast skill can also be significantly enhanced. Unlike statistical conclusions, climate theoretical knowledge has extensibility, and thus it can be used to create synthetic samples. This technique can significantly improve the interpretability of the forecasting models, thereby enhancing the forecast skills. The rolling forecast experiment results further confirm that the statistical forecasting models usually provide significantly fewer years of anomalous events, and the forecasting skills for abnormal events are generally low. Both the synthetic samples technique and amplifying anomalous events technique can improve the anomalous events forecasting. With the help of the aforementioned enhanced technologies, the final established forecasting system is generally stable and has good interpretability. The estimation of forecast skills suggests that this forecasting system has significant practical value, especially for anomalous events.

Author Contributions: X.S. designed this study. W.L. built the forecasting model, carried out the experiments, and created all figures. X.S. and W.L. wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by and the National Natural Science Foundation of China (grant nos. 41775095 and 42075145). The APC was funded by the same funders.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The precipitation data are downloadable from http://cmdp.ncc-cma.net/cn/index.htm (accessed on 20 January 2025). The original SST data are downloadable from https://psl.noaa.gov/data/gridded/data.kaplan_sst.html (accessed on 20 January 2025). The code of the forecast system and related results used in this study have been archived in a public repository https://doi.org/10.5281/zenodo.14690517 (accessed on 20 January 2025).

Acknowledgments: This study was conducted in the High-Performance Computing Center of Nanjing University of Information Science & Technology.

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

Appendix A: How to Set the Hyperparameters of NN Framework

There are four hyperparameters in the NN framework: one model training related hyperparameter (Tr) and three architecture hyperparameters (i.e., N0, N1, and N2). The Tr hyperparameter is a threshold. The model training process will stop when the loss below this threshold. A high value of Tr might prematurely terminate training, resulting in underfitting. In another side, a low value of Tr might induce overfitting, where the model excels on training set but falters on testing set. The more complex the framework architecture, the more training samples are needed. Because of mall sample size, the framework architecture should not be too complex. The optimal set of hyperparameters can be got through a lot of experiments. Here, the common-used

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 January 2025 doi:10.20944/preprints202501.1982.v1

leave-one-out cross validation experiments are used to estimate the NN model with a given set of hyperparameters. A group of forecast skills (Corr, Succ, and Wrong) correspond to a set of hyperparameters (Tr, N0, N1, and N2).

Since the combination of the four hyperparameters (Tr, N0, N1, and N2) are too large, the one-hidden-layer scenario (N2 = 0) is analyzed first. Table A1 lists the performances from different combinations of the rest three hyperparameters (Tr, N0, and N1). Taken overall, the combination (Tr = 0.28, N0 = 6, and N1 = 3) show best forecast skill (0.540, 25/42, 0/41). The two-hidden-layer scenario is also analyzed with fixed Tr (Tr =0.28). Table A2 shows that the framework architecture (N0 = 6, N1 = 6, and N2 = 3) provides the optimal forecast skill (0.537, 26/42, 0/40). The experiments with one-hidden-layer framework run faster than that with two-hidden-layer framework. To facilitate rapid experiments, the optimal set of one-hidden-layer hyperparameters is chosen for the reference experiment (i.e., OPT) used in this study. The above optimal set of two-hidden-layer hyperparameters is also tested as a sensitivity experiment (i.e., FMK).

Table A1. Forecast skills (Corr, Succ, and Wrong) from the one-hidden-layer framework under different hyperparameter combinations.

Tr\ (N0, N1)	(10, 5)	(8, 4)	(6, 3)	(4, 2)
0.50	0.468, 20/42, 4/41	0.432, 23/42, 3/43	0.483, 20/42, 3/40	0.475, 22/42, 2/41
0.40	0.505, 24/42, 3/39	0.477, 22/42, 1/40	0.513, 22/42, 3/43	0.517, 25/42, 2/42
0.36	0.503, 23/42, 0/37	0.478, 24/42, 2/41	0.522, 21/42, 0/38	0.522, 24/42, 1/40
0.32	0.528, 22/42, 1/38	0.494, 25/42, 1/41	0.526, 24/42, 0/42	0.527, 24/42, 1/43
0.28	0.522, 24/42, 0/38	0.519, 26/42, 1/39	0.540, 25/42, 0/41	0.528, 23/42, 1/40
0.24	0.526, 22/42, 0/41	0.508, 25/42, 0/39	0.531, 24/42, 0/43	0.510, 22/42, 1/42
0.20	0.517, 21/42, 1/40	0.478, 24/42, 1/37	0.516, 22/42, 1/42	0.504, 23/42, 1/43

Table A2. Forecast skills (Corr, Succ, and Wrong) from the two-hidden-layer framework under different architecture hyperparameter combinations.

N0\ (N1, N2)	(10, 5)	(8, 4)	(6, 3)	(4, 2)	(3, 1)
10	0.495, 17/42, 3/39	0.485, 22/42, 1/41	0.486, 21/42, 2/41	0.484, 21/42, 1/40	0.488, 24/42, 3/43
8	0.482, 20/42, 2/39	0.502, 22/42, 1/42	0.496, 24/42, 1/40	0.522, 22/42, 2/41	0.514, 20/42, 0/43
6	0.505, 21/42, 1/40	0.533, 23/42, 0/38	0.537, 26/42, 0/40	0.532, 24/42, 2/41	0.528, 23/42, 3/40
4	0.492, 22/42, 2/41	0.494, 23/42, 2/41	0.509, 25/42, 1/43	0.504, 24/42, 0/42	0.503, 23/42, 2/41

References

1. Bett, P. E.; Scaife, A.; Li, C.; Hewitt, C.; Golding, N.; Zhang, P.; Dunstone, N.; Smith, D. M.; Thornton, H. E.; Lu, R.; Ren, H. Seasonal Forecasts of the Summer 2016 Yangtze River Basin Rainfall. *Adv. Atmos. Sci.* **2018**, 35, 918–926. [CrossRef]
2. Ha, S.; Liu, D.; Mu, L. Prediction of Yangtze River streamflow based on deep learning neural network with El Niño–Southern Oscillation. *Sci. Rep.* **2021**, 11, 11738. [CrossRef]
3. Huang, A.; Gao, G.; Yao, L.; Yin, S.; Li, D.; Do, H. X.; Fu, B. Spatiotemporal variations of inter-and intra-annual extreme streamflow in the Yangtze River Basin. *J. Hydrol.* **2024**, 629, 130634. [CrossRef]
4. Zhou, L.T.; Tam, C.Y.; Zhou, W.; Chan, J. Influence of South China Sea SST and the ENSO on winter rainfall over South China. *Adv. Atmos. Sci.* **2010**, 27, 832–844. [CrossRef]
5. Ying, K.; Zheng, X.; Quan, X.; Frederiksen, C. Predictable signals of seasonal precipitation in the Yangtze–Huaihe River Valley. *Int. J. Climatol.* **2013**, 33, 3002–3015. [CrossRef]
6. Deng, Y.; Gao, T.; Gao, H.; Yao, X.; Xie, L. Regional precipitation variability in East Asia related to climate and environmental factors during 1979–2012. *Sci. Rep.* **2014**, 4, 5693. [CrossRef]

7. Zhang, Y.; Zhou, W.; Wang, X.; Chen, S.; Chen, J.; Li, S. Indian Ocean Dipole and ENSO's mechanistic importance in modulating the ensuing-summer precipitation over Eastern China. *npj Clim. Atmos. Sci.* **2022**, 5, 48. [CrossRef]
8. Wu, M.; Zhang, R.H.; Hu, J.; Zhi, H. Synergistic interdecadal evolution of precipitation over Eastern China and the Pacific decadal oscillation during 1951–2015. *Adv. Atmos. Sci.* **2024**, 41, 53–72. [CrossRef]
9. Wang, B.; Liu, J.; Yang, J.; Zhou, T.; Wu, Z. Distinct principal modes of early and late summer rainfall anomalies in East Asia. *J. Climate* **2009**, 22, 3864–3875. [CrossRef]
10. Su, Q.; Lu, R.Y.; Li, C.F. Large-scale circulation anomalies associated with interannual variation in monthly rainfall over South China from May to August. *Adv. Atmos. Sci.* **2014**, 31, 273–282. [CrossRef]
11. Xing, W.; Wang, B.; Yim, S. Long-lead seasonal prediction of China summer rainfall using an EOF–PLS regression-based methodology. *J. Climate* **2016**, 29, 1783–1796. [CrossRef]
12. Wei, F. Progresses on climatological statistical diagnosis and prediction methods—In commemoration of the 50th anniversaries of CAMS establishment. *J. Appl. Meteor. Sci.* **2006**, 17, 736–742. (In Chinese) [CrossRef]
13. Troccoli, A. Seasonal climate forecasting. *Meteorol. Appl.* **2010**, 17, 251–268. [CrossRef]
14. Wang, H.; Fan, K.; Sun, J.; Li, S.; Lin, Z.; Zhou, G.; Chen, L.; Lang, X.; Li, F.; Zhu, Y.; Chen, H.; Zheng, F. A review of seasonal climate prediction research in China. *Adv. Atmos. Sci.* **2015**, 32, 149–168. [CrossRef]
15. Fan, K.; Wang, H.; Choi, Y. A physically-based statistical forecast model for the middle-lower reaches of the Yangtze River Valley summer rainfall. *Chinese Sci. Bull.* **2008**, 53, 602–609. (In Chinese) [CrossRef]
16. Lang, X.M.; Wang, H. Improving extraseasonal summer rainfall prediction by merging information from GCMs and observations. *Weather Forecasting* **2010,** 25, 1263–1274. [CrossRef]
17. Guo, Y.; Li, J.; Zhu, J. A moving updated statistical prediction model for summer rainfall in the MLRYR valley. *J. Appl. Meteorol. Climatol.* **2017**, 56, 2275–2287. [CrossRef]
18. Li, W.; Gao, X.; Hao, Z.; Sun, R. Using deep learning for precipitation forecasting based on spatio-temporal information: A case study. *Clim. Dyn.* **2021**, 58, 443–457. [CrossRef]
19. Agarwal, A.; Guntu, R.; Banerjee, A.; Gadhawe, M.; Marwan, N. A complex network approach to study the extreme precipitation patterns in a river basin. *Chaos* **2022**, 32, 013113. [CrossRef]
20. Deepthi, B.; Sivakumar, B. General circulation models for rainfall simulations: Performance assessment using complex networks. *Atmos. Res.* **2022**, 278, 106333. [CrossRef]
21. Li, K.; Huang, Y.; Liu, K.; Wang, M.; Cai, F.; Zhang, J.; Boers, N. Key propagation pathways of extreme precipitation events revealed by climate networks. *npj Clim. Atmos. Sci.* **2024**, 7, 165. [CrossRef]
22. Chen, N.; Majda, A.J. Predicting observed and hidden extreme events in complex nonlinear dynamical systems with partial observations and short training time series. *Chaos* **2020**, 30, 033101. [CrossRef]
23. Ribeiro, R.P.; Moniz, N. Imbalanced regression and extreme value prediction. *Mach. Learn.* **2020**, 109, 1803–1835. [CrossRef]
24. Ham, Y.G.; Kim, J.H.; Luo, J.J. Deep learning for multi-year ENSO forecasts. *Nature* **2019**, 573, 568–572. [CrossRef]
25. Reichstein, M.; Camps-Valls, G.; Stevens, B.; Jung, M.; Denzler, J.; Carvalhais, N.; Prabhat. Deep learning and process understanding for data-driven Earth system science. *Nature* **2019**, 566, 195–204. [CrossRef]
26. Mansfield, L; Nowack, P.; Kasoar, M.; Everitt, R.; Collins, W.; Voulgarakis, A. Predicting global patterns of long-term climate change from short-term simulations using machine learning. *npj Clim. Atmos. Sci.* **2020**, 3, 44. [CrossRef]
27. Schneider, T.; Behera, S.; Boccaletti, G.; et al. Harnessing AI and computing to advance climate modelling and prediction. *Nature Clim. Change* **2023**, 13, 887–889. [CrossRef]
28. Schneider, T.; Behera, S.; Boccaletti, G.; Deser, C.; Emanuel, K.; Ferrari, R.; Leung, L. R.; Lin, N.; Müller, T.; Navarra, A.; Ndiaye, O.; Stuart, A.; Tribbia, J.; Yamagata, T. AI for climate impacts: Applications in flood risk. *npj Clim. Atmos. Sci.* **2023**, 6, 63. [CrossRef]
29. Becker, E.J.; van den Dool, H.; Peña, M. Short-term climate extremes: Prediction skill and predictability. J. Climate 2013, 26, 512–531. [CrossRef]
30. Trenberth, K.; Fasullo, J.; Shepherd, T. Attribution of climate extreme events. Nature Clim. Change 2015, 5, 725–730. [CrossRef]

31. Bellprat, O.; Guemas, V.; Doblas-Reyes, F.; Donat, M. G. Towards reliable extreme weather and climate event attribution. Nature Commun. 2019, 10, 1732. [CrossRef]
32. Xu, P.; Ji, X.; Li, M.; Lu, W. Small data machine learning in materials science. npj Comput. Mater. 2023, 9, 42. [CrossRef]
33. Eyring, V.; Collins, W. D.; Gentine, P.; Barnes, E. A.; Barreiro, M.; Beucler, T.; Bocquet, M.; Bretherton, C. S.; Christensen, H. M.; Dagon, K.; Gagne, D. J.; Hall, D.; Hammerling, D.; Hoyer, S.; Iglesias-Suarez, F.; Lopez-Gomez, I.; McGraw, M. C.; Meehl, G. A.; Molina, M. J.; Monteleoni, C.; Mueller, J.; Pritchard, M. S.; Rolnick, D.; Runge, J.; Stier, P.; Watt-Meyer, O.; Weigel, K.; Yu, R.; Laure. Pushing the frontiers in climate modelling and analysis with machine learning. Nature Clim. Change 2024, 14, 916–928. [CrossRef]
34. Hu, Z. Interdecadal variability of summer climate over East Asia and its association with 500 hPa height and global sea surface temperature. J. Geophys. Res. 1997, 102, 19403–19412. [CrossRef]
35. Nan, S.; Li, J. The relationship between the summer precipitation in the Yangtze River valley and the boreal spring Southern Hemisphere annular mode. Geophys. Res. Lett. 2003, 30, 2266. [CrossRef]
36. Zhu, Y.; Wang, H.; Zhou, W.; Ma, J. Recent changes in the summer precipitation pattern in East China and the background circulation. Clim. Dyn. 2011, 36, 1463–1473. [CrossRef]
37. Li, M.; Guan, Z.; Jin, D.; Han, J.; Zhang, Q. Anomalous circulation patterns in association with two types of daily precipitation extremes over southeastern China during boreal summer. J. Meteor. Res. 2016, 30, 183–202. [CrossRef]
38. Jin, D., S. N. Hameed, and L. Huo. Recent Changes in ENSO Teleconnection over the Western Pacific Impacts the Eastern China Precipitation Dipole. *Journal of Climate*, **2016**, 29, 7587-7598. [CrossRef]
39. Zhang, Z.; Wang, X.; Liu, M.; Huang, B.; Wu, Y.; Feng, G.; Sun, G. Contributions of Multiple Water Vapor Sources to the Precipitation in Middle and Lower Reaches of Yangtze River Based on Precipitation Recycle Ratio. *Atmosphere* **2022**, 13, 1957. [CrossRef]
40. Zhang, W.; Tao, W.; Huang, G.; Hu, K.; Qu, X.; Wang, Y.; Tang, H.; Zhang, S. Ongoing Intensification of Anomalous Western North Pacific Anticyclone during Post-El Niño Summer with Achieved Carbon Neutrality. *npj Clim. Atmos. Sci.* **2024**, 7, 317. [CrossRef]
41. Yihui, D., Chan, J. The East Asian Summer Monsoon: An Overview. *Meteorology and Atmospheric Physics*, **2005**, 89, 117–142. [CrossRef]
42. Cao, F.; Gao, T.; Dan, L.; Ma, Z.; Chen, X.; Zou, L.; Zhang, L. Synoptic-Scale Atmospheric Circulation Anomalies Associated with Summertime Daily Precipitation Extremes in the Middle–Lower Reaches of the Yangtze River Basin. *Clim. Dyn.* **2019**, 53, 3109–3129. [CrossRef]
43. Kaplan, A.; Cane, M. A.; Kushnir, Y.; Clement, A. C.; Blumenthal, M. B.; Rajagopalan, B. Analyses of Global Sea Surface Temperature **1856–1991**. *Journal of Geophysical Research*, **1998**, 103, 18,567-18,589. [CrossRef]
44. Latif, Mojib. The Ocean's Role in Modeling and Predicting Decadal Climate Variations. *International Geophysics*, **2013**, 103, 645-665. [CrossRef]
45. Kim, B.H., Ha, K.J. Observed Changes of Global and Western Pacific Precipitation Associated with Global Warming SST Mode and Mega-ENSO SST Mode. *Climate Dynamics*, **2015**, 45, 3067–3075. [CrossRef]
46. Iglesias, G.; Talavera, E.; González-Prieto, Á.; Mozo, A.; Gómez-Canaval, S. Data Augmentation Techniques in Time Series Domain: A Survey and Taxonomy. *Neural Computing and Applications*, **2023**, 35, 10123-10145. [CrossRef]
47. Taylor, L.; Nitschke, G. Improving Deep Learning with Generic Data Augmentation. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, **2018**, 1542-1547. [CrossRef]
48. Jiang, T.; Zhang, Q.; Zhu, D.; Wu, Y. Yangtze Floods and Droughts (China) and Teleconnections with ENSO Activities (**1470–2003**). *Quaternary International*, **2005**, 144(1), 29-37. [CrossRef]
49. Johnny, C.; Zhou, W. PDO, ENSO and the Early Summer Monsoon Rainfall over South China. *Geophys. Res. Lett.*, **2005**, 32, L08810. [CrossRef]
50. Xu, B.; Li, G.; Gao, C.; Yan, H.; Wang, Z.; Li, Y.; Zhu, S. Asymmetric Effect of El Niño—Southern Oscillation on the Spring Precipitation over South China. *Atmosphere* **2021**, 12, 391. [CrossRef]
51. Shaziya, H.; Zaheer, R. Impact of Hyperparameters on Model Development in Deep Learning. *Proceedings of International Conference on Computational Intelligence and Data Engineering*, **2020**, 56, 57-67. [CrossRef]

52. Dodge, J.; Ilharco, G.; Schwartz, R.; Farhadi, A.; Hajishirzi, H.; Smith, N. Fine-Tuning Pretrained Language Models: Weight Initializations, Data Orders, and Early Stopping. *arXiv*, **2020**, 06305. [CrossRef]
53. Pranava Madhyastha and Rishabh Jain. On Model Stability as a Function of Random Seed. *Proceedings of the 23rd Conference on Computational Natural Language Learning*, **2019**, 929–939. Association for Computational Linguistics. [CrossRef]
54. Cawley, G. C. Leave-One-Out Cross-Validation Based Model Selection Criteria for Weighted LS-SVMs. *The 2006 IEEE International Joint Conference on Neural Network Proceedings*, Vancouver, BC, Canada, 16-21 July 2006. [CrossRef]
55. Wu, J.; Mei, J.; Wen, S.; Liao, S.; Chen, J.; Shen, Y. A Self-Adaptive Genetic Algorithm-Artificial Neural Network Algorithm with Leave-One-Out Cross Validation for Descriptor Selection in QSAR Study. *J. Comput. Chem.*, **2010**, 31, 1956-1968. [CrossRef]
56. Snieder, E.; Abogadil, K.; Khan, U. T. Resampling and Ensemble Techniques for Improving ANN-Based High Streamflow Forecast Accuracy. *Hydrol. Earth Syst. Sci.*, **2020**, 24, 430-440. [CrossRef]
57. Maltare, N., Sharma, D., Patel, S. Rainfall Data-Based Time Series Forecasting Using Rolling Forecasting Model for Indian Geographic Area. In: Kaiser, M.S., Xie, J., Rathore, V.S. (eds) *Information and Communication Technology for Competitive Strategies* (ICTCS 2021). *Lecture Notes in Networks and Systems*, Springer, Singapore, **2023;** pp. 139–146. [CrossRef]
58. Mu, B.; Peng, C.; Yuan, S.; Chen, L. ENSO Forecasting over Multiple Time Horizons Using ConvLSTM Network and Rolling Mechanism. International Joint Conference on Neural Networks, Budapest, 14 July 2019. [CrossRef]
59. Yuan, S.; Wang, C.; Mu, B.; Zhou, F.; Duan, W. Typhoon Intensity Forecasting Based on LSTM Using the Rolling Forecast Method. *Algorithms*, **2021**, 14(3), 83. [CrossRef]
60. Wilks, D. Statistical Methods in the Atmospheric Sciences. 4th ed.; *Academic Press*, 2006; pp. 553-585. [CrossRef]
61. Frankcombe, L.; England, M; Kajtar, J.; Mann, M; Steinman, B. On the Choice of Ensemble Mean for Estimating the Forced Signal in the Presence of Internal Variability. *J. Climate*, **2018**, 31, 5681–5693. [CrossRef]
62. Christiansen, B. Analysis of Ensemble Mean Forecasts: The Blessings of High Dimensionality. *Mon. Wea. Rev.*, **2019**, 147, 1699–1712. [CrossRef]
63. Mansfield, L; Nowack, P.; Kasoar, M.; Everitt, R.; Collins, W; Voulgarakis, A. Predicting Global Patterns of Long-Term Climate Change from Short-Term Simulations Using Machine Learning. *npj Clim. Atmos. Sci.*, **2020**, 3, 44. [CrossRef]
64. Themeßl, M.; Gobiet, A., Heinrich, G. Empirical-Statistical Downscaling and Error Correction of Regional Climate Models and Its Impact on the Climate Change Signal. *Climatic Change*, **2012**, 112, 449–468. [CrossRef]
65. Alizadeh, O. Advances and Challenges in Climate Modeling. *Climatic Change*, **2022**, 170, 18. [CrossRef]