

Article

Multi-Model Ensemble Prediction of Summer Precipitation in China Based on Machine Learning Algorithms

Jie Yang ^{1,2} , Ying Xiang ¹, Jiali Sun ¹ and Xiazhen Xu ^{1,*}¹ Jiangsu Climate Center, Jiangsu Meteorology Bureau, Nanjing 210009, China² Institute of Physics Science and Technology, Yangzhou University, Yangzhou 225012, China

* Correspondence: xuxz0119@126.com

Abstract: The development of machine learning (ML) provides new means and methods for accurate climate analysis and prediction. This study focuses on summer precipitation prediction using ML algorithms. Based on BCC CSM1.1, ECMWF SEAS5, NCEP CFSv2, and JMA CPS2 model data, we conducted a multi-model ensemble (MME) prediction experiment using three tree-based ML algorithms: the decision tree (DT), random forest (RF), and adaptive boosting (AB) algorithms. On this basis, we explored the applicability of ML algorithms for ensemble prediction of seasonal precipitation in China, as well as the impact of different hyperparameters on prediction accuracy. Then, MME predictions based on optimal hyperparameters were constructed for different regions of China. The results showed that all three ML algorithms had an optimal maximum depth less than 2, which means that, based on the current amount of data, the three algorithms could only predict positive or negative precipitation anomalies, and extreme precipitation was hard to predict. The importance of each model in the ML-based MME was quantitatively evaluated. The results showed that NCEP CFSv2 and JMA CPS2 had a higher importance in MME for the eastern part of China. Finally, summer precipitation in China was predicted and tested from 2019 to 2021. According to the results, the method provided a more accurate prediction of the main rainband of summer precipitation in China. ML-based MME had a mean ACC of 0.3, an improvement of 0.09 over the weighted average MME of 0.21 for 2019–2021, exhibiting a significant improvement over the other methods. This shows that ML methods have great potential for improving short-term climate prediction.

Keywords: machine learning; climate models; multi-model ensemble; short-term climate prediction; precipitation



Citation: Yang, J.; Xiang, Y.; Sun, J.; Xu, X. Multi-Model Ensemble Prediction of Summer Precipitation in China Based on Machine Learning Algorithms. *Atmosphere* **2022**, *13*, 1424. <https://doi.org/10.3390/atmos13091424>

Academic Editor: Fei Ji

Received: 17 August 2022

Accepted: 31 August 2022

Published: 2 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Numerical climate models are important tools for the understanding of climatic phenomena, as well as for making climate predictions [1]. However, the capability for climate prediction is limited by internal atmospheric variability, which is largely unpredictable beyond the deterministic predictability limit of about two weeks [2]. It is fundamentally impossible to describe all the processes involved in the true climate system in a climate model, regardless of how complex the model is [3,4]. There are numerous uncertainties involved in constructing and applying these models, which are often classified as initial conditions, boundary condition parameters, structural uncertainties, etc. [5]. In this sense, knowledge of systematic errors occasioned by these uncertainties is of paramount importance in the realization of a suite of models, each of which carries a somewhat different representation of the above processes, which can be combined to reduce the collective local biases in space, time, and for different variables from the different models [6]. With the availability of climate predictions produced using various dynamical models, multi-model ensemble forecasting has gained more and more attention recently [7].

Multi-model ensembles (MME) are used to improve model predictions of temperature and precipitation where single-model prediction capabilities are limited [8–10]. In

particular, studies discuss a “super-ensemble approach”, in which a multi-model linear regression technique is used to improve deterministic forecasts locally [11]. In the sense of its construction, the super-ensemble is a post-processing product of multi-model forecasts. This super-ensemble can be used as a tool for making both deterministic and probabilistic predictions [12]. In addition, the technique of MME prediction has been explored by various researchers [13–15]. A combination of the results of various types of models, considering the performance of each model, can produce predictions of greater reliability [16]. The key challenge of MME is how to combine their complementary advantages.

With the development of artificial intelligence (AI), machine learning (ML) algorithms represented by deep learning (DL) have made breakthroughs and are now more widely used in the field of meteorology, covering many meteorological operations such as observation, forecasting, and services. ML provides an effective means of analyzing the massive amounts of observation and model simulation data, as well as providing technical support for weather forecasting [17]. ML can be used to learn laws and predict unknown data automatically from the data, and it also is suitable for all kinds of data. Therefore, ML is increasingly being applied to climate research [18]. Climate data can be efficiently analyzed and processed using ML methods, to extract valuable information from massive patterns of data and predict future climate conditions more accurately. With ML algorithms, new interrelated signals can be discovered and extracted from the climate system. For instance, SST data from a critical region can improve the climate prediction capability for a land region in the subsequent months [19]. It has been shown that a model based on convolutional neural networks can effectively predict ENSO events up to one and a half years in advance, with an accuracy of up to 80% [20]. In addition, a shallow neural network model was also effective in the identification of central and eastern El Niño events [21]. ML requires a large amount of data, to conduct appropriate training [22]. Hindcast samples for seasonal prediction models have been available for a few decades, but the volume of data is insufficient for ML, especially for DL. Due to the limited amount of data available, there have been relatively few studies on ML applied to MME prediction for seasonal precipitation. In spite of this, a limited amount of data does not prevent ML from making predictions, but early diagnosis and intervention are necessary [23].

The main objective of this study was to investigate the applicability of ML algorithms for MME prediction and to design an optimal ML ensemble forecast technique for summer precipitation in China. Decision tree (DT), random forest (RF), and adaptive boosting (AB) algorithms, which are easy to understand, were used to predict summer precipitation in China, based on the four dominant climate forecast systems. By varying the hyperparameters of each algorithm, we investigated the impact of hyperparameter changes on summer precipitation prediction, hence determining the optimal parameters for MME prediction across different regions. The importance of each model was quantitatively evaluated following the determination of the optimal parameters of the algorithms. On this basis, we applied the optimal MME algorithm to the prediction of summer precipitation for 2019–2021 and assessed its predictive ability.

2. Data and Methods

2.1. Data

Model data were derived from the current business models of the four major international climate prediction agencies, namely the Beijing Climate Center Climate System Model (BCC CSM), National Centers for Environmental Prediction Climate Forecast System version 2 (NCEP CFSv2), Japan Meteorological Agency Coupled Prediction System version 2 (JMA CPS2), and the European Centre for Medium-Range Weather Forecasts fifth long-range forecasting system (ECMWF SEAS5), hereafter referred to as BCC, NCEP, JMA, and ECMWF [24–27]. Each month’s data are the average set of all samples (different release times, different members) for the month. To take into account the need for the China Meteorological Administration’s climate prediction meeting at the end of March

each year, model precipitation data were used with a start date of March, with a data period of 1991–2021, and with a spatial resolution of $1^\circ \times 1^\circ$.

Observation of precipitation data was derived from the CPC Merged Analysis of Precipitation (CMAP). CMAP is a technique that produces pentad and monthly analyses of global precipitation in which observations from rain gauges are merged with precipitation estimates from several satellite-based algorithms (infrared and microwave). The analyses are on a 2.5×2.5 degree latitude/longitude grid and extend back to 1979. The time range and spatial resolution of all data, as well as the training and testing periods, are shown in Table 1. We interpolated the CMAP precipitation data to the model data resolution, to ensure spatial consistency.

Table 1. Summary of the forecast and observation products used in this study.

Name	Time Range	Spatial Resolution	Training Period	Testing Period
BCC CSM	1991–2021	$1^\circ \times 1^\circ$		
NCEP CFSv2	1982–2021	$1^\circ \times 1^\circ$		
JMA CPS2	1979–2021	$1^\circ \times 1^\circ$	1993–2018	2019–2021
ECMWF SEAS5	1993–2021	$1^\circ \times 1^\circ$		
CMAP	1979–2021	$2.5^\circ \times 2.5^\circ$		

Given the needs of the study, China was divided into eight regions: (A) South China (SC) contains Guangdong, the eastern part of Guangxi, Fujian, southern Jiangxi, and southern Hunan; (B) East China (EC) contains Jiangsu, Zhejiang, Anhui, Henan, Hubei; (C) North China (NC) contains Hebei, Shandong, eastern Shanxi, Beijing, and Tianjin; (D) Northeast China (NE) contains most of Northeast China, including Heilongjiang, Jilin, Liaoning, and Inner Mongolia; (E) western Northwest China (WN); (F) eastern Northwest China (EN); (G) the Tibetan Plateau (TP) contains Tibet and southwestern Qinghai Province; (H) Southwest China (SW) contains Yunnan, Guizhou, western Guangxi and Sichuan, Chongqing, western Hunan, and southern Shaanxi.

2.2. Methods

Classification and regression tree (CART) is a decision tree algorithm that was first introduced by Breiman [28]. This algorithm constructs an inverted tree-like graphical structure from data, comprising of a series of logical decisions at their root node, branches, and leaf nodes for classification or regression. The input and output data can be both categorical and continuous for classification and regression. Each node in CART represents a decision rule that splits the data into two or more homogeneous sets. The topmost node of the tree is known as the root node, which gives rise to internal nodes. The internal nodes have both parent and child nodes containing decision rules. The branches represent the outcome of the respective test or decision rule. The leaf node or terminal node represents the final output.

Suppose that X and Y are input and output variables, respectively, and Y are continuous variables. Given the training data set $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$, the input feature vector $x_i = \{x_i^1, x_i^2, \dots, x_i^n\}$. Considering that the input data has been divided into M cells (R_1, R_2, \dots, R_M) and that each cell R_m has a fixed output value c_m , the regression tree model can be expressed as follows:

$$f(x) = \sum_{m=1}^M c_m I(x \in R_m) \quad (1)$$

Following the division of the input data, the squared error can be used to express the prediction error of the regression tree, and the optimum output value on each cell will be determined by the least squared error criterion. The optimal value \hat{c}_m of c_m on a cell R_m is the mean value of the output y_i corresponding to all input instances x_i on R_m .

The CART regression tree divides the input data using a heuristic method, selecting the j -th variable x and the value s as the splitting variable and splitting point, and defining two

regions, $R_1(j, s) = \{x|x^j \leq s\}$ and $R_2(j, s) = \{x|x^j > s\}$. Then find the optimal splitting variable j and the optimal splitting point s with Equation (2):

$$\min_{j,s} \left[\min_{c_1} \sum_{x_i \in R_1(j,s)} (y_i - c_1)^2 + \min_{c_2} \sum_{x_i \in R_2(j,s)} (y_i - c_2)^2 \right] \tag{2}$$

Ensemble learning methods consist of a collection of weak learners, such as a decision tree, whose predictions are aggregated to determine the most popular result. The most well-known ensemble methods are bagging, also known as bootstrap aggregation, and boosting.

The random forest algorithm is an extension of the bagging method, as it utilizes both bagging and feature randomness to create an uncorrelated forest of decision trees [29]. A schematic illustration of the random forest concept can be found in Figure 1. This method generates a random subset of features, also known as feature bagging or the random subspace method, which ensures that decision trees have a low correlation among themselves. The random forest algorithm consists of a collection of decision trees, and each tree in the ensemble is comprised of a bootstrap sample, which is a data sample obtained from a training set with a replacement. Out of that training sample, one third is set aside as test data, called the out-of-bag (OOB) sample. Depending on the type of problem, the determination of the prediction will differ. For a regression problem, the individual decision trees will be averaged, whereas for a classification problem, the predicted class will be determined by a majority vote, or the most common categorical variable. Last but not least, the OOB sample is used for cross-validation to finalize the prediction.

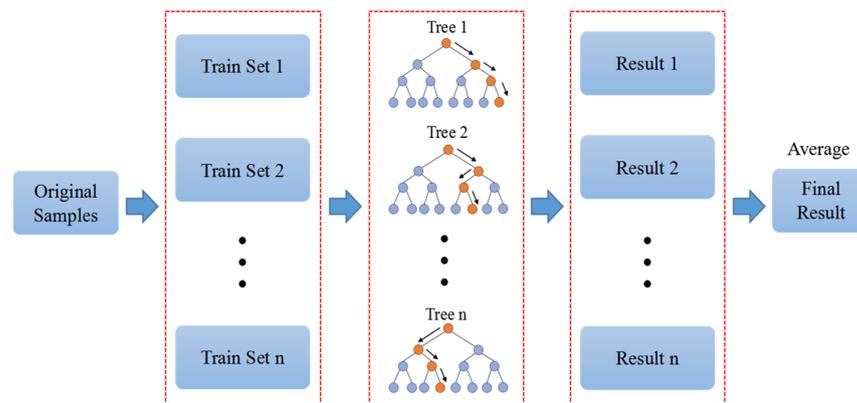


Figure 1. Flow chart of the random forest algorithm.

Freund and Schapire proposed the adaptive boosting algorithm [30]. It can be used to improve the performance of machine learning algorithms. It is best used with weak learners, and these models achieve high accuracy above random chance on regression problems. The common algorithms used with AdaBoost are decision trees of level one. Figure 2 illustrates a flow chart of the Adaboost algorithm. A weak learner is a classifier or predictor that performs relatively poorly in terms of accuracy. In addition, it can be implied that the weak learners are simple to compute, and many instances of algorithms are combined to create a strong learner through boosting.

As a next step, we describe the model training process. First, we extracted the precipitation forecasts for June, July, and August from each model and added them up, to obtain summer precipitation forecasts. After that, the summer precipitation prediction results were converted to anomaly percentage predictions of summer precipitation. Since ECMWF started in 1993, we chose 1993–2018 as the training period, and used 10-fold cross-validation to divide the training period into 10 subsets. Then, each subset was extracted in turn, as the validation period. The multi-year precipitation anomaly percentage predictions of the four models were substituted into the three machine learning models as features, respectively.

The mean value of the scores from the 10 cross-validation subsets was used as the prediction skill for the ML algorithms. This enabled the influence of hyperparameter variation on the prediction skill to be explored and the optimal hyperparameters to be determined.

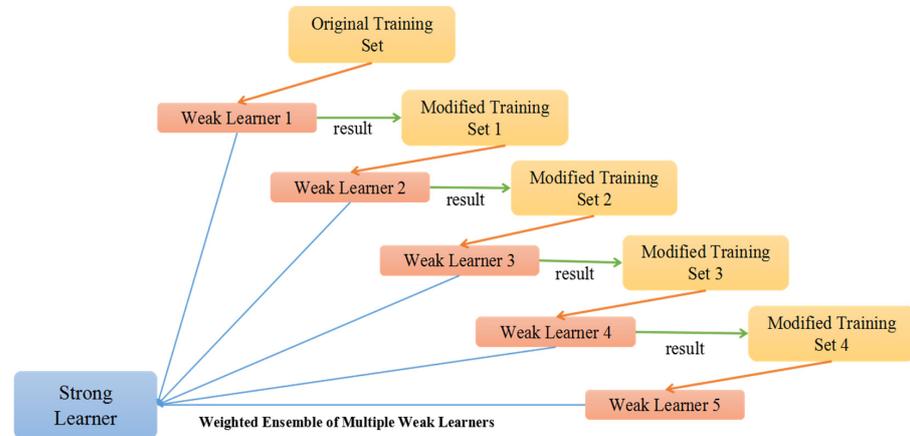


Figure 2. Flow chart of the Adaboost algorithm.

2.3. Assessment Criteria

The anomaly correlation coefficient (ACC) reflects the spatial similarity between the forecast and observation. This was established and recommended by WMO in 1996. ACC is calculated as follows:

$$ACC = \frac{\sum_{i=1}^N (\Delta x_i - \Delta \bar{x}_i) \times (\Delta f_i - \Delta \bar{f}_i)}{\sum_{i=1}^N (\Delta x_i - \Delta \bar{x}_i)^2 \times (\Delta f_i - \Delta \bar{f}_i)^2} \tag{3}$$

where Δf_i and $\Delta \bar{f}_i$ are the anomaly percentage (AP) prediction and the average value of AP prediction at the i th grid point, respectively. Δx_i and $\Delta \bar{x}_i$ are the PAP observation and the average value of AP observation at the i th grid point, respectively. N is the total number of grid points.

The temporal correlation coefficients (TCC) can provide a more accurate assessment of the forecasting capability of the model for each grid point anomaly from a statistical perspective. TCC is calculated as follows:

$$TCC = \frac{\sum_{i=1}^N (x_i - \bar{x}_i) \times (f_i - \bar{f}_i)}{\sum_{i=1}^N (x_i - \bar{x}_i)^2 \times (f_i - \bar{f}_i)^2} \tag{4}$$

where f_i and \bar{f}_i are the prediction and the average value of prediction at the i th grid point, respectively. x_i and \bar{x}_i are the observation and the average value of observation at the i th grid point, respectively. N is the total number of time series.

Root mean square error (RMSE) is the standard deviation of the prediction errors. RMSE is calculated as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (f_i - x_i)^2}{N}} \tag{5}$$

where f_i and x_i are the prediction and the observation at the i th grid. N is the total number of grid points.

3. Results

3.1. Prediction of the Skill of Models

To facilitate the comparison of the differences in forecasting skill between the models, Figure 3 depicts the distribution of each model's forecasting skill across the country. The time correlation coefficients (TCC) of the multi-year returns and actual precipitation for the BCC, ECMWF, NCEP, and JMA models are depicted in Figure 3a–d, respectively. The comparison of Figure 3a–d reveals that the forecasting accuracy varied considerably between regions. BCC (Figure 3a) had a high forecasting skill in the southwest, middle reaches of the Yangtze River, northern northeast region, and northern northwest region, whereas ECMWF (Figure 3b) had relatively large areas of high forecasting skill, with relatively high skill from the southwest to northern China and western Inner Mongolia. In the southwest to northern China region and western Inner Mongolia, the forecast accuracy was relatively high, whereas it was low in regions south of the Yangtze River, the northeast, and parts of the northwest. In Figure 3c, the distribution of NCEP forecasting skills was comparable to that of ECMWF, with the highest forecasting skills primarily in the southwest to Huang-huai regions, and the largest differences with ECMWF forecasting skills occurring in the Jiangnan and eastern northwest regions. ECMWF's predictive ability in these two regions was the exact opposite of NCEP's. Figure 3d demonstrates that the JMA had a high prediction skill in the region between the two river basins, but a low prediction skill in the region south of the Yangtze River and in the northeast.

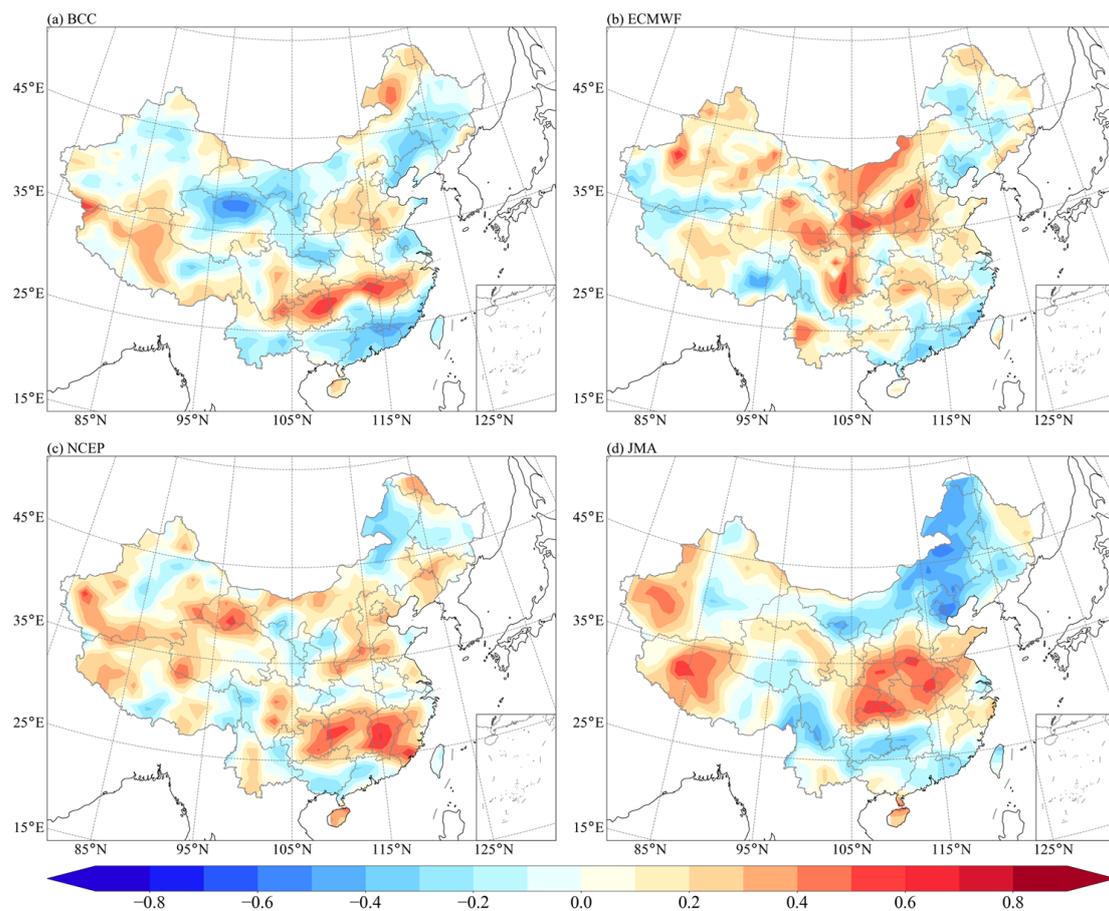


Figure 3. The time correlation coefficients between different models and observations (a) BCC, (b) ECMWF, (c) NCEP, and (d) JMA.

In Figure 4a–f, the characteristics of the TCC distributions for each of the four models are presented with respect to one another. Comparing the TCC between each model allows us to determine whether they are highly correlated. According to Figure 4a, the TCC

between BCC and ECMWF was higher in North China, the middle and lower Yangtze River, and South China, and its coefficient passed the 5% level of significance test, indicating that the forecasts of the two models were consistent in these areas; in Southwest China, the differences were larger, and the forecasts often contradicted each other. A high TCC between BCC and NCEP can be found in Figure 4b, mainly in the middle and upper reaches of the Yangtze River, in northern China, and in eastern northwest China. It can be seen from Figure 4c that regions with a higher TCC between BCC and JMA were primarily located in South China, Central China, and North China. Figure 4d illustrates regions with high TCC between ECMWF and NCEP that were primarily located in the middle and southern reaches of the Yangtze River. Despite the lower TCCs in eastern Yunnan, southern Northeast China, and the southeast coast, the correlation between ECMWF and JMA is strong in Figure 4e. In most areas, the TCC between the NCEP and the JMA models in Figure 4f is low, with the exception of the eastern northwest, western Yunnan, and some parts of the middle and lower reaches of the Yangtze River, where most TCCs were negative, indicating that the models made opposite predictions over time.

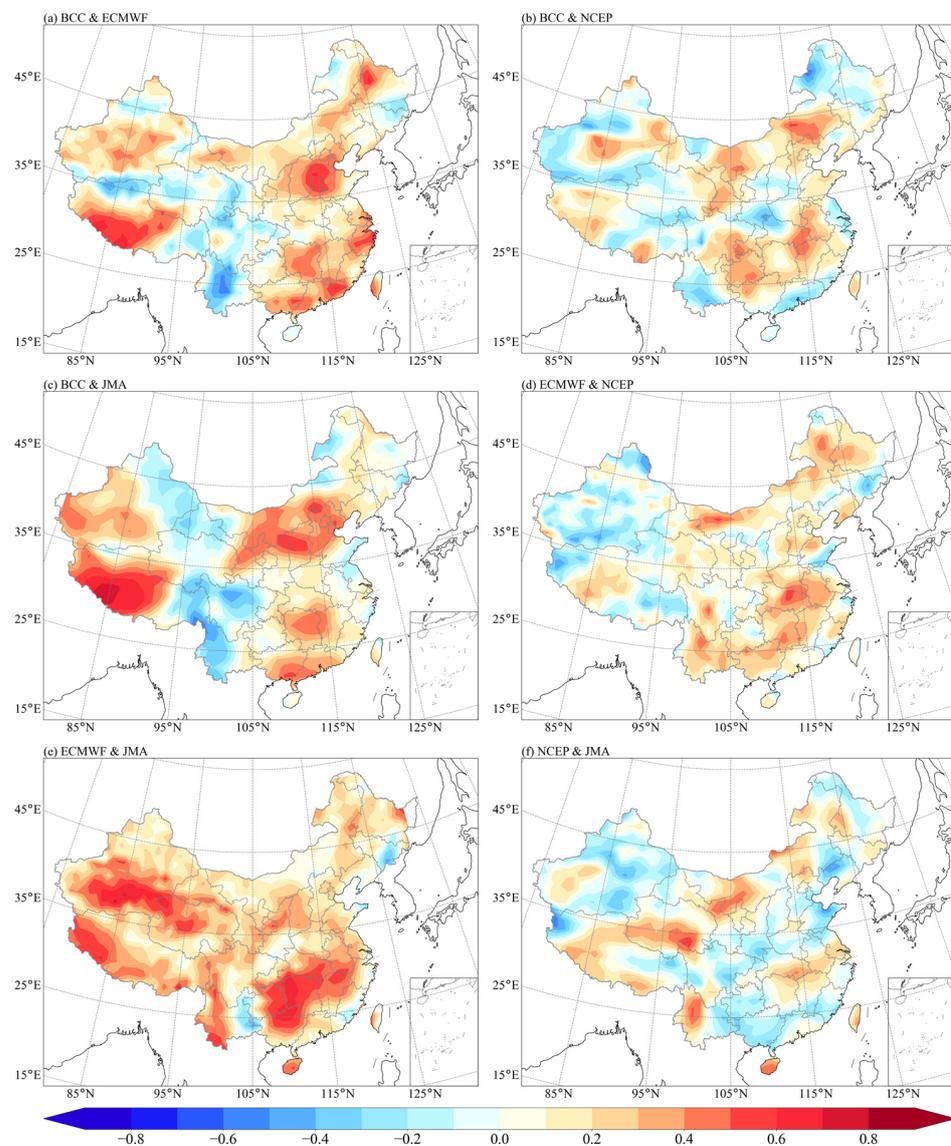


Figure 4. The time correlation coefficients between the different model prediction: (a) TCC between BCC and ECMWF, (b) TCC between BCC and NCEP, (c) TCC between BCC and JMA, (d) TCC between ECMWF and NCEP, (e) TCC between ECMWF and JMA, and (f) TCC between NCEP and JMA.

3.2. Parameter Optimization

The DT, RF, and AB algorithms have various hyperparameters, including feature selection criteria, feature partition criteria, maximum depth, leaf node minimum sample number, node partition minimum impure, maximum leaf node number, etc. Different configurations of hyperparameters have a significant effect on the simulation of the model. In essence, model learning refers to the process of adjusting the model parameters so that the observation data and the model prediction data become as close as possible. By adjusting the different configurations of parameters, the optimal model parameters are established, to achieve the optimal prediction effect. The maximum depth of the tree is the key tuning parameter in the CART, determining the complexity of the model.

As shown in Figure 5, a MME prediction with a maximum depth of three was constructed for summer precipitation in eastern China. As can be seen from the figure, the model used the JMA as the root node and whether the precipitation anomaly percentage was greater than -9.36% as the basis for node splitting. The ECMWF model and the NCEP model were used as the decision nodes, while ECMWF, NCEP, and the JMA were the decision nodes of the next layer. As can be seen from the figure, when the predicted value of JMA was greater than -9.36% , it entered the right-hand branch of the tree. For the right-hand branch, we determined whether the NCEP prediction was greater than 7.91% , in order to proceed to the next node. The terminal node and its corresponding precipitation prediction value were obtained. The same applied to the left-hand branch, where different configurations of the JMA and ECMWF models were used to obtain their corresponding precipitation predictions. From the tree structure, we can see that the BCC was not used in the tree, and therefore BCC was of relatively low reference value for the East China region. In this tree structure, the ECMWF, NCEP, and JMA do not act as nodes in the tree structure the same number of times. JMA was used as the root node of the whole tree structure and ECMWF acted as a leaf node many times in the branches for node splitting, whereas the NCEP only acted as two leaf nodes. Therefore, the importance of JMA and ECMWF was relatively higher in this tree structure.

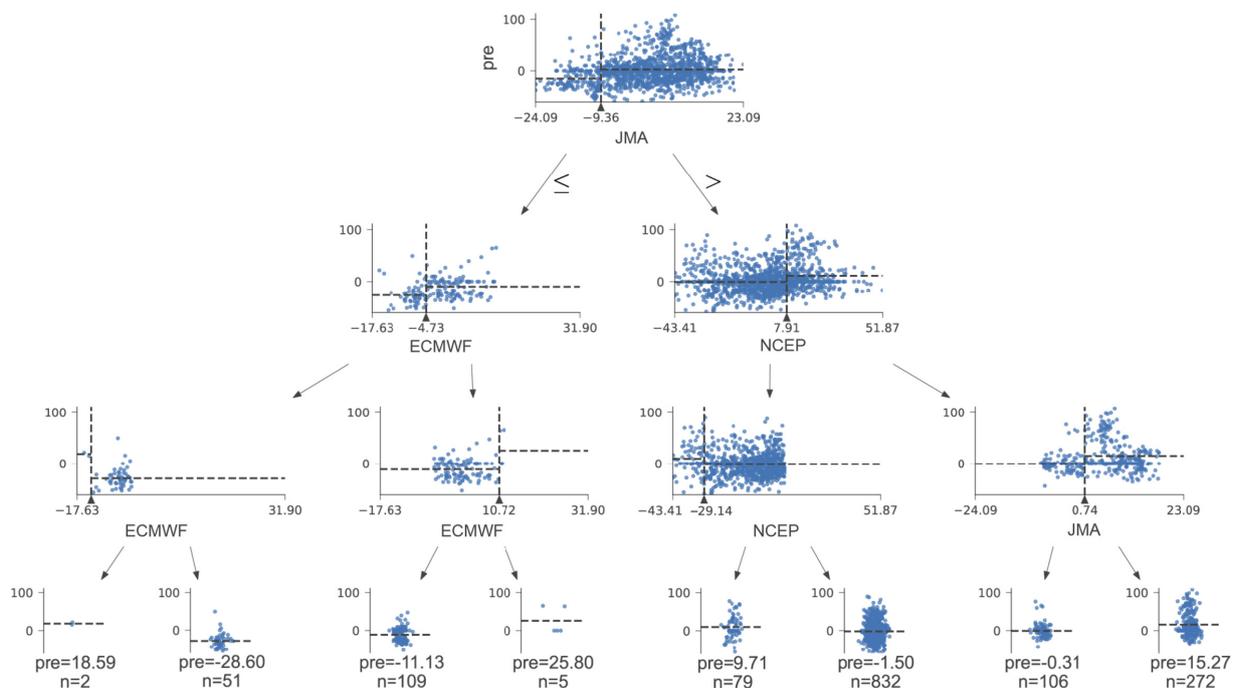


Figure 5. The decision tree structure with a maximum depth of three for summer precipitation prediction in eastern China. The dotted vertical line represents the value of node splitting and the dotted horizontal line represents the mean value of all points in the region.

To investigate the effect of the maximum depth of the tree on the prediction skills, cross-validation of the DT, RF, and AB algorithms was conducted using data from the last 30 years. For the four regions of South China, East China, North China, and Northeast China, Figure 6a–d show the variation of root mean square error (RMSE) with increasing maximum depth. A comparison of the RMSEs of the three ML algorithms for the South China region (Figure 6a) revealed that the RF algorithm had the lowest RMSE, regardless of the variation in maximum depth. DT algorithms had lower RMSEs than AB algorithms when the maximum depth was less than 6, and AB algorithms had lower RMSEs than DT algorithms when the maximum depth was greater. In both the RF and DT algorithms, the RMSE reached a minimum value when a maximum depth of 2 was reached, whereas in the AB algorithm, the RMSE reached its minimum value when a maximum depth of 1 was reached. According to the DT and RF algorithms, the optimal depth for South China is 2, while for the AB algorithm, the optimal depth is 1. For the East, North, and Northeast China regions, a comparison of the RMSE with maximum depth also showed that the RMSE was dependent on maximum depth. Based on the cross-validation results, the RMSEs of the three algorithms showed a trend of decreasing and then increasing according to the maximum depth of the tree, and the RMSE of the RF algorithm had the smallest RMSE across all regions. RMSE reached a global minimum for East and North China at a maximum depth of 2, and RMSE reached a global minimum for Northeast China at a maximum depth of 1. Generally, each algorithm minimized the RMSE at maximum depths below 2, which was a consequence of the small amount of model data. Consequently, when developing prediction models, the maximum depth of the tree should not exceed two.

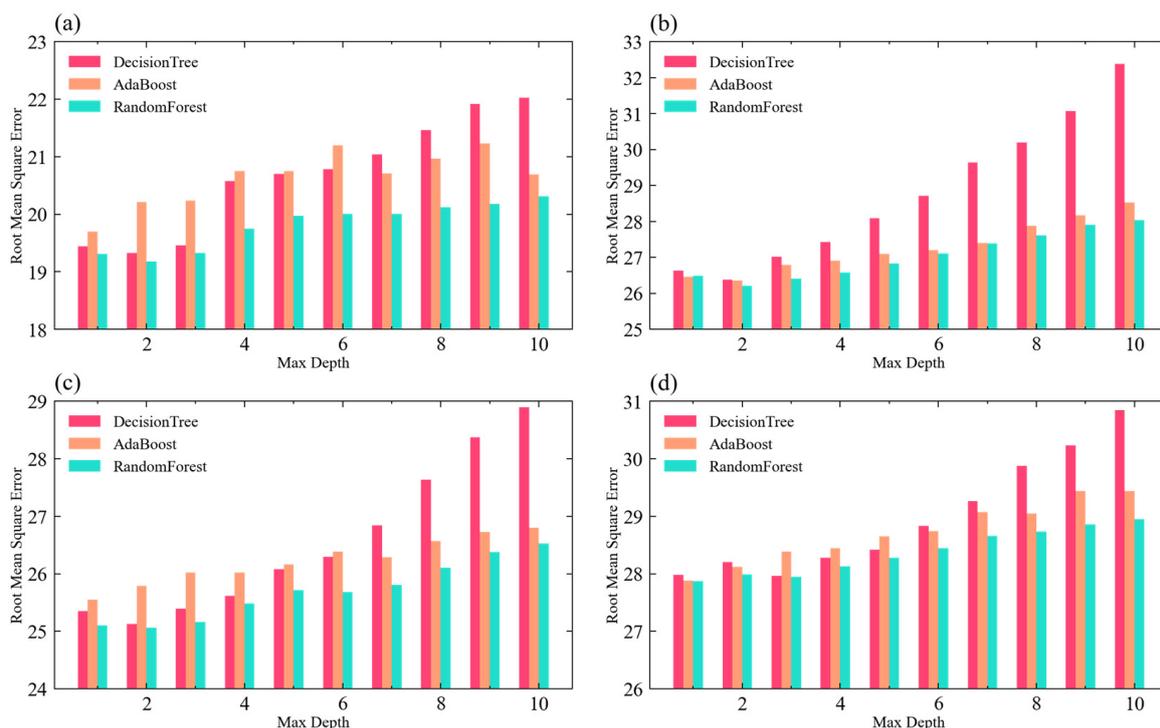


Figure 6. Variation of root mean square error with the maximum depth (a) South China, (b) East China, (c) North China, and (d) East North.

Additionally, when it comes to the AB and RF algorithms, the number of trees is an important parameter that affects the prediction skill. Figure 7a–d illustrates the effect of changing the number of trees in the AB algorithm and RF algorithm on the prediction skill in South China, East China, North China, and Northeast China, respectively. According to Figure 7a, the variation of RMSE was relatively similar for the AB and RF algorithms in South China as the number of trees increased. As the number of ensemble trees increased,

the cross-validation RMSE decreased rapidly when the number of trees was less than 10. In the presence of more than 10 trees, the decline in RMSE became significant and stabilized over time. RMSE essentially reached a minimum value when the number of trees reached 20. Further increases in the number of ensemble trees did not result in further decreases in the RMSE. The comparison of the three plots in Figure 7b–d shows that the effect of the number of integration trees on RMSE in East China, North China, and Northeast China was essentially the same as that in South China. As the number of ensemble trees increased, the RMSE for the three regions decreased rapidly at an early stage. Subsequently, with a further increase in the number of trees, the rate of decline decreased, and the RMSE reached a stable value when the number of trees reached 20 or more and no longer declined. Accordingly, the optimal number of integration trees for both the AB algorithm and the RF algorithm is about 20. Table 2 shows the hyperparameter selection of the three ML algorithms when carrying out independent prediction.

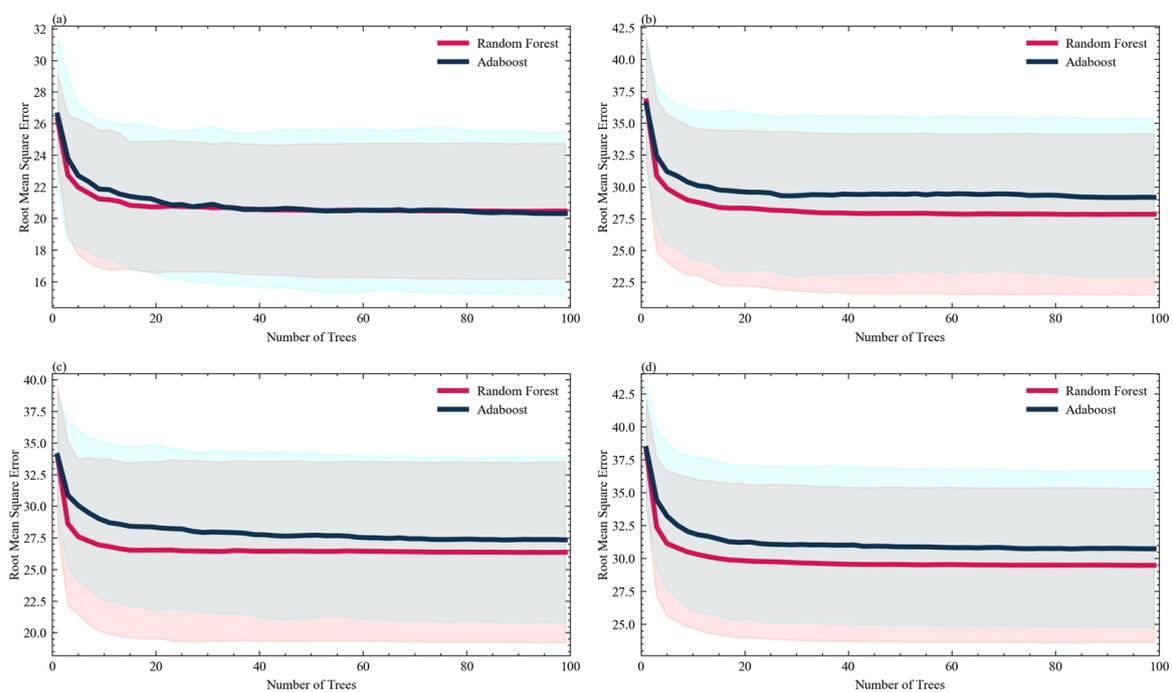


Figure 7. Variation of average RMSE of 10-fold cross-validation with the number of trees (a) South China, (b) East China, (c) North China, and (d) East North. The pink area is RMSE variation range of the 10-fold cross-validation for RF algorithm and the blue area is RMSE variation range of the 10-fold cross-validation for AB algorithm.

Table 2. The selection of key hyperparameters.

DT	Parameter	RF	Parameter	AB	Parameter
critierion	mse	critierion	mse	base estimator	decision tree
max depth	2	max depth	2	max depth	2
random state	0	n estimators	20	n estimators	20
min samples split	2	random state	0	random state	0

In the tree-based models, we can calculate the mean square error for each feature, and the feature importance is the normalized value of this mean square error reduction. This enabled us to calculate the importance of each model in the DT, RF, and AB algorithms. A distribution of the importance of each model in China could also be obtained, as shown in Figure 8.

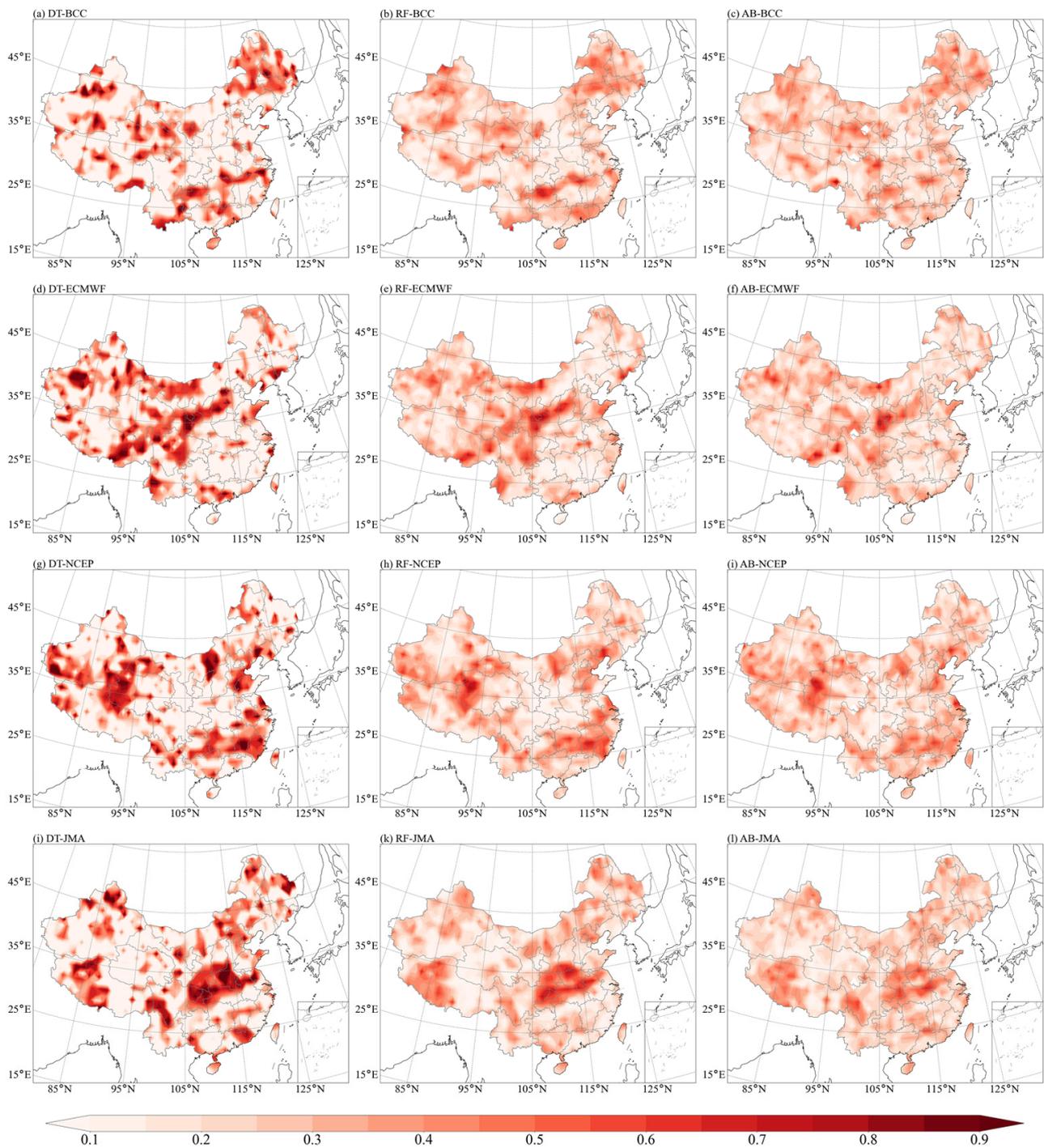


Figure 8. The importance of four models in the algorithms for decision trees, random forests, and AdaBoost: (a) BCC's importance in DT algorithm, (b) BCC's importance in RF algorithm, (c) BCC's importance in AB algorithm, (d) NCEP's importance in DT algorithm, (e) NCEP's importance in RF algorithm, (f) NCEP's importance in AB algorithm, (g) ECMWF's importance in DT algorithm, (h) ECMWF's importance in RF algorithm, (i) ECMWF's importance in AB algorithm, (j) JMA's importance in DT algorithm, (k) JMA's importance in RF algorithm, and (l) JMA's importance in AB algorithm.

As shown in Figure 8, each of the three ML algorithms, DT, RF, and AB, was evaluated based on its importance for each of the four regions. Figure 8a–c illustrates the importance of BCC in the three algorithms. Comparison of the three figures shows that the importance

of BCC in all three algorithms had similar distribution characteristics, all of them having a higher importance in the Yangtze River basin and the northeast region, while the importance of the DT algorithm was greater than the others. In part, this was due to the fact that both the RF and AB are ensemble algorithms, which build multiple decision trees, in order to utilize as much valid information as possible from all models without over-reliance on one model, as in the case of individual decision trees. Therefore, the difference in importance between the RF and AB algorithms was relatively low. Figure 8d–f shows the spatial distribution characteristics of the importance of ECMWF in each of the three algorithms, where it can be seen that ECMWF were highly significant in the Yellow River Basin, east of the northwest region, and west of the southwest region. According to Figure 8g–i, the importance of the NCEP model varied by region in the three algorithms, and the model mainly relied on the results of NCEP in most regions south of the Yangtze River, northern China, and western Northwest China. JMA mainly had a high importance in the Huaihe River basin, central China, and northern northeast China, as shown in Figure 8j–l. In general, the importance of each model varied widely from region to region. It can be seen from all graphs that both NCEP and JMA played a leading role in ensemble forecasting in eastern parts of China. Although BCC and ECMWF were less important than the other two models, they were of great importance in certain areas where the prediction accuracy was low, such as South China and Northeast China. ML algorithms can effectively evaluate the importance of each model for different regional zones and realize the complementary advantages of multiple models, resulting in an optimal MME prediction.

Comparing Figures 3 and 8, it can be seen that the importance of the model is more similar to that of TCC. The areas where a single model has high skill also has higher importance for MME. This is because the ML-based MME is essentially a regression of multi-models. If the prediction of a model is more accurate in a specific region, the machine learning method will inevitably use this prediction information more in the training process, thus increasing the importance of the model. The biggest difference between Figures 3 and 8 is that a model with lower predictive skill does not mean that the model is less important in the ML-based MME. For example, the prediction skill of a model is so low that predictions and observations are opposite in most years. It is also prediction information that can be used in ML algorithms. ML can obtain a relatively high accuracy by taking the inverse of the predictions of these models. Therefore, in such a training process, a model with low prediction skill could have a higher importance.

The validation results of the three ML algorithms were relatively similar, and in general, RF was the most stable of the three algorithms. Based on RF algorithms, Figure 9 shows the comparison between the MME prediction and actual summer precipitation in China for 2019–2021. According to Figure 9a, the summer precipitation in China in 2019 can be divided into two rain belts, from north to south, with the southern rain belt located in Jiangnan, while the northern rain belt is located in the northeast and east northwest, while North China, the Huaihe River basin, and Southwest China receive little rainfall. It can be observed that the integrated multi-model prediction results in Figure 9b were more accurate in predicting the overall distribution characteristics of the rainbands, as well as the general distribution characteristics of two rainbands in general. The precipitation predictions for Jiangnan and Northwest China are more accurate, while those for Northeast China are less accurate. Based on Figure 9c, China's summer precipitation in 2020 was relatively anomalous, with unusually high precipitation levels compared to climate averages, especially in the Yangtze River basin. It was only in a few small areas of southern China, southern northeast China, and central northwest China that the precipitation was less. Figure 9d depicts the general distribution of precipitation for 2020, with most areas in good agreement with the actual conditions, except for some areas in the northwest and northeast, where the actual conditions were reversed. However, the predictions differ significantly in magnitude from the actual conditions, and the anomalies in actual precipitation were not well predicted. Figure 9e represents the distribution of summer precipitation in China in 2021 as a rainband extending from northeast to southwest, mostly covering eastern

Inner Mongolia, north and central China, the Jianghuai basin, and the middle and lower Yangtze River reaches, with most other areas showing reduced precipitation. In Figure 9f, the prediction offered a relatively good interpretation of the rainband, although the extent of the rainband was larger than in reality. Forecasts and actuals differ mainly in the eastern part of the northwest region, in the northern part of the northeast region, and in the middle and lower reaches of the Yangtze River. In addition, the predictions were smaller than the actual results, when comparing the forecast and actual precipitation levels. The MME prediction for 2019–2021 better captured the overall distribution of summer precipitation in China, but the magnitude of the forecast differed significantly from that of the actual situation, which was significantly higher.

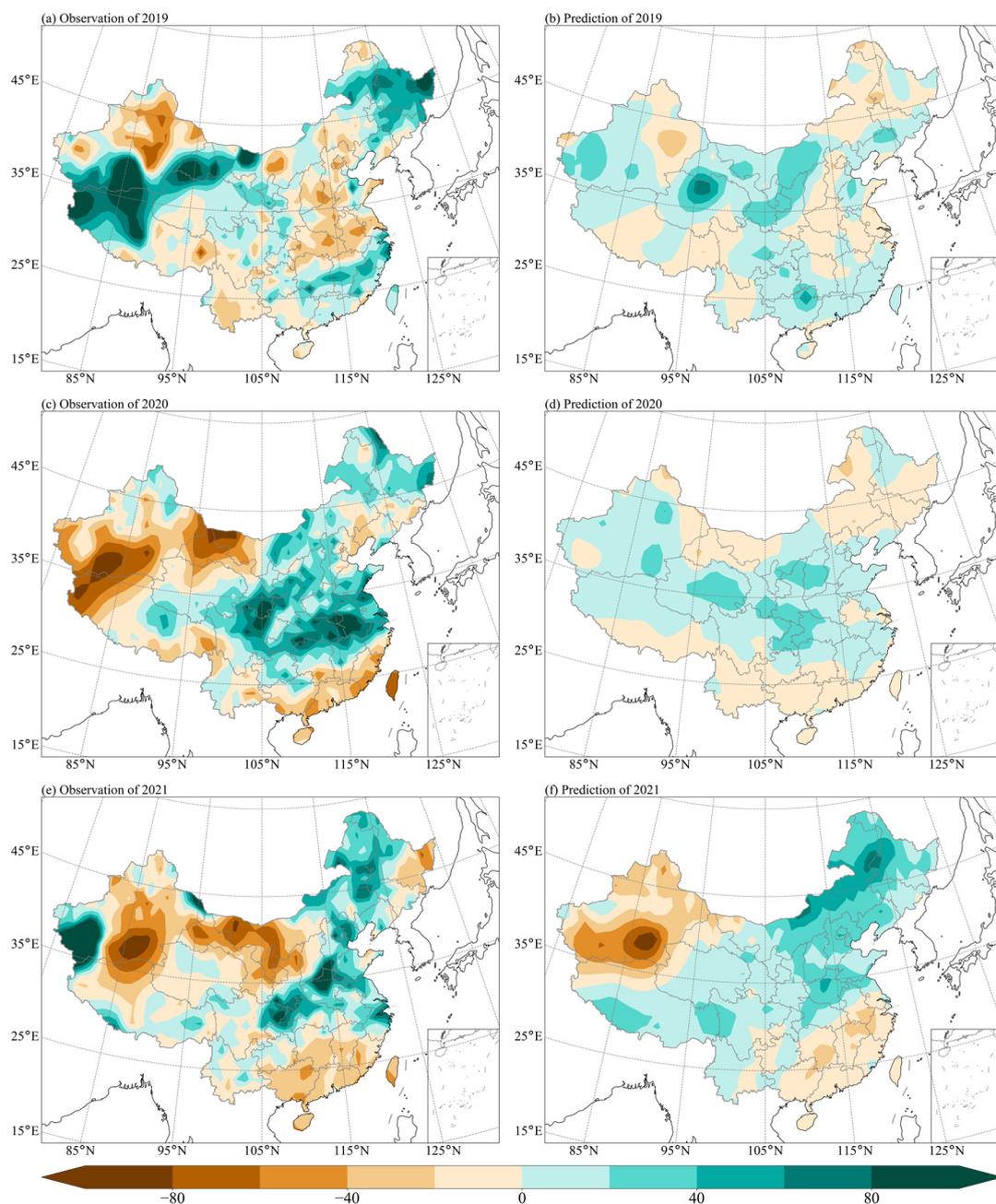


Figure 9. The comparison of summer precipitation anomaly percentage between the MME prediction and observations in China: (a) observation of 2019, (b) prediction of 2019, (c) observation of 2020, (d) prediction of 2020, (e) observation of 2021, and (f) prediction of 2021.

A quantitative evaluation of the forecasting skill was conducted by computing the ACC and RMSE of the four models, weighted average MME, and ML-based MME, as shown in Figure 10. When comparing the ACC of the various methods in Figure 10a, it can be seen that the single model's ACC was unstable and exhibited large interannual variations, but the two MME methods both had higher and more stable ACCs. It was determined that the mean ACC of ML-based MME was 0.3, an improvement of 0.09 over the weighted average MME of 0.21 for 2019–2021. The ML-based MME had a significant improvement over the other methods. RMSE comparison results in Figure 10b are similar to those of ACC. The ML-based MME had the lowest RMSEs. However, RMSE's improvement of ML-based MME was not as significant as ACC's.

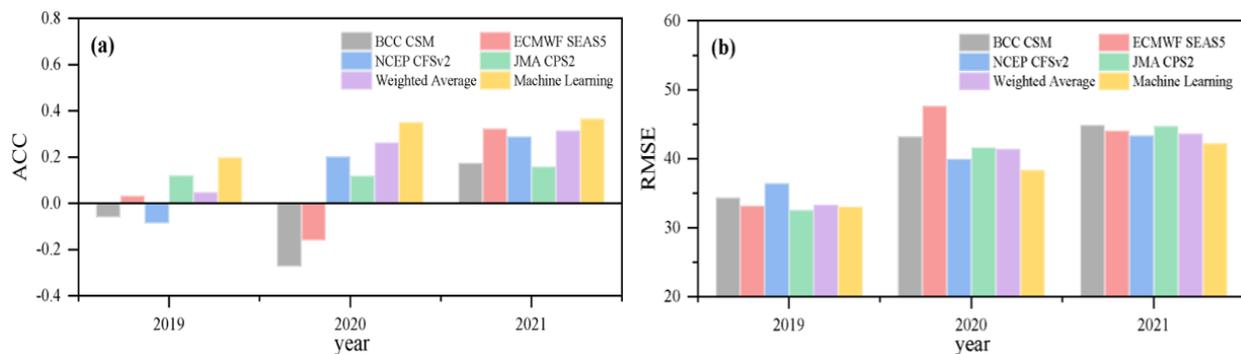


Figure 10. Comparison of the prediction accuracy among the four models, weighted average MME and ML-based MME: (a) ACC, (b) RMSE.

4. Conclusions and Discussion

This study was based on three tree-based ML algorithms for predicting summer precipitation in China with MME. Using BCC CSM1.1, ECMWF SEAS5, NCEP CFSv2, and JMA CPS2 model data, a MME prediction experiment was conducted of summer precipitation in China. On this basis, the applicability of ML algorithms for MME prediction was explored. The influence of hyperparameters, including maximum depth and number of trees, on the MME was investigated. Using optimum hyperparameters, MME forecasts were then created for various regions of China. The importance of each model in the ensemble prediction was quantified.

According to the cross-validation results, the RMSEs of the three algorithms declined and then increased, depending on the maximum depth of the tree, with the RF algorithm having the lowest RMSE across all regions. In light of the limited quantity of model data, each algorithm minimized RMSE at maximum depths below 2. Thus, based on the current data, the three algorithms were only able to predict precipitation anomalies that were positive or negative, and extreme precipitation was difficult to predict. In the case of the number of integration trees, the RMSE for all regions decreased rapidly as the number of ensemble trees increased at an early stage of the analysis. The rate of decline decreased as the number of trees was increased, and the RMSE reached a stable value when the number of trees reached 20 or more. Therefore, approximately 20 ensemble trees are optimal for both AB and RF algorithms.

In eastern parts of China, NCEP and JMA had a higher importance in MME. Despite the fact that BCC and ECMWF were less relevant than the other two models, they were of significant importance in certain areas where prediction accuracy is low, such as South China and Northeast China. MME prediction can be achieved with the use of machine learning algorithms that effectively evaluate the importance of each model for different regions, thereby utilizing the complementary advantages of multiple models.

The prediction results of summer precipitation from 2019 to 2021 implied that MME prediction with the ML algorithms represents a potential means for improving prediction skill. It is apparent that the ACC of a single model is unstable and exhibits large interannual

variations, and that MME have both higher and more stable ACCs. ML-based MME had a mean ACC of 0.3, an improvement of 0.09 over the weighted average MME of 0.21 for 2019–2021, which demonstrates a significant improvement over the other methods. However, throughout the study, it was evident that the improvement in prediction skill of ML-based MME was mainly due to a more accurate prediction of overall distribution characteristics, with little improvement in the prediction of extreme precipitations. In terms of development, MME prediction shows considerable promise in predicting summer precipitation in China.

Author Contributions: Conceptualization, Y.X.; methodology, J.Y.; supervision, X.X.; data curation, J.S.; writing—original draft preparation, review and editing, J.Y. and Y.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant numbers 42130610 and 41975098, Special Science and Technology Innovation Program for Carbon Peak and Carbon Neutralization of Jiangsu Province, grant numbers BE2022612, and the Key Fund of Jiangsu Meteorological Bureau, grant numbers KZ202102 and KZ202206.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The monthly precipitation data set were downloaded from CPC Merged Analysis of Precipitation (<https://psl.noaa.gov/data/gridded/data.cmap.html> accessed on 30 August 2022). The model prediction data in this paper were downloaded from the Beijing Climate Center (<http://forecast.bccsm.ncc-cma.net/web/channel-66.htm> accessed on 30 August 2022), the National Centers for Environmental Prediction (https://www.cpc.ncep.noaa.gov/products/CFSv2/CFSv2_body.html accessed on 30 August 2022), the European Centre for Medium-Range Weather Forecasts (<https://www.ecmwf.int/en/forecasts/datasets/set-v> accessed on 30 August 2022), and the Japan Meteorological Agency (<https://ds.da.jma.go.jp> accessed on 30 August 2022).

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

References

1. Rozante, J.R.; Moreira, D.S.; Godoy, R.C. Multi-model ensemble: Technique and validation. *Geosci. Model Dev.* **2014**, *7*, 2333–2343. [[CrossRef](#)]
2. Lorenz, E.N. A study of the predictability of a 28-variable atmospheric model. *Tellus* **1965**, *17*, 321–333. [[CrossRef](#)]
3. Chou, J.F. Short term climatic prediction: Present condition, problems and way out. *Bimon. Xinjiang Meteorol.* **2003**, *26*, 1–4.
4. Feng, G.L.; Zhao, J.H.; Yang, J. *Dynamic Statistical Prediction of Precipitation during Flood Season in China*; Science Press: Beijing, China, 2015.
5. Feng, G.L.; Yang, J.; Zhi, R. Improved prediction model for flood-season rainfall based on a nonlinear dynamics-statistic combined method. *Chaos Solitons Fractals* **2020**, *140*, 110160. [[CrossRef](#)]
6. Yun, W.T.; Stefanova, L.; Mitra, A.K. A multi-model superensemble algorithm for seasonal climate prediction using DEMETER forecasts. *Tellus A Dyn. Meteorol. Oceanogr.* **2005**, *57*, 280–289. [[CrossRef](#)]
7. Tebaldi, C.; Knutti, R. The use of the multi-model ensemble in probabilistic climate projections. *Philos. Trans. R. Soc.* **2007**, *365*, 2053–2075. [[CrossRef](#)]
8. Tebaldi, C.; Mearns, L.O.; Nychka, D. Regional probabilities of precipitation change: A Bayesian analysis of multimodel simulations. *Geophys. Res. Lett.* **2004**, *31*, L24213. [[CrossRef](#)]
9. Kharin, V.V.; Zwiers, F.W. Notes and correspondence: Climate predictions with multi-model ensembles. *J. Climate* **2002**, *15*, 793–799. [[CrossRef](#)]
10. Liu, C.Z.; Du, L.M.; Ke, Z.J.; Chen, L.J.; Jia, X.L.; Ai, W.X. Multi-model Downscaling Ensemble Prediction in National Climate Center. *J. Appl. Meteorol. Sci.* **2013**, *24*, 677–685.
11. Krishnamurti, T.N.; Kishtawal, C.M.; LaRow, T.E.; Bachiochi, D.R.; Zhang, Z. Improved weather and seasonal climate forecasts from multi-model superensemble. *Science* **1999**, *285*, 1548–1550. [[CrossRef](#)]
12. Krishnamurti, T.N.; Kishtawal, C.M.; Shin, D.W.; Williford, C.E. Multi-model superensemble forecasts for weather and seasonal climate. *J. Climate* **2000**, *13*, 4196–4216. [[CrossRef](#)]
13. Johnson, C.; Swinbank, R. Medium-range multimodel ensemble combination and calibration. *Q. J. R. Meteorol. Soc.* **2009**, *135*, 777–794. [[CrossRef](#)]
14. Roy, S.K.; Durai, V.R. Application of multimodel ensemble techniques for real time district level rainfall forecasts in short range time scale over Indian region. *Meteorol. Atmos. Phys.* **2010**, *106*, 19–35.

15. Kotal, S.D.; Bhowmik, S.K. A multimodel ensemble (MME) technique for cyclone track prediction over the North Indian Sea. *Geofizika* **2011**, *28*, 275–291.
16. Chen, L.J.; Xu, L.; Wang, Y.G. Application of Superensemble to Precipitation Prediction in China during Flood Season. *Meteor Mon.* **2005**, *31*, 52–54.
17. He, S.P.; Wang, H.J.; Li, H.; Zhao, J.Z. Machine learning and its potential application to climate prediction. *Trans. Atmos. Sci.* **2021**, *44*, 26–38.
18. Weyn, J.A.; Durran, D.R.; Canxana, R. Can machines learn to predict weather? using deep learning to predict gridded 500-hPa geopotential height from historical weather data. *J. Adv. Model Earth Syst.* **2019**, *11*, 2680–2693. [[CrossRef](#)]
19. Huntingford, C.; Jeffers, E.S.; Bonsall, M.B. Machine learning and artificial intelligence to aid climate change research and preparedness. *Env. Res. Lett.* **2019**, *14*, 124007. [[CrossRef](#)]
20. Ham, Y.G.; Kim, J.H.; Luo, J.J. Deep learning for multi-year ENSO forecasts. *Nature* **2019**, *573*, 568–572. [[CrossRef](#)]
21. Toms, B.A.; Barnes, E.A.; Ebert, U.I. Physically interpretable neural networks for the geosciences: Applications to earth system variability. *J. Adv. Model Earth Syst.* **2020**, *12*, e2019MS002002. [[CrossRef](#)]
22. Najafabadi, M.M.; Villanustre, F.; Khoshgoftaar, T.M. Deep learning applications and challenges in big data analytics. *J. Big Data* **2015**, *2*, 1–21. [[CrossRef](#)]
23. Zhang, C.S. Challenges in machine learning. *Sci. China Inf. Sci.* **2013**, *43*, 1621–1623.
24. Saha, S. The NCEP Climate Forecast System Version 2. *J. Climate* **2014**, *27*, 2185–2208. [[CrossRef](#)]
25. Johnson, S.J.; Stockdale, T.N.; Ferranti, L. SEAS5: The new ECMWF seasonal forecast system. *Geosci. Model Dev.* **2019**, *12*, 1087–1117. [[CrossRef](#)]
26. Takaya, Y.S.; Hirahara, T.; Yasuda, S.; Matsueda, T.; Toyoda, Y.; Fujii, H.; Sugimoto, C.; Matsukawa, I.; Ishikawa, H.; Mori, R.; et al. Japan Meteorological Agency/Meteorological Research Institute-Coupled Prediction System version 2 (JMA/MRI-CPS2): Atmosphere-land-ocean-sea ice coupled prediction system for operational seasonal forecasting. *Clim. Dyn.* **2018**, *3*, 751–765. [[CrossRef](#)]
27. Wu, T.; Lu, Y.; Fang, Y. The Beijing Climate Center climate system model (BCC-CSM): The main progress from CMIP5 to CMIP6. *Geosci. Model Dev.* **2019**, *12*, 1573–1600. [[CrossRef](#)]
28. Loh, W.Y. Classification and regression trees. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2011**, *1*, 14–23. [[CrossRef](#)]
29. Freund, Y.; Schapire, R.E. Experiments with a new boosting algorithm. *Icml* **1996**, *96*, 148–156.
30. Smith, P.F.; Ganesh, S.; Liu, P. A comparison of random forest regression and multiple linear regression for prediction in neuroscience. *J. Neurosci. Methods* **2013**, *220*, 85–91. [[CrossRef](#)]