

Dynamical and Machine Learning Hybrid Seasonal Prediction of Summer Rainfall in China

Jialin WANG, Jing YANG, Hong-Li REN, Jinxiao LI, Qing BAO, Miaoni GAO

Citation: , 2021: Dynamical and Machine Learning Hybrid Seasonal Prediction of Summer Rainfall in China. *J. Meteor. Res.*, **35**(4), 583–593, doi: [10.1007/s13351-021-0185-0](https://doi.org/10.1007/s13351-021-0185-0)

View online: <http://jmr.cmsjournal.net/article/doi/10.1007/s13351-021-0185-0>

Related articles that may interest you

[A Multivariable Approach for Estimating Soil Moisture from Microwave Radiation Imager \(MWRI\)](#)

Journal of Meteorological Research. 2020, 34(4), 732 <https://doi.org/10.1007/s13351-020-9203-x>

[Seasonal Rainfall Forecasts for the Yangtze River Basin of China in Summer 2019 from an Improved Climate Service](#)

Journal of Meteorological Research. 2020, 34(5), 904 <https://doi.org/10.1007/s13351-020-0049-z>

[Seasonal Prediction of Boreal Winter Rainfall over the Western Maritime Continent during ENSO](#)

Journal of Meteorological Research. 2020, 34(2), 294 <https://doi.org/10.1007/s13351-020-9181-z>

[Forecasting Different Types of Convective Weather: A Deep Learning Approach](#)

Journal of Meteorological Research. 2019, 33(5), 797 <https://doi.org/10.1007/s13351-019-8162-6>

[Seasonal Climate Prediction Models for the Number of Landfalling Tropical Cyclones in China](#)

Journal of Meteorological Research. 2019, 33(5), 837 <https://doi.org/10.1007/s13351-019-8187-x>

[Upper-Ocean Dynamical Features and Prediction of the Super El Nio in 2015/16: A Comparison with the Cases in 1982/83 and 1997/98](#)

Journal of Meteorological Research. 2017, 31(2), 278 <https://doi.org/10.1007/s13351-017-6194-3>

Dynamical and Machine Learning Hybrid Seasonal Prediction of Summer Rainfall in China

Jialin WANG¹, Jing YANG^{1,2*}, Hong-Li REN³, Jinxiao LI⁴, Qing BAO⁴, and Miaoni GAO⁵

¹ State Key Laboratory of Earth Surface Process and Resource Ecology/Key Laboratory of Environmental Change and Natural Disaster, Faculty of Geographical Science, Beijing Normal University, Beijing 100875

² Southern Marine Science and Engineering Guangdong Laboratory, Guangzhou 511458

³ State Key Laboratory of Severe Weather, Chinese Academy of Meteorological Sciences, Beijing 100081

⁴ State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics (LASG), Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029

⁵ Institute for Disaster Risk Management, School of Geographical Sciences, Nanjing University of Information Science & Technology, Nanjing 210044

(Received November 5, 2020; in final form March 28, 2021)

ABSTRACT

Seasonal prediction of summer rainfall is crucial to reduction of regional disasters, but currently it has a low prediction skill. We developed a dynamical and machine learning hybrid (MLD) seasonal prediction method for summer rainfall in China based on circulation fields from the Chinese Academy of Sciences (CAS) Flexible Global Ocean–Atmosphere–Land System Model finite volume version 2 (FGOALS-f2) operational dynamical prediction model. Through selecting optimum hyperparameters for three machine learning methods to obtain the best fit and least overfitting, an ensemble mean of the random forest and gradient boosting regression tree methods was shown to have the highest prediction skill measured by the anomalous correlation coefficient. The skill has an average value of 0.34 in the historical cross-validation period (1981–2010) and 0.20 in the 10-yr period (2011–2020) of independent prediction, which significantly improves the dynamical prediction skill by 400%. Both reducing overfitting and using the best dynamical prediction are important in applications of the MLD method and in-depth analysis of these warrants a further investigation.

Key words: seasonal rainfall prediction, statistical–dynamical model, machine learning

Citation: Wang, J. L., J. Yang, H.-L. Ren, et al., 2021: Dynamical and machine learning hybrid seasonal prediction of summer rainfall in China. *J. Meteor. Res.*, **35**(4), 583–593, doi: 10.1007/s13351-021-0185-0.

1. Introduction

Summer rainfall in China shows a prominent year-to-year variability (Wang et al., 2009) and can cause disastrous droughts and flooding (Chen and Huang, 2008; Li et al., 2011). Accurate prediction of seasonal rainfall is valuable in regional disaster prevention and policy making. Although global dynamical models have been improved in recent years (Yang et al., 2008; MacLachlan et al., 2015; Alessandri et al., 2018; Chevuturi et al., 2019) and multimodel ensembles have been widely applied in seasonal prediction (Alessandri et al., 2018; Ren et al., 2019), the pure dynamical seasonal prediction of

rainfall usually has a considerably lower skill than predictions based on the circulation fields, despite decades of efforts (Wang et al., 2009; Alessandri et al., 2018). To overcome the current limitations of dynamical seasonal prediction, researchers used statistical methods to increase the prediction skill (Badr et al., 2014; Bett et al., 2018).

Two major types of statistical strategies have been applied to improve the seasonal prediction of rainfall: empirical statistical methods and statistical–dynamical methods. Empirical statistical methods usually establish a statistical relationship between the precursors and predictands in observational data (Jia and Zhu 2010; Pour et

Supported by the National Natural Science Foundation of China (42022034, 41775071, and U1811464) and China National Key Research and Development Program [Early Warning and Prevention of Major Natural Disaster (2018YFC1506005)].

*Corresponding author: yangjing@bnu.edu.cn

© The Chinese Meteorological Society and Springer-Verlag Berlin Heidelberg 2021

al., 2020), which usually have a higher prediction skill with low computational costs in the historical fitting years. However, its prediction skill tends to be unstable in different periods as a result of unstable precursors or shortages in physical linkages (Goddard et al., 2001; Gao et al., 2018).

Statistical–dynamical methods aim to correct model outputs based on the statistical relationship between the model outputs and predictands (Lee et al., 2013; Lim et al., 2015). Linear statistical methods (e.g., linear regression) have been widely used (Xing et al., 2016; Rana et al., 2018; Strazzo et al., 2019), but are usually unable to capture realistic nonlinear relationships (Goddard et al., 2001). Although some nonlinear statistical models have been applied for regional seasonal predictions (Saha et al., 2017) and to predict the El Niño–Southern Oscillation (Ham et al., 2019), nonlinear statistical methods, including machine learning combined with dynamical outputs, have not been applied to the prediction of seasonal rainfall in China. Overfitting, which has not been widely recognized or well solved, is the most challenging problem in machine learning methods (Tetko et al., 1995; Lever et al., 2016).

We attempted to improve the seasonal prediction skill of rainfall in China by selecting the optimum machine learning method from three nonparametric methods, determining the optimum hyperparameters for each machine learning method to obtain the best fit and the least overfitting, and selecting suitable dynamical circulation fields; namely, a dynamical and machine learning hybrid (MLD) approach. In addition, the following scientific questions were investigated. To what extent does the nonparametric MLD improve the skill of the dynamical prediction? To what extent does the MLD depend on the dynamical model performance?

2. Models, datasets, and methods

2.1 Dynamical models

Dynamical seasonal prediction datasets are retrieved from an operational dynamical prediction system using the Flexible Global Ocean–Atmosphere–Land System Model finite volume version 2 (FGOALS-f2), which is a coupled global climate model developed by the State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics, Institute of Atmospheric Physics, Chinese Academy of Sciences (Bao et al., 2019; He et al., 2019). The prediction system of FGOALS-f2 (hereafter referred to as FGOALS) has successfully predicted a weak El Niño event with a non-

classical type of El Niño or a central Pacific El Niño in 2019 (Bao et al., 2019).

The seasonal prediction used in this study initializes on March 20 and predicts the following six months of each year. Both a seasonal reforecast and a real-time seasonal prediction are used. The reforecast covers the period 1981–2017, combined with real-time prediction from 2018 to 2020, with 24 ensemble members and $1^\circ \times 1^\circ$ global grid (Bao et al., 2019). Seven dominant meteorological fields influencing China—including the sea-level pressure (SLP), the U and V wind fields at 850 hPa (U850 and V850), the geopotential height at 500 hPa (Z500), and the U and V wind fields and temperature at 200 hPa (U200, V200, and T200)—are chosen based on previous research on the mechanisms of summer rainfall in China (Gong and Ho, 2002; Shen et al., 2011; Zhu et al., 2014) and seasonal prediction studies (Wang and Zhu, 2001; Nan and Li, 2003; Fan, 2006; Yin and Wang, 2016). The ensemble mean for each variable is applied before preprocessing.

The North American Multi-Model Ensemble (NMME) hindcast and real-time monthly precipitation prediction initializing on March 8 provided by the Climate Prediction Center of the National Weather Service (www.cpc.ncep.noaa.gov/products/NMME/) are also applied for comparison (Kirtman et al., 2014). The multimodel ensemble mean is used for comparison. Table 1 briefly describes the selected models in the NMME.

2.2 Observational datasets

Rainfall datasets from 160 stations in China for 1981–2020 obtained from the China Meteorological Administration (<http://data.cma.cn>) are used to feed the machine learning algorithm and evaluate the MLD results. To evaluate the circulation prediction of the output from the FGOALS model, the observational monthly mean circulation (including SLP, U850, V850, Z500, U200, V200, and T200) from the ECMWF Re-Analysis Interim (ERA-Interim) dataset (Dee et al., 2011) (<https://apps.ecmwf.int/datasets/data/interim-full-mod/levtype=pl/>) are retrieved for 1981–2018. To evaluate the precipitation prediction of FGOALS model output, the monthly mean global rainfall is obtained from the Global Precipitation Climatology Project (GPCP) Version 2.3 combined precipitation dataset (Adler et al., 2003) provided by the Physical Sciences Division of the Earth System Research Laboratory of the NOAA Office of Oceanic and Atmospheric Research (<ftp://ftp.cdc.noaa.gov/Datasets/gpcp/precip.mon.mean.nc>).

2.3 Machine learning methods

We used three nonparametric machine learning meth-

Table 1. Descriptions of the NMME models used in this study

Model name	Institution	No. of ensemble members	Period
CMC1-CanCM3	Canada's Centre for Climate Modeling and Analysis	10	1982–2009
CMC2-CanCM4	Canada's Centre for Climate Modeling and Analysis	10	1982–2009
COLA-RSMAS-CCSM3	National Center for Atmospheric Research	10	1982–2009
COLA-RSMAS-CCSM4	National Center for Atmospheric Research	24	1982–2009
GFDL-CM2p1-aer04	NOAA's Geophysical Fluid Dynamics Laboratory	10	1982–2009
GFDL-CM2p5-FLOR-A06	NOAA's Geophysical Fluid Dynamics Laboratory	12	1982–2009
GFDL-CM2p5-FLOR-B01	NOAA's Geophysical Fluid Dynamics Laboratory	12	1982–2009
NCEP-CFSv2	NOAA's National Centers for Environmental Prediction	24	1982–2009
CMC1-CanCM3	Canada's Centre for Climate Modeling and Analysis	10	2012–2020
CMC2-CanCM4	Canada's Centre for Climate Modeling and Analysis	10	2012–2020
GFDL-CM2.1	NOAA's Geophysical Fluid Dynamics Laboratory	10	2011–2020
GFDL-CM2p5-FLOR-A06	NOAA's Geophysical Fluid Dynamics Laboratory	12	2014–2020
GFDL-CM2p5-FLOR-B01	NOAA's Geophysical Fluid Dynamics Laboratory	12	2014–2020
NASA-GEOSv2	NASA's Goddard Space Flight Center	11	2018–2020
NCAR-CCSM4	National Center for Atmospheric Research	10	2014–2020
NCEP-CFSv2	NOAA's National Centers for Environmental Prediction	32	2011–2020

ods: epsilon-sensitive support vector regression (SVR; Drucker et al., 1996), random forest regression (RF; Breiman, 2001; Liaw and Wiener, 2002), and gradient boosting regression trees (GBRT; Friedman, 2001, 2002). SVR is often used for machine learning projects with high dimension input features. RF and GBRT are the two most frequently used decision tree based machine learning algorithms and feature the bagging and boosting of regression (decision) trees, respectively. The three methods are implemented in Python's machine learning library Scikit-learn (Pedregosa et al., 2011). Table 2 lists the major hyperparameters considered.

2.3.1 Preprocessing and input features

Before building machine learning models, both the input features (i.e., the dynamical model output) and the original regression target (i.e., the summer precipitation anomaly) are preprocessed. Global gridded circulation fields are used instead of manually selecting the specific domain for each circulation variable. Before inputting the circulation data in the machine learning models, the output circulation variables of FGOALS are preprocessed to reduce overfitting and to improve the speed of training. First, the grid size of the circulation field is interpolated

from 384×192 to 60×31 using bilinear interpolation. This decreases the number of input features from 73,728 to 1860 for every variable. The field data are then standardized by removing the mean of all grids between 1981 and 2010 and scaling to unit variance for each variable. The values of the circulation variables in all the grids are then flattened to vectors, such that each original grid becomes an input feature in the machine learning method.

2.3.2 Original regression target and grid search

Based on the June–July–August mean precipitation anomaly from 1981 to 2010, empirical orthogonal functions (EOF) analysis is used to extract the leading dominant pattern modes (Figs. 1a–c) of the rainfall anomaly in China. Given the leading pattern modes and corresponding principal components, the precipitation anomaly can be estimated by accumulating the product of the pattern modes and the corresponding principal components. The regression of the precipitation anomaly of 160 stations can be approximately transformed into the regression of the leading principal components. Figure 1d shows the spatial percentage variance distribution of the reconstructed rainfall anomaly from the first three EOF modes against the total variance. The result shows that the first

Table 2. Hyperparameters considered in SVR, RF, and GBRT machine learning methods

Machine learning method	Major hyperparameter
SVR	Kernel (radial based function, sigmoid, polynomial, linear) C (penalty parameter) ϵ (ϵ in the ϵ -SVR model)
RF	Gamma (kernel coefficient, exclude polynomial kernel) or degree (for polynomial kernel only) Number of estimators Maximum depth of tree Minimum number of samples required to split an internal node Minimum number of samples required to be at a leaf node Number of features to consider when looking for the best split
GBRT	Learning rate Maximum depth of tree Minimum number of samples required to split an internal node Minimum number of samples required to be at a leaf node Number of features to consider when looking for the best split

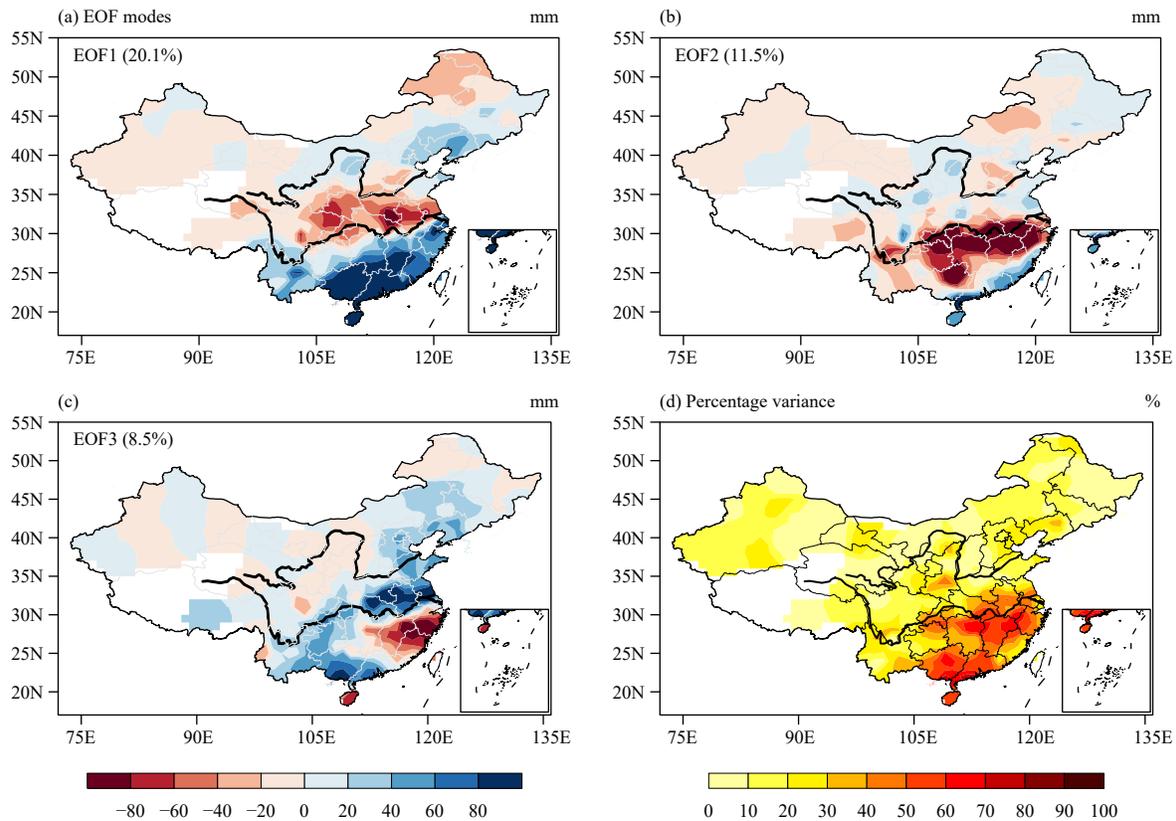


Fig. 1. (a–c) First three EOF modes using the 1981–2010 summer rainfall anomaly with observational rainfall from 160 China Meteorological Administration stations. (d) Spatial percentage variance of the reconstructed rainfall anomaly from the first three EOF modes against the total variance.

three EOF modes can explain a large fraction of the variance over the mid-to-lower reaches of the Yangtze River basin and southern China, which account for more than 60% of the variance.

In machine learning, hyperparameters are a group of model configurations or parameters that need to be pre-set before learning begins. Overfitting is often used to describe an analysis that fits too closely with a particular set of data, but is unable to fit additional data or predict future observations. To examine the fitting and overfitting of machine learning models, a single dataset is often split into a training set and a testing set. The training set is used to fit the model and the testing set is used to examine the skill of that model with unknown data.

To reach the best fitting and the least overfitting possible, we intentionally choose the optimum combinations of the gridded circulation variables and the corresponding hyperparameters for each machine learning method. A grid search that simply exhausts all possible combinations of hyperparameter options is applied for this optimum selection (Bergstra et al., 2011). During each search, we apply six-fold cross-validation (Picard and Cook, 1984; Zeng et al., 2011) and use the associ-

ated R^2 scores (Colin Cameron and Windmeijer, 1997) to calculate both the training set and the testing set in the search:

$$R^2 = 1 - \frac{\sum_i^n (O_i - P_i)^2}{\sum_i^n (O_i - \bar{O})^2},$$

where O_i represents the observational value in year i , P_i is the predicted value in year i and n is the total number of years; \bar{O} is the multiyear averaged observation; and R^2 is a general metric for the goodness-of-fit that varies from $-\infty$ to 1 and a larger R^2 means a better fit ($R^2 = 1$ if the prediction is perfectly fitted; $R^2 = 0$ if the prediction is the climatology).

2.3.3 Validation

Using the optimum combination of circulation variables/hyperparameters, the MLD seasonal prediction is validated for the independent prediction of individual years from 2011 to 2020. The independent prediction contains no future information. In this instance, we use the temporal correlation coefficient (TCC) to validate the prediction performance of the principal components. The

ensemble mean of the MLDs refers to the ensemble mean of the MLD-predicted principal components. Because the final prediction target is the pattern of rainfall anomalies against the climatology (1981–2010), we eventually reconstruct the rainfall anomaly distribution using the observed dominant EOF spatial patterns and the MLD-predicted principal components. We use the pattern correlation coefficient (PCC; sometimes called the anomaly correlation coefficient) and root mean square error (RMSE) to validate the final prediction performance of the MLD for the rainfall anomaly of 160 stations.

3. Results

To improve the computational efficiency, the first three dominant EOF modes of rainfall in China, which can account for 40.1% of the total variance (Figs. 1a–c), are extracted for machine learning regression. According to previous studies (Pang et al., 2014; Xing et al., 2016), the first three principal modes are physically related to the dominant tropical ocean interannual variation and remain stable regardless of the selected years. Further sensitivity tests for the number of leading EOF modes also show that the three leading EOF modes have best independent skills (figure omitted).

3.1 Determination of the optimum hyperparameters and circulation variables

We first need to determine the optimum hyperparameters in each machine learning method. In this instance, the optimum hyperparameters simultaneously produce the best fit and least overfitting. The R^2 scores are applied as a measurement and are calculated for both the training and testing sets of each principal component for the three machine learning methods during the grid search.

To select the optimum circulation variables, we first select the five top realistic variables (i.e., SLP, U850, Z500, U200, and T200) according to their dynamical prediction skill ($> 90\%$ confidence level; Fig. 2). There are a total of 31 combinations of these 5 circulation variables. A grid search is then applied to examine all combinations of these circulation variables. A six-fold cross-validation is applied for each hyperparameter and circulation variable.

Thirty-year data from 1981 to 2010 (the historical cross-validation period) are used in each cross-validation. Taking the first fold in the cross-validation as an example, the first 25 years of data are used as the training set to train a machine learning model and the remaining 5 years are used as the testing set. The machine learning

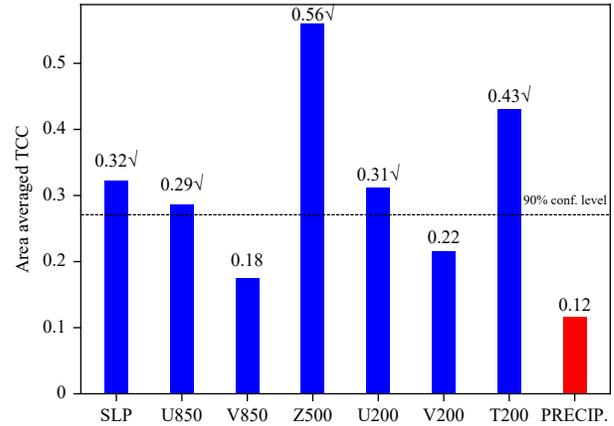


Fig. 2. TCC between the FGOALS model output and the observations for the dominant circulation variables and precipitation averaged over China (0° – 55° N, 70° – 160° E) during 1981–2018. The TCCs that pass Student’s t -test at the 90% confidence level are suffixed with the symbol $\sqrt{}$.

model makes predictions on both the training and the testing sets and R^2 is used to evaluate the model on both sets to obtain the training and testing scores in this fold. This process is applied six times until the data from each year have been used in the testing set.

Figure 3 shows the mean training and testing R^2 scores of the six folds for all combinations of hyperparameters and circulation variables. R^2_{train} and R^2_{test} are the R^2 scores of the training and testing sets, respectively. To obtain a better fit, R^2_{train} and R^2_{test} need to be as large as possible. To obtain less overfitting, ΔR^2 (i.e., $R^2_{\text{train}} - R^2_{\text{test}}$) needs to be as small as possible. To satisfy the conditions of both better fitting and less overfitting, the optimum results chosen are when $(R^2_{\text{train}} + R^2_{\text{test}})$ reaches a maximum. The criterion $\max(R^2_{\text{train}} + R^2_{\text{test}})$ is used to exclude some outliers that have a decent R^2_{test} result, but a very poor R^2_{train} result. The sensitivity experiments show that the results of the criterion $\max(R^2_{\text{train}} + R^2_{\text{test}})$ with a simple R^2_{train} constraint (e.g., $R^2_{\text{train}} \geq 0.6$) are similar to the criterion $\max(R^2_{\text{train}} + R^2_{\text{test}})$ for the RF and GBRT methods. The optimum selection of the most suitable circulation variables with the corresponding hyperparameters (not listed), which are dependent on the different principal components and machine learning methods, are therefore determined (Table 3).

Table 3 lists the optimal input variables obtained by using this selection for each machine learning method and Table 4 gives the TCC of each principal component calculated for each machine learning method. The SVR method uses 0.15 for PC1, 0.545 for PC2, and -0.11 for PC3; the RF method uses 0.52 for PC1, 0.22 for PC2, and 0.33 for PC3; and the GBRT uses 0.70 for PC1, 0.35

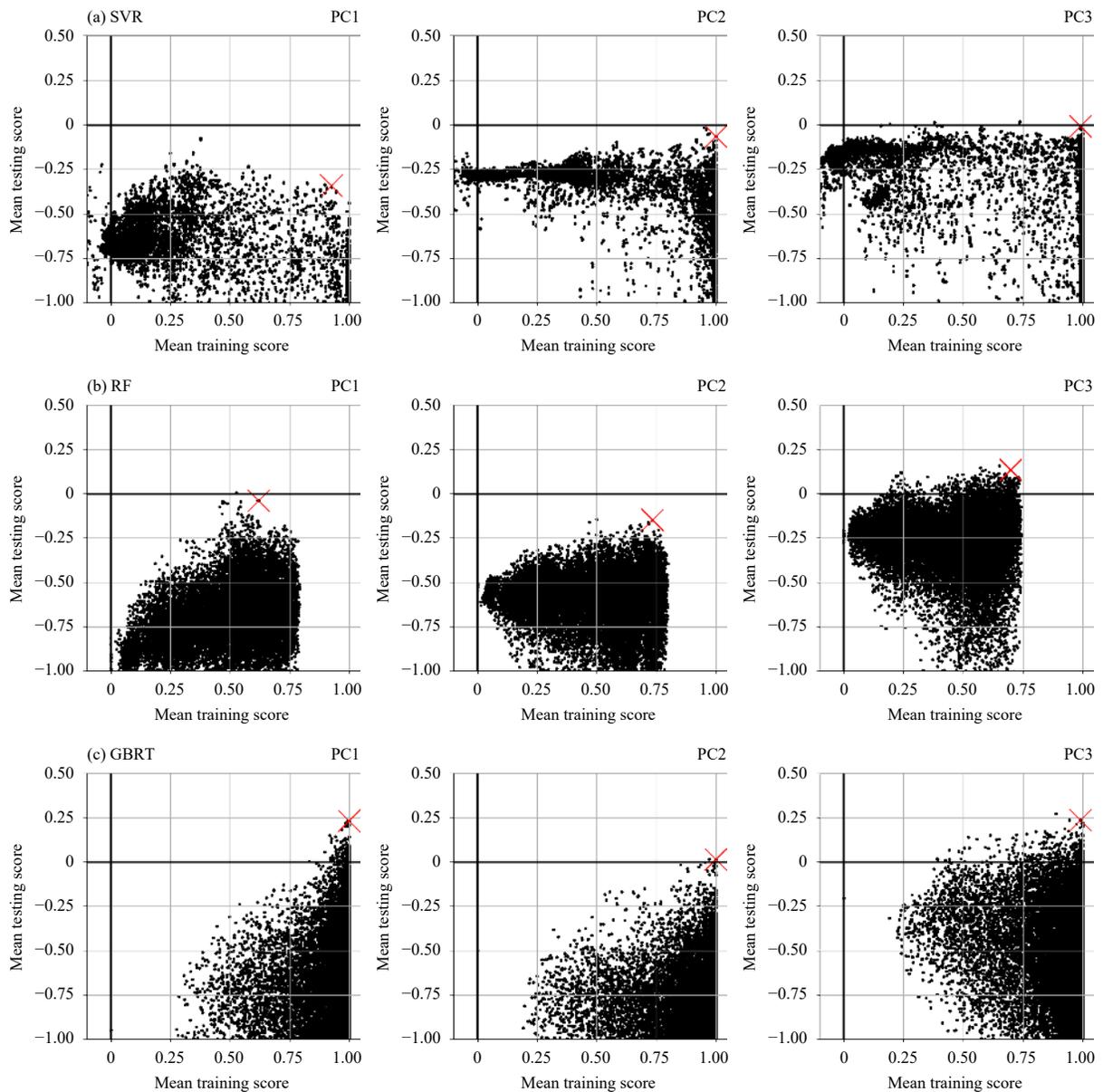


Fig. 3. Mean training and testing scores of the six-fold cross-validation during the grid search with different hyperparameters and different circulation variables (31 variable combinations for 5 selected variables) for the (a) SVR, (b) RF, and (c) GBRT methods. There are 2745 hyperparameter combinations for the SVR method and 25,000 combinations for the RF and GBRT methods for each combination of variables. The red crosses represent the optimum combinations of variables with the corresponding selected hyperparameters.

Table 3. Selected optimum input variables from the FGOALS prediction

Method	PC1	PC2	PC3
SVR	U850, T200	U850	U850, T200, U200
RF	SLP, U850, U200	SLP, U850, Z500, T200, U200	U200
GBRT	SLP, U850, T200, U200	SLP, U850, Z500	U850

for PC2, and 0.47 for PC3. Therefore, compared with the other two machine learning methods, the GBRT method gives the best fit for the historical cross-validation epoch.

3.2 Reconstruction of precipitation anomaly patterns and validation

Using the selected dynamical circulation variables/optimum hyperparameters retrieved from the historical cross-validation period (1981–2010), we first apply the predicted principal components to reconstruct the precipitation anomaly pattern over China in the historical cross-validation period (1981–2010) and validate the prediction skill by PCC. We also use the optimum machine learning models to independently predict individual years from 2011 to 2020. The validations are carried out based

Table 4. The TCCs for the three machine learning methods for each principal component prediction during 1981–2010

Method	PC1	PC2	PC3
SVR	0.15	0.55	-0.12
RF	0.52	0.22	0.33
GBRT	0.70	0.35	0.47

on the following three aspects: a parallel comparison among the three machine learning methods and their ensemble; comparison with the pure dynamical prediction results including FGOALS-f2 (the dynamical model, which the MLD is based on) and NMME; and comparison between the results with and without the least overfitting.

Figure 4a shows that the idealized PCC is 0.56 between the reconstructed rainfall pattern using the first three principal components and the observational total rainfall pattern in the historical cross-validation period. In this instance, “idealized” means that the prediction of the first three principal components is perfect. Although the three leading modes only account for 40.1% of the total variance, the mean PCC of 0.56 is an acceptable result.

Figures 4b–d show the prediction skills from the three MLD methods. During the historical cross-validation period (1981–2010), the average PCC is 0.10 for the SVR, 0.18 for the RF method, and 0.33 for the GBRT method. In the independent prediction years from 2011 to 2020, the average PCC is 0.06 for the SVR, 0.19 for the RF method, and 0.19 for the GBRT method. Figure 4e shows the prediction skill of the ensemble mean of the RF and GBRT methods. The average PCC during the historical cross-validation period (1981–2010) is 0.34 and the average PCC in the independent prediction years from 2011 to 2020 is 0.20. The GBRT method therefore shows the best performance among the three MLD methods in terms of the PCC metric, whereas the ensemble mean of the RF and GBRT methods performs better than any MLD alone.

We further compare the MLD methods with the pure dynamical prediction skill, including both the FGOALS-f2 ensemble mean and the NMME (Figs. 4f, g). The results show that the average prediction skill in FGOALS is nearly -0.02 in the historical cross-validation period and 0.04 in the prediction period, which is much lower than most MLD predictions in both periods. In terms of the independent years of 2011–2020, the skill of the ensemble prediction of the RF and GBRT methods based on the output of the FGOALS-f2 model [FGOALS + ens(RF + GBRT)] is 400% higher than the FGOALS-f2 model direct prediction. The NMME averaged prediction skill is 0.01 for the hindcast years from 1982 to 2010 and 0.16 for the real-time forecast from 2012 to 2020, which is

also considerably lower than the MLD prediction skill. Although the PCC aims to measure the sign of the predicted rainfall anomaly, the RMSE measures the magnitude biases of rainfall prediction. Table 5 shows that the machine learning methods did not give a significant improvement in the RMSE in either the historical cross-validation period or the independent prediction period compared with the ensemble mean of the original FGOALS prediction.

4. Conclusions and discussion

To examine whether reducing the overfitting to the largest extent possible is required for MLD, we compare the results with and without considering overfitting. Two groups of prediction schemes are selected for comparison. Group 1 is selected with the consideration of better fitting and less overfitting simultaneously, from which the combination of variables and hyperparameters are selected for final MLD prediction (Table 3). Group 2 has better fitting, but more serious overfitting than Group 1. In this instance, the extent of fitting is measured by R^2_{train} , where a larger R^2_{train} means a better fit, whereas the extent of overfitting is measured by ΔR^2 (i.e., $R^2_{\text{train}} - R^2_{\text{test}}$), where a larger ΔR^2 denotes more overfitting. The GBRT method is taken as an example (Fig. 5). The training/fitting of each principal component in Group 1 is worse than that in Group 2, whereas the overfitting of each principal component in Group 1 is better than that in Group 2. Group 1 clearly has a higher cross-validation prediction skill (measured by TCC) than Group 2 (i.e., 0.70 vs. 0.08 in PC1, 0.35 vs. 0.20 in PC2, and 0.47 vs. 0.21 in PC3), which suggests that reducing overfitting significantly improves the MLD prediction skill.

Does the dynamical prediction skill influence the MLD prediction results in individual years? To investigate this question, we evaluated the relationship of the prediction skills between the MLD and its corresponding dynamical circulation variable in 21 rolling 9-yr windows. PC3 is taken as an example because it is dependent on a single variable (Table 3). Figure 6 shows that the epochs with better dynamical prediction usually have a

Table 5. Mean RMSE of the PCC for the historical cross-validation (1981–2010) and independent prediction (2011–2020) periods

Method	Mean RMSE of 1981–2010	Mean RMSE of 2011–2020
FGOALS	152.7	154.7
FGOALS + SVR	145.0	156.9
FGOALS + RF	142.8	151.9
FGOALS + GBRT	138.2	153.5
FGOALS + ens(RF + GBRT)	139.1	151.9

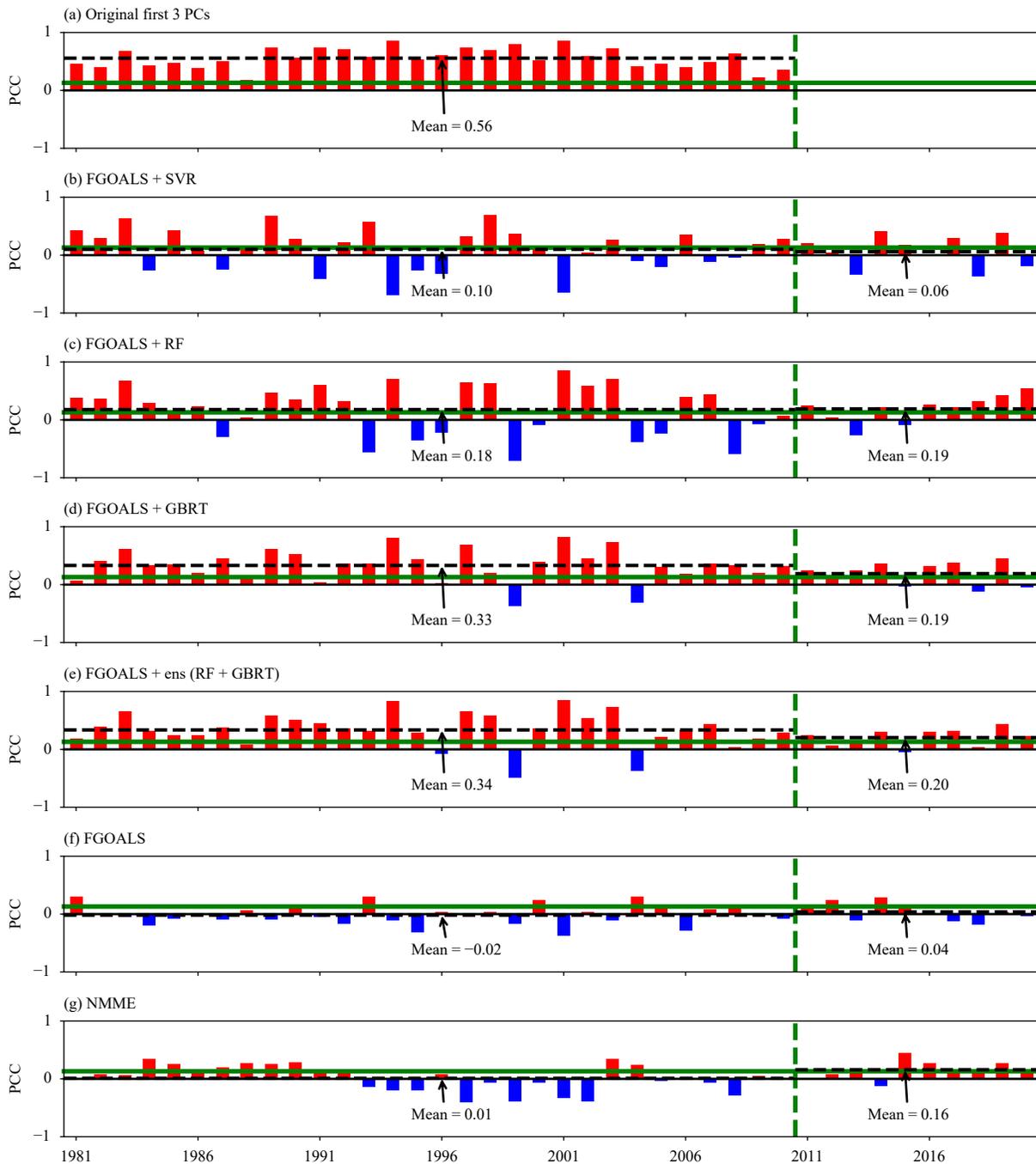


Fig. 4. PCC between (a) the reconstructed precipitation anomaly using the first three observational principal components, (b–d) the reconstructed precipitation anomaly using three types of MLD-predicted principal components, (e) the ensemble of the RF and GBRT methods, (f) the FGOALS ensemble precipitation prediction, and (g) the NMME ensemble precipitation prediction and the observed summer precipitation anomaly in China. The horizontal solid green lines denote the 90% confidence level and the dashed green lines separate the historical cross-validation period (1981–2010) and the independent prediction period (2011–2020).

higher MLD prediction skill, which indicates that the MLD method largely depends on the dynamical prediction performance. Therefore, developing state-of-the-art dynamical prediction is a prerequisite for MLD predictions.

We developed a method for the seasonal prediction of

summer rainfall in China based on machine learning and a dynamical model. Three types of nonparametric machine learning method are examined. We chose the optimum hyperparameters for each machine learning method to reach the best fit and the least overfitting after a grid search. We also chose the dynamical circulation fields

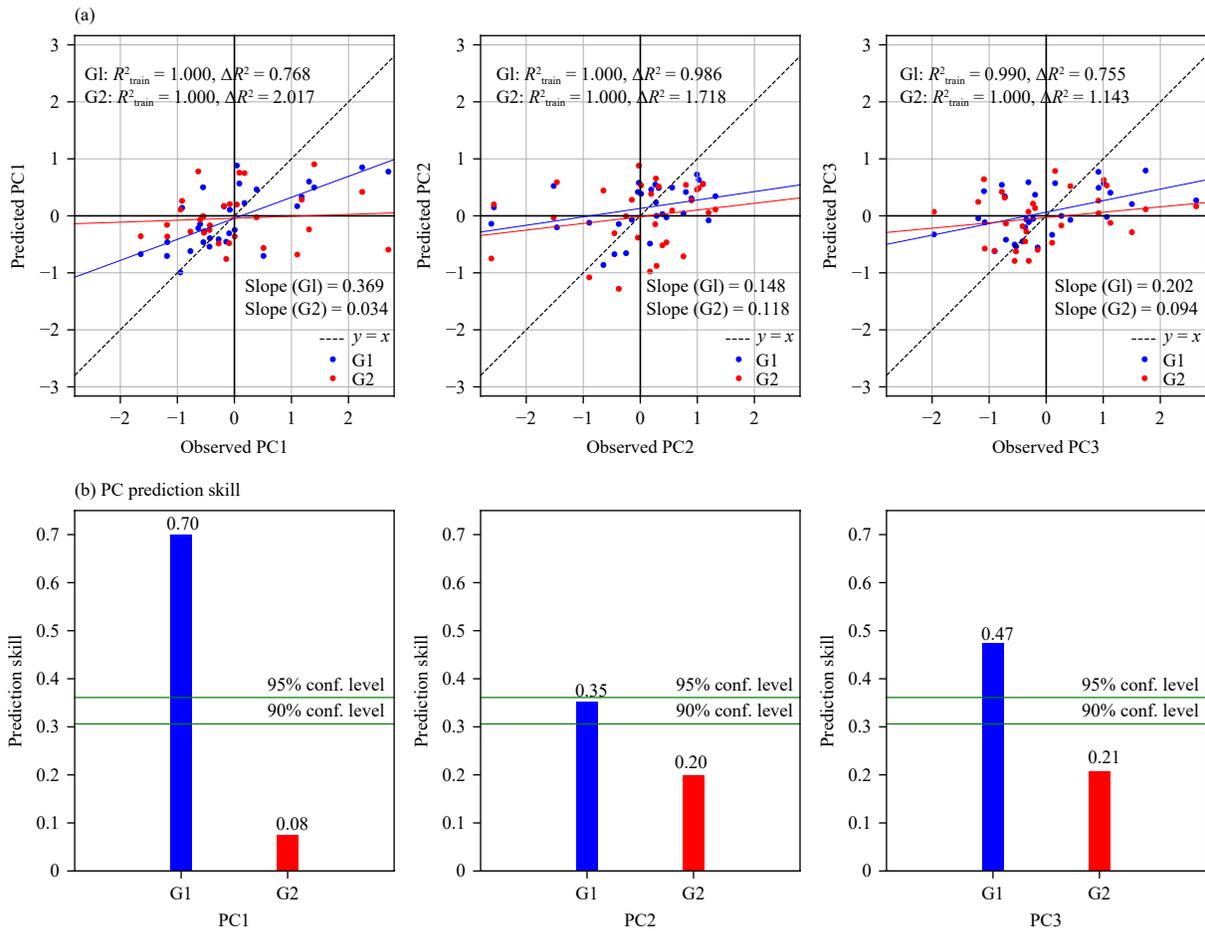


Fig. 5. (a) Scatter diagrams of the FGOALS + GBRT predicted principal components and the observed principal components with the contrasting hyperparameters of Groups 1 (G1) and 2 (G2). (b) Bar charts of the prediction skill (TCC) in the historical cross-validation period for the first three principal components in G1 and G2.

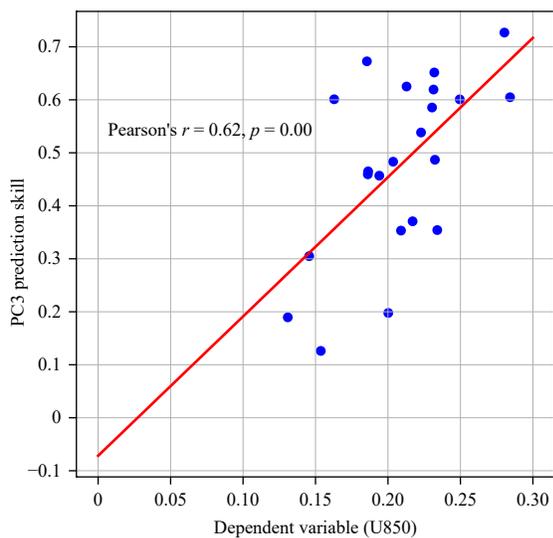


Fig. 6. Scatter diagram representing the relationship between the prediction skill (measured by TCC) of the MLD-predicted principal component (PC3) and the dynamical prediction skill of the dependent variable (U850; measured by the multiyear mean PCC). The rolling window is 9 yr, and the total number of rolling windows is 21.

with a higher prediction skill for MLD. The GBRT method showed the best performance among the three MLD methods. The ensemble means of the random forest and GBRT methods performed better than any MLD alone and showed the best prediction skill for the prediction of summer rainfall in both the historical cross-validation period (1981–2010) and 10-yr independent predictions (2011–2020). This study indicates that MLD could be an efficient method to improve the current dynamical prediction of summer rainfall in China and the combination of several efficient MLD methods could give a better performance in the correction of dynamical prediction. This study also emphasizes that reducing overfitting and using a “best” dynamical model are essential preconditions for MLD.

Several unsolved issues still need to be clarified. The improvement here primarily refers to the sign of the rainfall anomaly and further correction for the biases in rainfall intensity will be investigated in future work. Although physical relationships between some selected

variables and summer rainfall are clearly stated in previous papers (e.g., Z500 used for PC2 prediction; Tao and Xu, 1962; Zhou et al., 2009), the physical linkage between the statistically selected optimum circulation variables and rainfall in China still needs investigation. This study only used three machine learning methods and other potential machine learning methods will be examined in the future.

Acknowledgments. We are grateful for support with the high-performance computing from the Center for Geodata and Analysis, Faculty of Geographical Science, Beijing Normal University (<https://gda.bnu.edu.cn/>). The GPCP precipitation data were provided by the Physical Sciences Division of the Earth System Research Laboratory of the NOAA Office of Oceanic and Atmospheric Research (Boulder, CO, USA; www.esrl.noaa.gov/psd/).

REFERENCES

- Adler, R. F., G. J. Huffman, A. Chang, et al., 2003: The Version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–present). *J. Hydrometeorol.*, **4**, 1147–1167, doi: [10.1175/1525-7541\(2003\)004<1147:TVGPCP>2.0.CO;2](https://doi.org/10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2).
- Alessandri, A., M. D. Felice, F. Catalano, et al., 2018: Grand European and Asian-Pacific multi-model seasonal forecasts: maximization of skill and of potential economical value to end-users. *Climate Dyn.*, **50**, 2719–2738, doi: [10.1007/s00382-017-3766-y](https://doi.org/10.1007/s00382-017-3766-y).
- Badr, H. S., B. F. Zaitchik, and S. D. Guikema, 2014: Application of statistical models to the prediction of seasonal rainfall anomalies over the Sahel. *J. Appl. Meteor. Climatol.*, **53**, 614–636, doi: [10.1175/JAMC-D-13-0181.1](https://doi.org/10.1175/JAMC-D-13-0181.1).
- Bao, Q., X. F. Wu, J. X. Li, et al., 2019: Outlook for El Niño and the Indian Ocean Dipole in autumn–winter 2018–2019. *Chinese Sci. Bull.*, **64**, 73–78, doi: [10.1360/N972018-00913](https://doi.org/10.1360/N972018-00913). (in Chinese)
- Bergstra, J., R. Bardenet, Y. Bengio, et al., 2011: Algorithms for hyper-parameter optimization. Proceedings of the 24th International Conference on Neural Information Processing Systems, ACM, Granada, Spain, 2546–2554.
- Bett, P. E., A. A. Scaife, C. F. Li, et al., 2018: Seasonal forecasts of the summer 2016 Yangtze River basin rainfall. *Adv. Atmos. Sci.*, **35**, 918–926, doi: [10.1007/s00376-018-7210-y](https://doi.org/10.1007/s00376-018-7210-y).
- Breiman, L., 2001: Random forests. *Mach. Learn.*, **45**, 5–32, doi: [10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324).
- Chen, J.-L., and R.-H. Huang, 2008: Interannual and interdecadal variations of moisture transport by Asian summer monsoon and their association with droughts or floods in China. *Chinese J. Geophys.*, **51**, 352–359, doi: [10.3321/j.issn:0001-5733.2008.02.007](https://doi.org/10.3321/j.issn:0001-5733.2008.02.007). (in Chinese)
- Chevuturi, A., A. G. Turner, S. J. Woolnough, et al., 2019: Indian summer monsoon onset forecast skill in the UK Met Office initialized coupled seasonal forecasting system (GloSea5-GC2). *Climate Dyn.*, **52**, 6599–6617, doi: [10.1007/s00382-018-4536-1](https://doi.org/10.1007/s00382-018-4536-1).
- Colin Cameron, A., and F. A. G. Windmeijer, 1997: An *R*-squared measure of goodness of fit for some common nonlinear regression models. *J. Econom.*, **77**, 329–342, doi: [10.1016/S0304-4076\(96\)01818-0](https://doi.org/10.1016/S0304-4076(96)01818-0).
- Dee, D. P., S. M. Uppala, A. J. Simmons, et al., 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, doi: [10.1002/qj.828](https://doi.org/10.1002/qj.828).
- Drucker, H., C. J. C. Burges, L. Kaufman, et al., 1996: Support vector regression machines. Proceedings of the 9th International Conference on Neural Information Processing Systems, MIT Press, Denver, CO, USA, 155–161.
- Fan, K., 2006: Atmospheric circulation in Southern Hemisphere and summer rainfall over Yangtze River valley. *Chinese J. Geophys.*, **49**, 599–606, doi: [10.1002/cjg2.873](https://doi.org/10.1002/cjg2.873).
- Friedman, J. H., 2001: Greedy function approximation: A gradient boosting machine. *Ann. Statist.*, **29**, 1189–1232, doi: [10.1214/aos/1013203451](https://doi.org/10.1214/aos/1013203451).
- Friedman, J. H., 2002: Stochastic gradient boosting. *Computat. Statist. Data Anal.*, **38**, 367–378, doi: [10.1016/S0167-9473\(01\)00065-2](https://doi.org/10.1016/S0167-9473(01)00065-2).
- Gao, M. N., B. Wang, J. Yang, et al., 2018: Are peak summer sultry heat wave days over the Yangtze–Huaihe River basin predictable? *J. Climate*, **31**, 2185–2196, doi: [10.1175/JCLI-D-17-0342.1](https://doi.org/10.1175/JCLI-D-17-0342.1).
- Goddard, L., S. J. Mason, S. E. Zebiak, et al., 2001: Current approaches to seasonal to interannual climate predictions. *Int. J. Climatol.*, **21**, 1111–1152, doi: [10.1002/joc.636](https://doi.org/10.1002/joc.636).
- Gong, D. Y., and C. H. Ho, 2002: Shift in the summer rainfall over the Yangtze River valley in the late 1970s. *Geophys. Res. Lett.*, **29**, 1436, doi: [10.1029/2001GL014523](https://doi.org/10.1029/2001GL014523).
- Ham, Y. G., J. H. Kim, and J. J. Luo, 2019: Deep learning for multi-year ENSO forecasts. *Nature*, **573**, 568–572, doi: [10.1038/s41586-019-1559-7](https://doi.org/10.1038/s41586-019-1559-7).
- He, B., Q. Bao, X. C. Wang, et al., 2019: CAS FGOALS-f3-L Model datasets for CMIP6 historical Atmospheric Model Intercomparison Project simulation. *Adv. Atmos. Sci.*, **36**, 771–778, doi: [10.1007/s00376-019-9027-8](https://doi.org/10.1007/s00376-019-9027-8).
- Jia, X. J., and P. J. Zhu, 2010: Improving the seasonal forecast of summer precipitation in China using a dynamical–statistical approach. *Atmos. Oceanic Sci. Lett.*, **3**, 100–105, doi: [10.1080/16742834.2010.11446849](https://doi.org/10.1080/16742834.2010.11446849).
- Kirtman, B. P., D. Min, J. M. Infanti, et al., 2014: The North American Multimodel Ensemble: Phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. *Bull. Amer. Meteor. Soc.*, **95**, 585–601, doi: [10.1175/BAMS-D-12-00050.1](https://doi.org/10.1175/BAMS-D-12-00050.1).
- Lee, J.-Y., S.-S. Lee, B. Wang, et al., 2013: Seasonal prediction and predictability of the Asian winter temperature variability. *Climate Dyn.*, **41**, 573–587, doi: [10.1007/s00382-012-1588-5](https://doi.org/10.1007/s00382-012-1588-5).
- Lever, J., M. Krzywinski, and N. Altman, 2016: Model selection and overfitting. *Nat. Methods*, **13**, 703–704, doi: [10.1038/nmeth.3968](https://doi.org/10.1038/nmeth.3968).
- Li, Q., F. Y. Wei, and D. L. Li, 2011: Interdecadal variation of East Asian summer monsoon and drought/flood distribution over eastern China in the last 159 years. *J. Geogr. Sci.*, **21**, 579–593, doi: [10.1007/s11442-011-0865-2](https://doi.org/10.1007/s11442-011-0865-2).
- Liaw, A., and M. Wiener, 2002: Classification and regression by random forest. *R News*, **2–3**, 18–22.

- Lim, Y., J. Lee, H.-S. Oh, et al., 2015: Independent component regression for seasonal climate prediction: An efficient way to improve multimodel ensembles. *Theor. Appl. Climatol.*, **119**, 433–441, doi: [10.1007/s00704-014-1099-x](https://doi.org/10.1007/s00704-014-1099-x).
- MacLachlan, C., A. Arribas, K. A. Peterson, et al., 2015: Global seasonal forecast system version 5 (GloSea5): A high-resolution seasonal forecast system. *Quart. J. Roy. Meteor. Soc.*, **141**, 1072–1084, doi: [10.1002/qj.2396](https://doi.org/10.1002/qj.2396).
- Nan, S. L., and J. P. Li, 2003: The relationship between the summer precipitation in the Yangtze River valley and the boreal spring Southern Hemisphere annular mode. *Geophys. Res. Lett.*, **30**, 2266, doi: [10.1029/2003GL018381](https://doi.org/10.1029/2003GL018381).
- Pang, Y. S., C. W. Zhu, and K. Liu, 2014: Analysis of stability of EOF modes in summer rainfall anomalies in China. *Chinese J. Atmos. Sci.*, **38**, 1137–1146, doi: [10.3878/j.issn.1006-9895.1402.13274](https://doi.org/10.3878/j.issn.1006-9895.1402.13274). (in Chinese)
- Pedregosa, F., G. Varoquaux, A. Gramfort, et al., 2011: Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.*, **12**, 2825–2830.
- Picard, R. R., and R. D. Cook, 1984: Cross-validation of regression models. *J. Amer. Statist. Assoc.*, **79**, 575–583, doi: [10.1080/01621459.1984.10478083](https://doi.org/10.1080/01621459.1984.10478083).
- Pour, S. H., A. K. A. Wahab, and S. Shahid, 2020: Physical-empirical models for prediction of seasonal rainfall extremes of Peninsular Malaysia. *Atmos. Res.*, **233**, 104720, doi: [10.1016/j.atmosres.2019.104720](https://doi.org/10.1016/j.atmosres.2019.104720).
- Rana, S., J. Renwick, J. McGregor, et al., 2018: Seasonal prediction of winter precipitation anomalies over Central Southwest Asia: A canonical correlation analysis approach. *J. Climate*, **31**, 727–741, doi: [10.1175/JCLI-D-17-0131.1](https://doi.org/10.1175/JCLI-D-17-0131.1).
- Ren, H. L., Y. J. Wu, Q. Bao, et al., 2019: The China multi-model ensemble prediction system and its application to flood-season prediction in 2018. *J. Meteor. Res.*, **33**, 540–552, doi: [10.1007/s13351-019-8154-6](https://doi.org/10.1007/s13351-019-8154-6).
- Saha, M., P. Mitra, and R. S. Nanjundiah, 2017: Deep learning for predicting the monsoon over the homogeneous regions of India. *J. Earth Syst. Sci.*, **126**, 54, doi: [10.1007/s12040-017-0838-7](https://doi.org/10.1007/s12040-017-0838-7).
- Shen, B. Z., Z. D. Lin, R. Y. Lu, et al., 2011: Circulation anomalies associated with interannual variation of early- and late-summer precipitation in Northeast China. *Sci. China Earth Sci.*, **54**, 1095–1104, doi: [10.1007/s11430-011-4173-6](https://doi.org/10.1007/s11430-011-4173-6).
- Strazzo, S., D. C. Collins, A. Schepen, et al., 2019: Application of a hybrid statistical–dynamical system to seasonal prediction of North American temperature and precipitation. *Mon. Wea. Rev.*, **147**, 607–625, doi: [10.1175/MWR-D-18-0156.1](https://doi.org/10.1175/MWR-D-18-0156.1).
- Tao, S. Y., and S. Y. Xu, 1962: Some aspects of the circulation during the periods of the persistent drought and flood in Yangtze and Hwai-ho valleys in summer. *Acta Meteor. Sinica*, **32**, 1–10, doi: [10.11676/qxb1962.001](https://doi.org/10.11676/qxb1962.001). (in Chinese)
- Tetko, I. V., D. J. Livingstone, and A. I. Luik, 1995: Neural network studies. 1. Comparison of overfitting and overtraining. *J. Chem. Inf. Comput. Sci.*, **35**, 826–833, doi: [10.1021/ci00027a006](https://doi.org/10.1021/ci00027a006).
- Wang, B., J.-Y. Lee, I.-S. Kang, et al., 2009: Advance and prospect of seasonal prediction: Assessment of the APCC/CLIPAS 14-model ensemble retrospective seasonal prediction (1980–2004). *Climate Dyn.*, **33**, 93–117, doi: [10.1007/s00382-008-0460-0](https://doi.org/10.1007/s00382-008-0460-0).
- Wang, S. W., and J. H. Zhu, 2001: A review on seasonal climate prediction. *Adv. Atmos. Sci.*, **18**, 197–208, doi: [10.1007/s00376-001-0013-5](https://doi.org/10.1007/s00376-001-0013-5).
- Xing, W., B. Wang, and S.-Y. Yim, 2016: Long-lead seasonal prediction of China summer rainfall using an EOF–PLS regression-based methodology climate. *J. Climate*, **29**, 1783–1796, doi: [10.1175/JCLI-D-15-0016.1](https://doi.org/10.1175/JCLI-D-15-0016.1).
- Yang, S., Z. Q. Zhang, V. E. Kousky, et al., 2008: Simulations and seasonal prediction of the Asian summer monsoon in the NCEP Climate Forecast System. *J. Climate*, **21**, 3755–3775, doi: [10.1175/2008JCLI1961.1](https://doi.org/10.1175/2008JCLI1961.1).
- Yin, Z. C., and H. J. Wang, 2016: Seasonal prediction of winter haze days in the north central North China Plain. *Atmos. Chem. Phys.*, **16**, 14843–14852, doi: [10.5194/acp-16-14843-2016](https://doi.org/10.5194/acp-16-14843-2016).
- Zeng, Z., W. W. Hsieh, A. Shabbara, et al., 2011: Seasonal prediction of winter extreme precipitation over Canada by support vector regression. *Hydro. Earth Syst. Sci.*, **15**, 65–74, doi: [10.5194/hess-15-65-2011](https://doi.org/10.5194/hess-15-65-2011).
- Zhou, T. J., R. C. Yu, J. Zhang, et al., 2009: Why the western Pacific subtropical high has extended westward since the late 1970s. *J. Climate*, **22**, 2199–2215, doi: [10.1175/2008JCLI2527.1](https://doi.org/10.1175/2008JCLI2527.1).
- Zhu, Z. W., T. Li, and J. H. He, 2014: Out-of-phase relationship between boreal spring and summer decadal rainfall changes in southern China. *J. Climate*, **27**, 1083–1099, doi: [10.1175/JCLI-D-13-00180.1](https://doi.org/10.1175/JCLI-D-13-00180.1).