Ensemble Forecasting for Basin Extreme Precipitation

XUE-ZHEN LI[1], KUN-FEI LU[1], HAO-JIA CHEN[2], ZI-YU CHEN[2]

[1]School of Information Engineering, Guangdong Eco-Engineering Polytechnic, China
[2]College of Mathematics and Informatics, South China Agricultural University, China
E-MAIL: xuezhen_li@aliyun.com, lkf0918@163.com, e18476557853@gmail.com, yutsuch@163.com

Abstract:

Basin extreme precipitation plays a critical role in regulating reservoir operations and water allocation strategies. Machine learning methods provide new opportunities for precipitation prediction. However, due to complex factors, forecasting extreme precipitation in the basin remains a formidable challenge. In this paper, based on the extensive feature engineering, an ensemble learning framework is proposed for forecasting extreme precipitation in basins. In particular, a meta-learned method based on stacked generalization is adopted for ensemble forecasting by constructing a two-layer prediction model and feeding the prediction results of the basic classifiers (layer 0 models) into a layer 1 model. The results of forecast experiments on Dongjiang river basin show that our method is more effective then traditional time series methods, single machine learning models, and even deep learning time series model.

Keywords:

Extreme precipitation forecasting; Ensemble learning; Machine learning

1. Introduction

Effective management of river basin systems [1, 2] necessitates sophisticated hydrological forecasting capabilities. Advanced prediction of precipitation, especially extreme events, plays a critical role in regulating reservoir operations and water allocation strategies to mitigate disastrous consequences like catastrophic flooding [3, 4]. Nevertheless, accurate projection of severe precipitation events persists as a significant scientific challenge due to multifaceted influencing parameters.

Contemporary precipitation prediction methodologies primarily employ synoptic analysis and statistical modeling techniques. Synoptic meteorological approaches, while valuable, exhibit inherent limitations as they depend on human interpretation, potentially leading to inter-forecaster variability and inadequate spatial resolution for differentiating rainfall quantities across localized regions. Statistical prediction frameworks encompass three principal categories: mathematical modeling [5], physical simulation [6], and dynamic statistical analysis [7]. Mathematical statistics enables probabilistic forecasting through historical climate pattern analysis, with its application in short-term climate prediction tracing back to early 1900s. Recent decades have witnessed substantial advancements in these techniques, particularly in dynamic statistical methodologies, driven by enhanced global observation networks and computational model refinement. However, conventional statistical methods demonstrate limited effectiveness in extreme event prediction due to their inherent neglect of precipitation formation mechanisms.

Prior research predominantly emphasized identifying correlative relationships and hierarchical organization of precipitation-influencing variables. The intricate dynamical interplay between these factors has hindered development of physics-based predictive models [8]. Both conventional physical approaches and existing statistical techniques exhibit restricted accuracy, especially with respect to the prediction of extreme events. In recent decades, with continuous accumulation of high-quality observational data, statistical machine learning methods provide new opportunities for precipitation prediction.

Statistical learning architectures are generally divided into two paradigms: generative models and discriminative models [9]. Generative frameworks (e.g., Naive Bayes classifiers, hidden Markov models) model joint probability distributions, whereas discriminative approaches (including neural networks, support vector machines, decision trees, and maximum entropy models) focus on conditional probability relationships. Current applications of neural networks and support vector machines span both station-specific precipitation forecasting [10] and annual rainfall prediction [3], frequently incorporating outputs from atmospheric circulation models. While recent emphasis has shifted toward dynamical models, their limitations remain incompletely understood. Discriminative models offer distinct advantages in scenarios involving unclear extreme precipitation mechanisms, demonstrating reduced sensitivity to latent characteristic correlations and enhanced capacity for utilizing comprehensive feature sets compared to generative counterparts.

Considering that combining different types of models can be expected to make the final decision more reliable, this paper proposes and implements ensemble forecasting based on multiple single models to integrate the advantages of them. The proposed model with extensive feature engineering aims to predict 24-hour extreme precipitation events in the Dongjiang river basin. The developed forecasting framework enables subsequent hydrological computations for optimized reservoir management and water resource allocation.

This paper is organized as follows: Section 2 details the proposed ensemble forecasting framework. Section 3 provides experimental validation through comparative analysis of the proposed model against traditional time series methods, single machine learning models and deep learning time series model, and the conclusion is presented in Section 4.

2 Methodology

The overall framework of the proposed ensemble forecasting for basin extreme precipitation is shown in Figure 1, which includes four levels: data level, feature level, model level, and decision level.

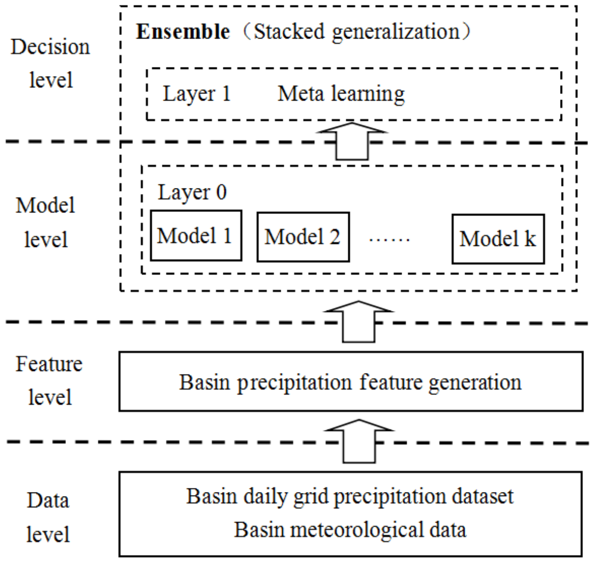

FIGURE 1. The overall structure of our method

2.1 Data level

The raw data selected has certain explanatory significance for extreme precipitation in the Dongjiang river basin, including daily grid data of precipitation, temperature and pressure data, and typhoon data. More details about the datasets in our study is shown in Subsection 3.1.

2.2 Feature generation

The features used for machine learning modeling are generated from the datasets. We use function-based feature generation methods, such as aggregate, synthesis, statistical, boolean operators, and concept hierarchies by data reduction: (1) Aggregate functions typically include the mean, total, maximum, minimum, count, first value, last value, median, and standard deviation. (2) Synthesis functions are functions that use original features or aggregate features as input to calculate a new meaningful feature. The basic synthetic functions, such as add, subtract, multiply, and divide, were used in our framework. (3) Four statistical features, i.e., level number, range, fairness index, and volatility index, are applied from the serialization precipitation data. (4) In addition, concept hierarchies by data reduction were used, such as wet/dry periods, pre flood season, and later flood season.

2.3 Machine learning models

Three machine learning models are used in the model level of our framework, including one generative model and two discriminative models.

2.3.1 Naïve Bayesian model

Naïve Bayesian algorithm (NB) [9] is based on the Bayesian theorem, which is a generative model that focuses on the total probability of all variables. Although it is not a complex model, it can achieve better prediction performance than traditional time series methods and similar to discriminative models under the condition of a rich feature set. Naïve Bayes model does not have inherent difficulties such as overfitting in neural network algorithms, human intervention in support vector machine applications, and difficulty in adjusting parameters.

2.3.2 Neural network model

Neural network (NN) model [11] is a discriminative model based on observations of variables and conditional probability of a target variable under certain premises. Neural networks, inspired by biological cognitive mechanisms, have demonstrated significant efficacy in predictive analytics. For this research, we adopted a backpropagation (BP) neural network, which is a hierarchically structured feedforward system utilizing error backpropagation methodology that has gained predominant application in industrial implementations. Distinctively, BP networks employ differentiable S-shaped activation functions (typically sigmoid-type transfer units) to establish nonlinear correlations between input and output variables. For example, Tansig/Purelin network [11], a two-layer BP network possesses universal function approximation capabilities, theoretically able to model any measurable function with limited discontinuities to arbitrary precision when provided with adequate hidden layer neurons.

2.3.3 Support vector machine model

Support vector machines (SVMs) [12] are alternative discriminative models that possess a high level of generalization and can therefore be used for precipitation forecasting [13]. SVMs are statistical learning methods for structural risk minimization discriminant models, which have better expected generalization ability compared to neural networks. As a discriminative model, support vector machines are also not troubled by potential correlations between features. The key to achieving good predictive performance for support vector machines lies in how to adjust key parameters such as the selection of kernel function and the determination of C-value.

2.4 Ensemble by stacked generalization

The meta learner method of stacked generalization [14] is adopted for ensemble forecasting. By constructing a two-layer prediction model and inputting the prediction results of the basic classifiers (layer 0 models) into a layer 1 model, multiple layer 0 models are fused at the decision level through the layer 1 model to achieve combined forecasting. When outputting from a single model at layer 0 to a layer 1 model, using probability output to carry the prediction confidence of each single model for each prediction instance will be beneficial for the integration of the layer 1 model with the layer 0 models. As this paper focuses on extreme precipitation forecasting, the output of each layer 0 model will be two probabilities, namely the probability of extreme precipitation and the probability of no extreme precipitation.

2.4.1 Probability output of layer 0 models

Naïve Bayes, neural networks, and support vector machine models are used in the layer 0 models in the paper, and the probability output by these models can vary.

- Naïve Bayes model: For each predicted instance x, the output probability of each category $P_i(x)$ in the Naive Bayes model is obtained by the ratio of the posterior probability of each category to the total probability of all categories:

$$P_i(x) = P(i|x) / \sum_{j=1}^{2} P(j|x) \tag{1}$$

- BP neural network model: The output of BP neural network does not have probability. For this purpose, for each predicted instance x, we first define a model output confidence $C_i(x)$, where i is the corresponding prediction category, and calculate the output probability estimate $P_i(x)$ based on the confidence:

$$P_i(x) = C_i(x) / \sum_{j=1}^{2} C_j(x) \tag{2}$$

When training the neural network model, we used small precipitation levels, 0-50 levels, with one level every 5mm, except for [0,0.1mm) representing no rain. Therefore, for each predicted instance x, the neural network outputs a predicted small category, which we then convert into the corresponding large category, that is, with or without extreme precipitation. The level difference between the model output level and the actual precipitation level is used as the model output confidence measure, and mapped to the range of 0.5-1. A level difference of 0 corresponds to a confidence level of 1, and the maximum level difference corresponds to a confidence level of 0.5. Based on this, the output confidence of instance x regarding category i is obtained.

Due to the fact that the actual precipitation level of instance x is still unknown during prediction, for each predicted data in the dataset, we first find the most similar five instance data from the previous five years. The similarity is obtained based on the feature set and using cosine as an example. Then, using similarity as a weight, the weighted output confidence of the 5 instance data is used as the output confidence of the predicted category of the data to be predicted.

- Support Vector Machine Model: The output of support vector machine also does not carry probability. Therefore, for each predicted instance x, we need to first define a model output confidence $C_i(x)$, where i is the corresponding prediction category, and calculate the output probability estimate $P_i(x)$ based on the confidence. The calculation formula is the same as equation (2).

  In support vector machine models, the farther the data is from the hyperplane in high-dimensional space, the lower the likelihood of misclassification. Therefore, the confidence measure of the model is directly taken as the distance between each predicted instance x in the sixth year and the hyperplane of the support vector machine model trained on the data from the previous five years, and mapped to the range of 0.5-1. The maximum distance corresponds to a confidence level of 1, and the distance of 0 corresponds to a confidence level of 0.5, in order to obtain the output confidence of $C_i(x)$.

2.4.2 Layer 1 modeling

As an arbitrator in the selection of layer 1 model, some researchers have suggested that choosing a simple model that can integrate the outputs of layer 0 models is more appropriate [14]. Considering that most of the early research was based on relatively similar layer 0 models, such as linear models, while the three layer 0 models used in this paper include linear and nonlinear models, we adopted a nonlinear layer 1 model for the synthesis of layer 0 models of different classes. Specifically, we take the output probabilities of three layer 0 models for different prediction categories as the input of a layer 1 SVM. The data from the first five years are used to establish the model using cross validation.

3 Experiments

3.1 Datasets

The data sets used to forecast the basin extreme precipitation include the basin precipitation time series, the meteorological data, and the typhoon data.

- Basin Daily Grid Precipitation Dataset. This study utilized daily gridded precipitation records (2008-2013) covering the Dongjiang river basin, acquired from the National Climatic Centre (NCC) of the China Meteorological Administration (CMA). The locations of the Dongjiang river basin and the corresponding geographical grid considered in our study are shown in Figure 2. The dataset features $0.25°\times0.25°$ spatial resolution, and in this paper, mean area precipitation is used to analyze the daily precipitation of the river basin.
- Basin Meteorological Data. We obtained the basin area and associated meteorological data, including surrounding temperature and pressure from the daily surface climate dataset of NCC to generate a feature set for forecasting purposes.
- Typhoon Data. We also obtained daily typhoon forecast data including the path, speed and 10-gale ring area from the Flood Release System of Guangdong Province to generate feature for forecasting purposes.

3.2 Baselines

We compare our model ESG (Ensemble by stacked generalization) with the following methods:

- Simple moving average method (SMA) [15]: A traditional time series method in which the mean precipitation value of the evidence window serves as the prediction for the future precipitation.
- Auto-regression method (AR) [15]: Another commonly used model in time series modeling and analysis. Use the same variable, such as the precipitation of previous periods, to predict the performance of the future precipitation, and assume that they have a linear relationship.

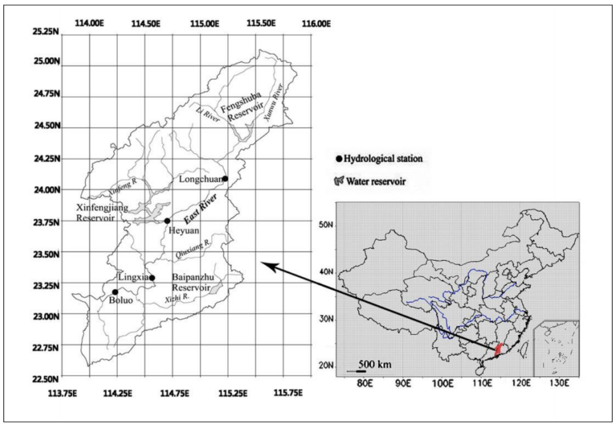

FIGURE 2. Location of the study region and geographical grid

- Naïve Bayesian algorithm (NB) [9]: A generative model based on the Bayesian theorem, which is a generative model that focuses on the total probability of all variables.

- Neural network model (NN) [11]: a discriminative model inspired by biological cognitive mechanisms, have demonstrated significant efficacy in predictive analytics.

- Support vector machines (SVMs) [12, 13]: Another discriminative model, which is one of the most popular machine learning methods that belongs to the structural risk minimization.

- Autoformer [16]: A deep learning model for processing time series prediction tasks using precipitation series data, which innovatively combines the Decomposition Architecture and Auto Correlation Mechanism.

3.3 Metrics

Model performance are measured using the Metrics of F-Measure and Bias Score.

3.3.1 F-Measure

Precision (P), Recall (R) and F1-score are widely used to evaluate the quality of forecasting results

Precision denotes the proportion of predicted positive cases (extreme precipitation) that are correct real positives:

$$P = TP/(TP + FP) \tag{3}$$

where TP is the True Positive and FP is the False Positive

Conversely, Recall is defined as the proportion of real positive cases that are correctly predicted as positive:

$$R = TP/(TP + FN) \tag{4}$$

where FN is the False Negative.

P and R represent the false alarm rate and missing report rate, respectively, in extreme precipitation forecasting. The larger the P value, the lower the false alarm rate, while the larger the R value, the lower the missing report rate.

The F1-score is defined as a harmonic mean of the precision and recall:

$$F1 - score = 2PR/(P + R) \tag{5}$$

3.3.2 Bias Score

Bias Score (B-value) is a commonly used scoring method in precipitation forecasting system, that is:

$$B = (TP + FN)/(TP + FN) \tag{6}$$

where a B-value close to 1 represents a small forecast deviation

3.4 Results and Analysis

The performance comparison of different models on F-Measure are shown in Table 1. As we can see from Table 1, our ESG model obtains the best results of F1-score, while NB obtains the best result of precision, and SVM obtains the best result of recall. By applying ensemble framework, ESG model combines the complementary advantages of individual machine learning models in precision and recall, achieving relatively good precision and recall simultaneously, thus obtaining the best F1-score. The relative improvement of F1-score of our ESG compared to Autoformer, SVM, NN, NB, AR and SMA are 42.1%, 5.9%, 10.2%, 25.6%, 157.1% and 260%, respectively. We also see that for the time series models using only precipitation series data, Autoformer significantly outperforms the traditional time series models.

We then observe the comparison of forecast deviation B-value of the single machine learning models and the ensemble ESG. From Figure 3, it can be seen that the ESG model achieved the best prediction bias, with a B-value of only 1.30, while the B-value of the SVM model is as high as 1.91. An excessively high B-value also represents too

TABLE 1. F-Measure performance comparison of different models

Models	Metrics		
	Precision(P)	Recall (R)	F1- score
SMA	0.24	0.11	0.15
AR	0.35	0.15	0.21
NB	0.65	0.32	0.43
NN	0.57	0.43	0.49
SVM	0.39	0.74	0.51
Autoformer	0.62	0.28	0.38
ESG (ours)	0.48	0.62	0.54

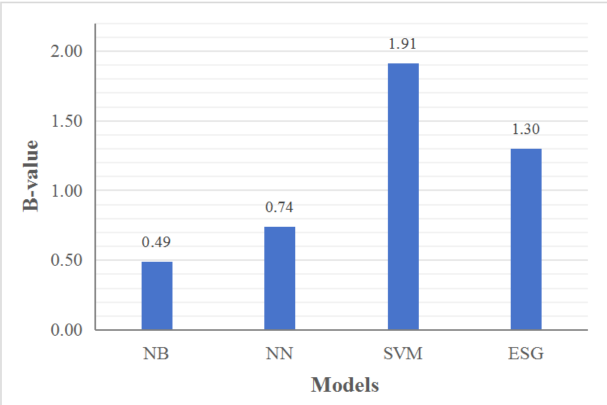

FIGURE 3. B-value comparison of different models

many false forecast. Both NN and NB models tend to be conservative, resulting in forecast biases below 1. In particular, the B-value of the NB is only 0.49, which means that the total number of extreme precipitation forecasts (including accurate and inaccurate forecasts) is only half of the actual number of extreme precipitation forecasts, leading to a low recall rate of only 0.32 and serious underreporting.

4 Conclusion

In this paper, an ensemble learning framework using stacked generalization is proposed for basin extreme precipitation forecasting. Experiments on Dongjiang river basin show that our approach with extensive feature engineering is more effective compared with traditional time series methods, single machine learning models and even deep learning time series model. Although only precipitation series data was used, the deep learning time series model showed competitive performance. Therefore, future work will mainly focus on how to integrate the deep learning time series model into the ensemble framework.

Acknowledgements

We acknowledge the National Climatic Centre (NCC) of the China Meteorological Administration (CMA) for providing daily grid precipitation dataset for analysis. This study is supported by the Educational Commission of Guangdong Province of China (No. 2018GKTSCX061)

References

[1] H. M. Wang and J. P. Tong, "An Adaptive Method for Water Resource Allocation and Its Application", Science Press, Beijing, China, 2011.

[2] K. Baran-Gurgul, A. Rutkowska, "Water Resource Management: Hydrological Modelling, Hydrological Cycles, and Hydrological Prediction", Water, 2024, Vol. 16, No. 24, p3689

[3] A. Kalra and S. Ahmad, "Estimating Annual Precipitation for the Colorado River Basin Using Oceanic-atmospheric Oscillations", Water Resources Research, 2012, vol.48, W06527, pp.1-24.

[4] H. D. Lyu, H. F. Xing, T. X. Duan, "Optimizing Water Resource Allocation for Food Security: An Evaluation of China's Water Rights Trading Policy", Sustainability, 2024, Vol.16, No. 23, p10443

[5] Z. P. Yu, P. Chu and T. Schroeder, "Predictive Skills of Seasonal to Annual Rainfall Variations in the U.S. Affiliated Pacific Islands: Canonical Correlation Analysis and Multivariate Principal Component Regression Approaches", Climate, 1997, no.12, pp. 2586-2599.

[6] E. J. Becker, H. Van Den Dool and M. Peña, "Short-term Climate Extremes: Prediction Skill and Predictability", Journal of Climate, 2013, vol.26, no.2, pp.512-531.

[7] J. Yang, "Predictive Reliability of Summer Precipitation in China Based on Error Distribution of Numerical Model", Acta Phys. Sin., 2014, vol.63, no.14, 149202, pp.1-12.

[8] G. F. Lin, G.R. Chen, M.C. Wu, Y.C. Chou, "Effective Forecasting of Hourly Typhoon Rainfall Using Support Vector Machines", Water Resources Research, 2009, Vol.45, W08440, pp.1-11.

[9] H. Li, “Machine learning methods”, Beijing: Tsinghua University Press, 2022.

[10] Q. F. Xiong and X.Q. Zeng, “Applicat ion and Improvement of SVM Method in Precipitation Forecast”, Meteorological Monthly, 2008, vol.34, no.12, pp.90-95.

[11] S. Haykin, “Neural networks and learning machine”, 3rd edn, Prentice Hall, 2008.

[12] V. Vapnik, “Statistical Learning Theory”,Wiley, New York, 1998.

[13] J. C. Lei, Q. Quan, P. Z. Li, et al., “Research on Monthly Precipitation Prediction Based on the Least Square Support Vector Machine with Multi-Factor Integration”, Atmosphere, 2021, Vol. 12, No. 8, p1076.

[14] K. M. Ting, I. H. Witten, “Stacked Generalization: when does it work”, In Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence (IJCAI 1997), pp. 866–871.

[15] J. D. Hamilton, “Time Series Analysis”, Princeton University Press, 1994.

[16] H. X. Wu, J. H. Xu, J. M. Wang, et al. “Autoformer: Decomposition Transformers withAuto-Correlation for Long-Term Series Forecasting”, In NeurIPS 2021: Advances in Neural Information Processing Systems. New York: Curran Associates, 2021: 22419-22430.