Y. Li[1,2], S. Y. Huo[1,2], B. Ma[3], B.B.Pei[1,2], Q.K.Tan[1,2], D.Wang[2,4]

[1] Hubei Key Laboratory of Petroleum Geochemistry and Environment, Yangtze University, Wuhan 430100, China.

[2] College of Resources and Environment, Yangtze University, Wuhan 430100, China.

[3] School of Environmental Studies, China University of Geosciences, Wuhan 430074, China.

[4] Six Geological team of Hubei Geological Bureau, Xiaogan 432000, China.

Corresponding author: Siyuan Huo (syhuo@yangtzeu.edu.cn)

Key Points:

- Deep learning can predict precipitation isotope distributions over multi-site and multi-year.
- CNN-Bi-LSTM has better prediction results than CNN and Bi-LSTM.
- Clustering first and then prediction can improve the prediction accuracy to some extent.

Abstract

The combined utilization of spatiotemporal clustering and deep learning neural network models were designed to evaluate the applicability of the multi-year and multi-sites precipitation ^{18}O forecasting method based on the precipitation isotope data of 10 stations in Germany from 1988 to 2012. In the overall forecasting, the performance of single-site multi-year forecasting is in the order of the Bi-directional Long Short-Term Memory (Bi-LSTM), CNN-Bi-LSTM, and the Convolutional Neural Network (CNN), with CNN-Bi-LSTM being the optimal model for multi-site multi-year forecasts. The seasonal forecasting does not demonstrate a significant improvement compared to the overall forecasting. For forecasting based on spatiotemporal clustering, cluster 1 improved accuracy, and cluster 2 improved error reduction and variance consistency. Nevertheless, the accuracy of forecasts depends solely on the amount of input data when the proportion of forecasting increases to a certain level. Overall, the seasonal forecasting is more appropriate for improving forecasting within a specific season, while spatiotemporal clustering can improve forecasting accuracy to some degree. In addition, optimal solutions exist for the type and number of model clusters. In terms of model types, CNN-Bi-LSTM generally has better forecasting performance than CNN and Bi-LSTM.

Plain Language Summary

The joint prediction model of spatio-temporal clustering and deep learning provides a new approach for precipitation isotope prediction.

1 Introduction

The stable hydrogen and oxygen isotopes in water molecules are ideal tracers because their behavior and changes reflect the origin of natural water bodies and the hydrological and geochemical processes they undergo [*Gower*, 1967]. As a result of isotopic fractionation, rainfall exhibits distinct spatial and temporal distributions of ^{18}O. There are two reasons for the different spatial and temporal distributions of precipitation ^{18}O: First, the regional environmental background, that is, the source and nature of the precipitation air mass and the entire process from the generation to the transport to the occurrence of the precipitation event. The second is the local geographical factors, including various meteorological elements during precipitation (precipitation, temperature, humidity, etc.) and the local latitude, altitude, etc. [*Guo et al.*, 2019]. These two factors' interaction and combined influence result in the spatial and temporal distributions of precipitation ^{18}O.

The spatial and temporal distributions of precipitation isotopes can reveal changes in hydrometeorological processes at local and larger scales [*Craig*, 1961; *Klaus and McDonnell*, 2013; *J. Klaus et al.*, 2015]. Currently, precipitation isotope data have been used for estimating regional recharge processes [*Koeniger et al.*, 2016], infiltration and mixing processes [*Christine Stumpp and Maloszewski*, 2010; *Zhao et al.*, 2013], plant water uptake processes [*Gaines et al.*, 2016; *Koeniger et al.*, 2010], evaporation and soil water transfers [*Gonfiantini et al.*, 2018], and the mutual transformation of surface water and atmospheric water processes [*Y Q Li et al.*, 2021; *Yang et al.*, 2018]. The above studies require the knowledge of the isotopic composition of modern atmospheric precipitation at one or more locations. Globally, however, only a few hundred sites have been able to determine stable isotopic compositions of precipitation over an extended period. For some specific applications, direct measurements of local precipitation over one or several years can be used to estimate their average isotopic composition. However, direct measurements are not desirable for studies conducted at the regional or global level. Overall, these issues demonstrate the need for methods to estimate ^{18}O using available data.

Based on the information from measured stations, three types of methods are used to study the spatial and temporal distributions of atmospheric precipitation ^{18}O: one is a statistical method, including autoregressive integrated moving average (ARIMA) [*Julian Klaus et al.*, 2015] and multiple linear regression [*Zhu and Burzykowski*, 2013]. The model simulates the spatial or temporal patterns of precipitation ^{18}O through establishing a relationship between topography, location, and meteorological factors; the second method is numerical, the most common of which is spatial interpolation. The model uses a variety of mathematical modeling methods to fit the regional precipitation ^{18}O composition, including contour method, trend surface method, Tyson polygon method, inverse distance weighting method [*X Liu et al.*, 2015], kriging interpolation method [*J Li et al.*, 2014] and sample method, etc.; the third method is the integrated method, which combines a statistical model with spatial interpolation, such as a BW regression model [*Bowen and Wilkinson*, 2002; *Bowen and Revenaugh*, 2003]. However, these methods do not adequately account for the effects of

precipitation ^{18}O 's temporal variation on its spatial distribution. As of right now, there is no model that can predict the spatial and temporal distribution of precipitation ^{18}O simultaneously.

Deep learning is a powerful machine learning method that has been widely utilized in various fields related to hydrology. For example, in comparison with the traditional Support Vector Regression (SVR) model, the CNN-based groundwater potential map forecasting model performs better [*Panahi et al.*, 2020]; The Long-Short-Term Memory model (LSTM) has greater potential for predicting rainfall runoff than the traditional Sacramento Soil Moisture Accounting Model (SAC-SMA) [*Kratzert et al.*, 2018; *Xiang et al.*, 2020]; In addition, the Bi-LSTM based forecasting model is more accurate than LSTM and gated recursive unit (GRU) models in predicting daily precipitation data [*Latifoglu*, 2022]. Despite the promise of deep learning to improve the predictive power of models, individual networks cannot address more complex problems with the same accuracy. Many researchers have investigated the combination of multiple deep learning networks [*Belayneh et al.*, 2014; *H Liu et al.*, 2013; *Mohammadi and Mehdizadeh*, 2020]. The widely popular CNN-LSTM combines the advantages of CNN and LSTM. The CNN can extract features from grid data effectively, and the LSTM can take advantage of time-series data to extract useful features. For example, a CNN-LSTM model for predicting multi-hour, multi-location air quality indices in Beijing utilizes an LSTM to analyze the features extracted by the CNN [*Yan et al.*, 2021]. In comparison with CNN and LSTM, the CNN-LSTM hybrid model significantly reduced forecasting error for short-term water quality variables [*Barzegar et al.*, 2020]. Moreover, the joint forecasting of daily precipitation data based on instantaneous frequency features and Bi-LSTM has a higher forecasting accuracy than single LSTM, gated recursive unit (GRU) and Bi-LSTM models [*Latifoglu*, 2022]. There are, however, no data regarding the spatial and temporal distribution characteristics of hydrological processes prior to their establishment for these combined models.

Cluster analysis is a common method used for analyzing internal patterns within an unordered dataset by dividing the data into a number of similar categories [*Gower*, 1967]. In this manner, training samples for a model can be provided that are highly similar, which reduces training time and increases generalization ability [*Cui et al.*, 2019]. A study of the Tibetan Plateau by fuzzy clustering combined with Regionalized Cluster-based Water Isotope Prediction (RCWIP) and Online Isotopes in Precipitation Calculator (OIPC) shows that the forecasting accuracy varies by region [*Shi et al.*, 2020]. Nevertheless, a change in the timing of precipitation isotopes can also affect the precipitation isotope forecasting. Research on precipitation isotope forecasting based on spatial and temporal clustering is limited.

The objective of this study is to develop multi-year multi-site deep learning models (CNN, Bi-LSTM, CNN-Bi-LSTM) based on spatiotemporal clustering for month-by-month precipitation ^{18}O forecasting. This entails (1) characterizing the spatiotemporal distribution of precipitation ^{18}O; (2) developing and

comparing the overall forecasting based on CNN, Bi-LSTM, and CNN-Bi-LSTM models; (3) combining CNN, Bi-LSTM, and CNN-Bi-LSTM models with spatiotemporal clustering and comparing them with the overall forecasting; and (4) applying the proposed models to month-by-month precipitation ^{18}O forecasting. Based on the precipitation isotope data in Germany, this study combines spatiotemporal clustering analysis with deep learning to examine the spatiotemporal distribution characteristics of precipitation ^{18}O and their impact on forecasting.

2 Materials and Methods

2.1 Overall framework

The objective of this study is to develop multi-year multi-site deep learning models (CNN, Bi-LSTM, CNN-Bi-LSTM) based on spatiotemporal clustering for month-by-month precipitation ^{18}O forecasting. Following the spatiotemporal analysis of German precipitation ^{18}O, all data are clustered by either spatial or temporal dimension. The forecasting model for precipitation ^{18}O based on total data and data from different spatiotemporal clusters were developed. Furthermore, different models were evaluated and compared regarding their forecasting performance.

2.2 Spatiotemporal analysis and clustering

On the temporal scale, the mean and standard deviation of ^{18}O were calculated at various time scales, and the temporal variation characteristics of ^{18}O in Germany were examined on a seasonal and monthly scale. On the integrated spatial and temporal scales, K-means ++ clustering was employed to analyze the spatial and temporal distribution characteristics of ^{18}O in precipitation at multiple sites in Germany on a month-by-month basis.

The K-means ++ clustering algorithm consists of five steps [*Takashi Onoda*, 2010]: (1) randomly select a sample point c_1 1 from the sample set $\overrightarrow{X}$ as the 1st cluster center; (2) calculate the distance $d(x)$ from the other sample points x to the nearest cluster center; (3) with probability $\frac{d(x)^2}{\sum \in x d(x)^2}$ to select a new sample point c_i to be added to the set of cluster centroids, where the larger the distance value $d(x)$, the higher the probability of being selected; (4) repeat steps (2) and (3) to select k cluster centers; (5) perform k-means operations based on these k cluster centers.

Based on the spatial and temporal distribution characteristics of precipitation ^{18}O in Germany, the dataset was divided into different clusters. On the spatial scale, four clusters were obtained according to the seasons. On the integrated spatial and temporal scales, the 10 gauging stations were divided into multiple classes by K-means ++ clustering. Since the size of the convolution kernel is mostly 3×3 or above, the number of stations in each cluster should be more than 3 to complete the convolution process. In this regard, CNN-based, Bi-LSTM-based, and CNN-Bi-LSTM-based forecasting models were developed using seasonal and spatiotemporal data.

2.3 CNN

CNN is a neural network: an algorithm for recognizing data patterns, with a main structure consisting of an input layer, a convolutional layer, a pooling layer, a fully connected layer, and an output layer[*Alzubaidi et al.*, 2021; *Kattenborn et al.*, 2021; *Wu et al.*, 2020]. By using convolutional kernels on the convolution layer and down sampling on the pooling layer, the potential features of the data are extracted and mapped to the output signal (Fig. 1a).

The convolutional layer is the core building block of a CNN, consisting of learnable kernels or filters that extend to the entire depth of the input. In this layer, each cell receives a small neighborhood from the previous layer of input, and during signal forward pass, each filter is convolved with the input of the generated graph. When multiple filters from these feature maps are combined, a convolution layer is formed. In addition, the generated weight vector feature maps are shared, which greatly reduces the complexity of the model. The formula for calculating each element in the feature map is as follows:

$$x_{i,j}^{\text{out}} = f_{\text{cov}}\left(\sum_{m=0}^{k}\sum_{n=0}^{k} w_{m,n}x_{i+m,j+n}^{\text{in}} + b\right) \quad (2)$$

Where $x_{i,j}^{\text{out}}$ is the output value in row i and column j of the feature map; $x_{i+m,j+m}^{\text{in}}$ is the value in row i and column j of the input matrix; $f_{\text{cov}}(\cdot)$ is the selected activation function; $w_{m,n}$ is the weight in row m and column n for the convolution kernel; and b is the bias of the convolution kernel. In general, multiple kernels are used for convolution operations on the input matrix.

On this basis, by downsampling the previous feature mapping, the pooling layer reduces its length and width, and thus improves its accuracy. Due to CNN's good feature extraction capabilities, m variables (precipitation ^{18}O, precipitation, temperature, longitude, latitude, and elevation) of each type for n stations were expanded to obtain a matrix of m rows and n columns. Multiple channels were used to input data for c types (precipitation ^{18}O, meteorological elements, spatial and temporal properties). By transforming each time slice in the input, the time distribution layer can provide more information about the historical time step of the data and its long-term characteristics. Thus, the time distribution layer was chosen to wrap the data for the past t months. In this case, the input to the CNN is a four-dimensional array of dimensions t × m × n × c (12 × 6 × 10 × 3). Following convolution and pooling, a one-dimensional array was constructed from the extracted features. Lastly, a full output concatenation layer generated the precipitation ^{18}O forecasting for t stations for the next t months.

2.4 Bi-LSTM

The Recurrent Neural Network (RNN) is a class of neural networks used to process sequential data. In essence, it is the weight sharing between neural networks in temporal order (CNN is equivalent to weight sharing in a spatial order) [*LeCun et al.*, 2015]. The RNN can learn the nonlinear features of sequences

with high efficiency since it has characteristics of memorability and parameter sharing. However, the traditional RNN implicit layer has a simple internal structure, and the network does not filter the input of the current moment from the input of the previous moment, preventing key data information from being transmitted, causing large bias in forecasting results [*Hochreiter*, 2011].

As a variant of RNN, LSTM adds input gates, output gates, and forgetting gates to the implicit layer, as well as memory storage units. Fig. 1b illustrates the internal structure of the implicit layer of a typical LSTM network. The functional relationship between the variables is as follows:

$$i_t = \sigma\left(\mathbf{W}_{\mathrm{xf}} x_t + \mathbf{W}_{\mathrm{hi}} h_{t-1} + \mathbf{W}_{\mathrm{ci}} c_{t-1} + b_i\right) \quad (3)$$

$$f_t = \sigma\left(\mathbf{W}_{\mathrm{xf}} x_t + \mathbf{W}_{\mathrm{hf}} h_{t-1} + \mathbf{W}_{\mathrm{cf}} c_{t-1} + b_f\right) \quad (4)$$

$$c_t = f_t c_{t-1} + i_t \tanh\left(\mathbf{W}_{\mathrm{xc}} x_t + \mathbf{W}_{\mathrm{hc}} h_{t-1} + b_c\right) \quad (5)$$

$$o_t = \sigma\left(\mathbf{W}_{\mathrm{xo}} x_t + \mathbf{W}_{\mathrm{ho}} h_{t-1} + \mathbf{W}_{\mathrm{co}} c_t + b_o\right) \quad (6)$$

$$h_t = o_t \tanh\left(c_t\right) \quad (7)$$

Where: i, f, c, o denotes the input gate, forgetting gate, cell state, and output gate, respectively; b denotes the corresponding bias term; W denotes the weight matrix between the layer and each gate; is the sigmoid activation function; tanh is the hyperbolic tangent activation function.

As the output of one-way LSTM next moment forecasting is jointly influenced by the inputs from multiple past moments, it may result in a loss of useful information when extracting data features. Bi-LSTM, a DNN model, is based on a combination of LSTM and LSTM backwards, which is an advanced variant of the standard LSTM model [*Shahid et al.*, 2020]. By combining forward and backward instantaneous inputs, Bi-LSTM provides more accurate forecasting results [*Hernández-Pérez et al.*, 2020].

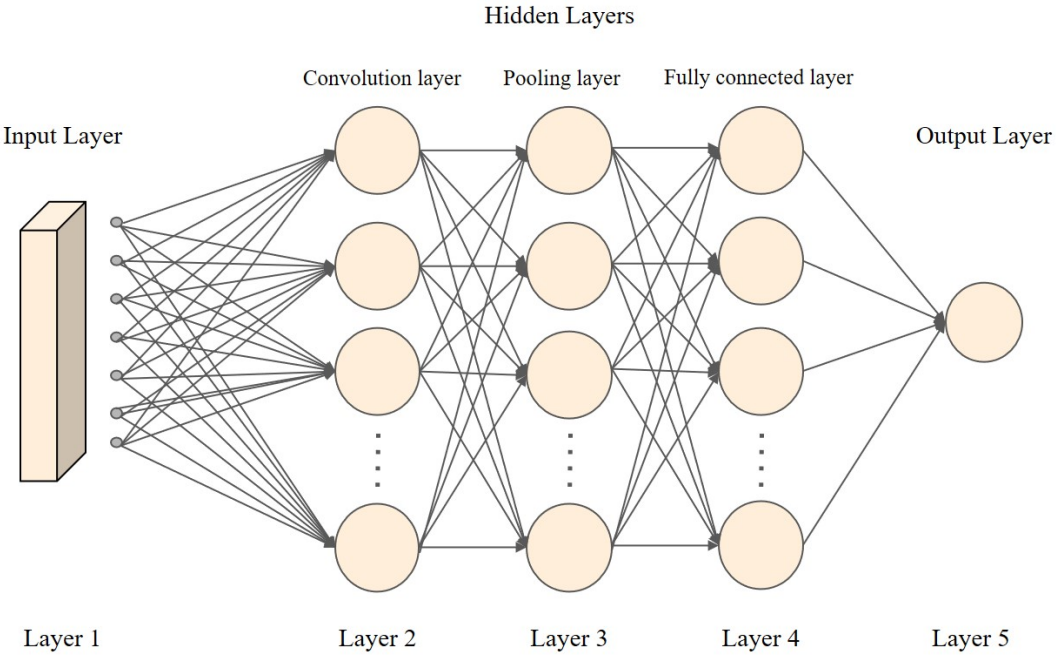
Considering the capability of Bi-LSTM to handle time series data, the m × c variables of n stations in the past t months are expanded into a two-dimensional array of t × v (v = m × n × c), i.e., 12 × 180 (180 = 6 × 10 × 3) dimensions, and these variables are added to the network as inputs. Moreover, the fully connected output layer generated the precipitation ^{18}O for n stations in the next t' months.

2.5 CNN-Bi-LSTM

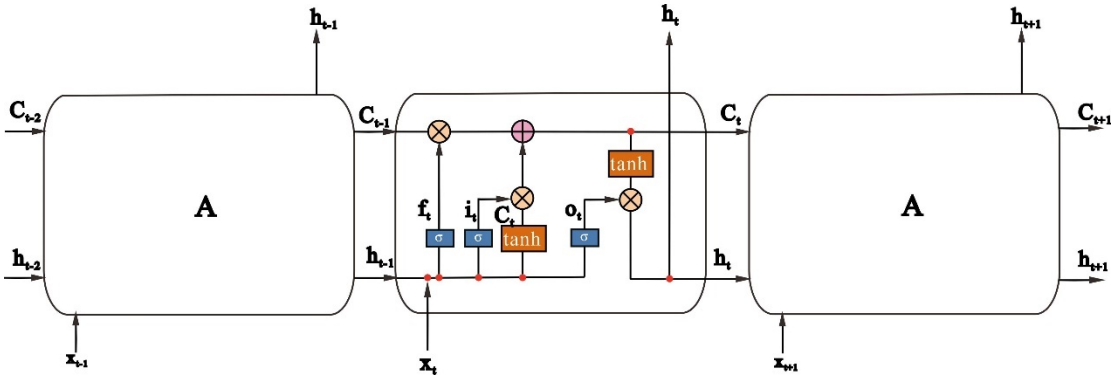
The CNN can learn the dependencies in the data, but its ability to capture the dependencies decreases as the input sequence lengthens. In contrast, the Bi-LSTM model can capture long-term dependencies between sequences, making them complementary.

In this study, CNN-Bi-LSTM is composed of several processes. The m × c variables of n sites in the past t months were used as input data. As with CNN, the input dimensions were t × m × n × c, i.e., 12 × 6 × 10 × 3. To extract features from the input data, the input, convolutional, and pooling layers of

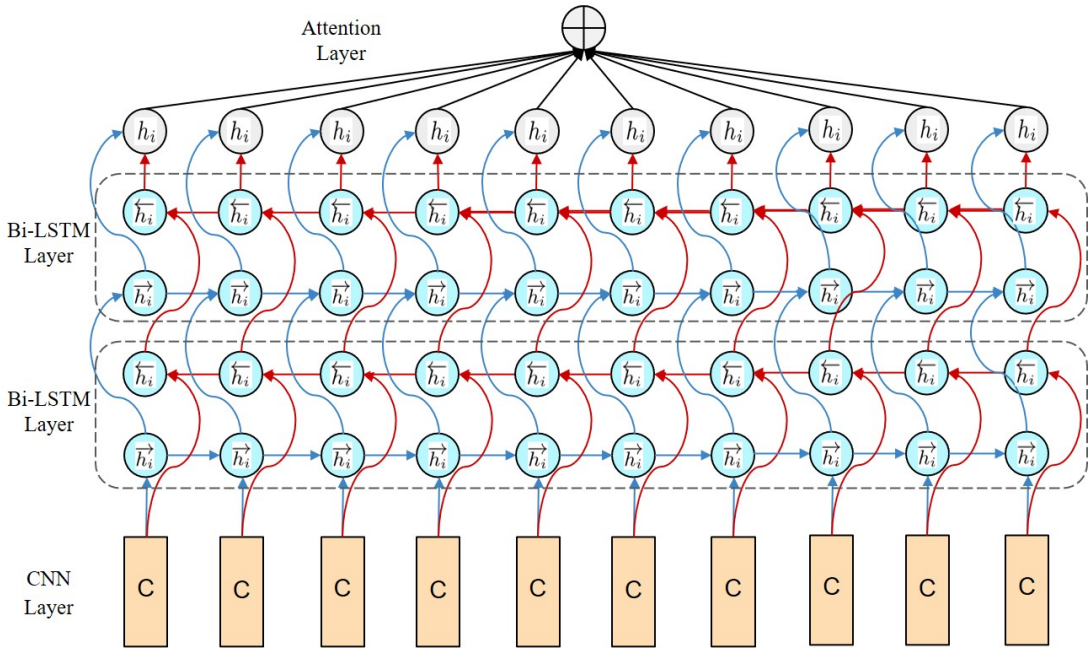
CNN were retained. The obtained features were tiled into a one-dimensional array and fed into the Bi-LSTM layer as time series to analyze their time series characteristics. Lastly, the output layer and the fully connected layer were used to calculate precipitation ^{18}O for n stations in the following t' months. The CNN-Bi-LSTM model's internal structure is shown in Fig. 1c.



(a) Structure of Convolutional Neural Network.



(b) The memory cell of LSTM.



(c) Schematic diagram of CNN-Bi-LSTM forecasting model.

Fig. 1 Structure of CNN, Bi-LSTM, and CNN- Bi-LSTM.

It is worth noting that, without any processing, raw data have a different order of magnitude from each other, and too large a difference can cause network forecasting errors, so input data are normalized and processed data are restricted to 0 to 1 [*White and Carty*, 2010]. To normalize the training data, we use minmax normalization, namely:

$$X' = \frac{X-min}{max-min} \quad (8)$$

Where: max represents the maximum value in the training sample; min represents the minimum value in the training sample X represents the original data; X' represents the normalized data; X' represents the normalized data.

2.6 Performance evaluation indices

In this paper, root mean squared error (RMSE), mean absolute error (MAE), correlation coefficient (R), and coefficient of determination (R^2) were used to evaluate the precipitation ^{18}O time variables forecasting model. The root mean square error is the difference between the predicted (test) value and the actual (true) value open root sign. The mathematical expression of RMSE is given in Equation (9). The mean absolute error is the average of the absolute error between the predicted and true values, and the mathematical expression of MAE is shown in Equation (10). The smaller the value of MAE, the closer the predicted value is to the true value, i.e., the better the forecasting is [*Chun et al.*, 2016].

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(\text{ test }_t - \text{ real }_t\right)^2} \quad (9)$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{t=1}^{N}\left|\text{ test }_t - \text{ real }_t\right| \quad (10)$$

Where N denotes the number of samples.

The correlation coefficient reveals the degree, direction, and importance of the relationship between the time series data observed by the proposed model and the predicted data. The mathematical expression of R is given in Equation (11). The larger the correlation coefficient, the stronger the relationship between the two variables.

$$R = \frac{1}{N-1}\sum_{i=1}^{N}\left(\frac{X_{\text{observed},i}-\mu_X}{\sigma_X}\right)\left(\frac{Y_{\text{estimated },i}-\mu_Y}{\sigma_Y}\right) \quad (11)$$

Where $X_{\text{observed },i}$ is the observed time series data, μ_X is the average, and σ_X is the standard deviation of the observed time series data; $Y_{\text{estimated },i}$ is estimated data, μ_Y is the average, and σ_y is the standard deviation of the estimated data.

Comparison of regression model forecasting error with Y = sample point average is performed via the coefficient of determination R^2. The mathematical expression of R^2 is given in Equation (12). A larger R^2 indicates a better fit of the data.

$$R^2 = 1 - \frac{\sum_{i=1}^{N}\left[X_{\text{observed},i}-Y_{\text{estimated},i}\right]^2}{\sum_{i=1}^{N}\left[X_{\text{observed},i}-\mu_X\right]^2} \quad (12)$$

3 Data

Due to its extensive isotope records, Germany was selected as the study area. For this study, monthly precipitation isotope (^{18}O), precipitation (P), temperature (T), latitude, longitude, and elevation data for 10 stations between 1988 and 2012 were selected as input variables for the forecasting model. The region, where the data were gauged, is situated between 5° and 15°E, 47° and 55°N in Germany. Fig. 2 illustrates the sites where the data were collected. The dataset was obtained from the third version of the WISER database updated in October 2017 (https://nucleus.iaea.org/wiser).All isotope data are expressed in per mil (‰) relative to the Vienna Standard Mean Ocean Water standard sample:

(1)

The elevation data were derived from the GTOPO1 digital elevation model (doi:10.7289/V5C8276M) of the U.S. Geological Survey, which provides 1 arc second spatial resolution.

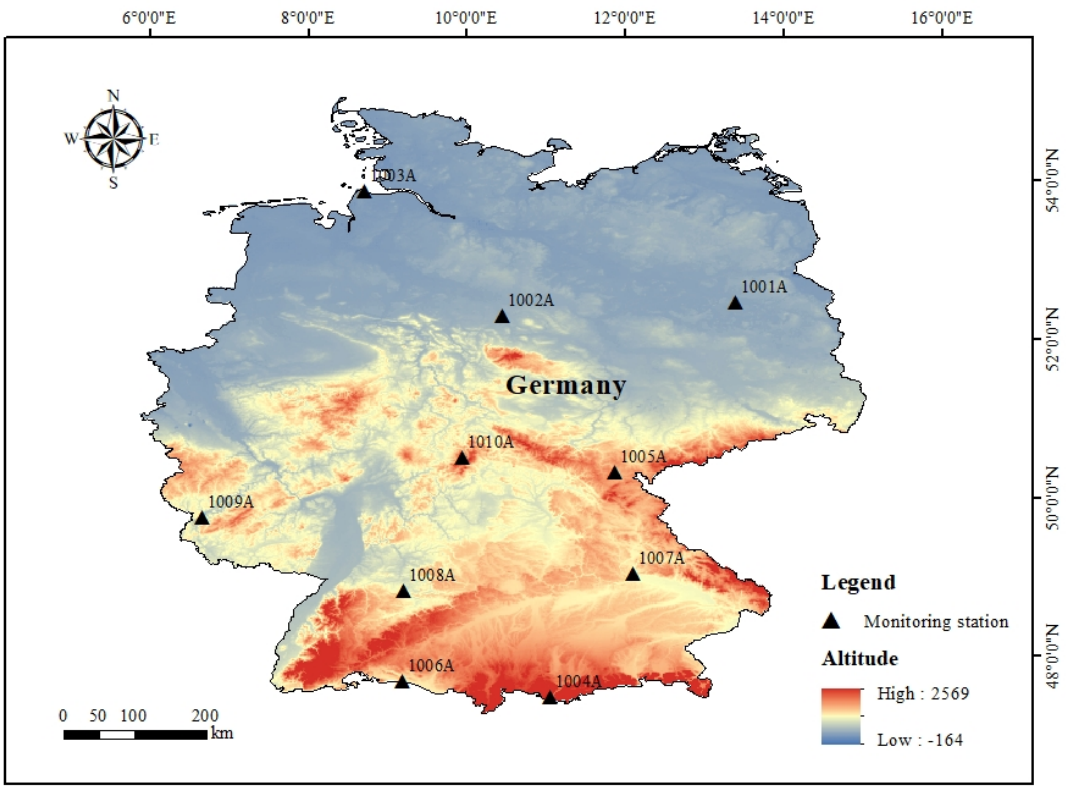


Fig. 2 Location of the study site.

With less than 5% missing data for each site, the data from all sites were retained despite the variable quality of the data. For some variables, however, average data from other sites were used since individual site data were missing at a particular time. At a particular point in time, if all site data for a variable were missing, the average data for each site during that time period was used to fill in the gaps.

Table 1

Investigated the precipitation isotope distribution monitoring sites in Germany.

Monitoring station	Latitude (°N)	Longitude (°E)	Altitude (m)	District	Group
1001A	52.4672	13.4019	48	Berlin	GNIP
1002A	52.2914	10.4464	81	Niedersachsen	GNIP
1003A	53.8713	8.7058	5	Niedersachsen	GNIP
1004A	47.4828	11.0622	719	Bayern	GNIP
1005A	50.3119	11.8758	565	Bayern	GNIP
1006A	47.6772	9.19	443	Baden-Württemberg	GNIP

Monitoring station	Latitude (°N)	Longitude (°E)	Altitude (m)	District	Group
1007A	49.0422	12.1019	365	Bayern	GNIP
1008A	48.8281	9.2	314	Baden-Württemberg	GNIP
1009A	49.7478	6.6581	265	Rheinland-Pfalz	GNIP
1010A	50.4972	9.9428	921	Hessen	GNIP

4 Results

4.1 Spatiotemporal distribution characteristics of the precipitation ^{18}O

On the temporal scale, variations in precipitation ^{18}O have been analyzed at the annual, seasonal, and monthly levels (Fig. 3). From 1988 to 2012, the monthly average precipitation ^{18}O in Germany was -25.2‰. In terms of seasonal differences, the average precipitation ^{18}O during the year decreased in the order of summer, spring, autumn, and winter. Monthly average precipitation ^{18}O in summer (-18.6±14.05‰) was significantly higher than the overall average, and monthly average precipitation ^{18}O in winter (-30.4±23.17‰) was significantly lower than the overall average. From 1988 to 2012, the monthly average precipitation ^{18}O variation ranged from -21.21‰ to -0.78‰. With an inverted U-shaped variation, precipitation ^{18}O was lower at the beginning and end of the year, and higher in the middle. The average precipitation ^{18}O is highest in July (-6.19±4.61‰) and is more stable. The mean precipitation ^{18}O is lower and more discrete in January and December. Temperature and precipitation amount influence precipitation isotope distribution, which accounts for these changes. In summer, rainfall is frequent, and heavy isotope ^{18}O falls with raindrops during a rainstorm, while light isotopes are enriched in the clouds [*Gower*, 1967]. In addition, the high temperatures in summer and the strong evaporation will cause secondary fractionation of isotopes as raindrops fall, enriching the underlying surface with heavy isotopes [*Barnes and Allison*, 1983]. By contrast, in winter, rainfall reduces the enrichment of heavy isotope ^{18}O in clouds. Additionally, low temperatures in winter cause less secondary fractionation of isotopes as raindrops fall, resulting in fewer heavy isotope ^{18}O to fall.

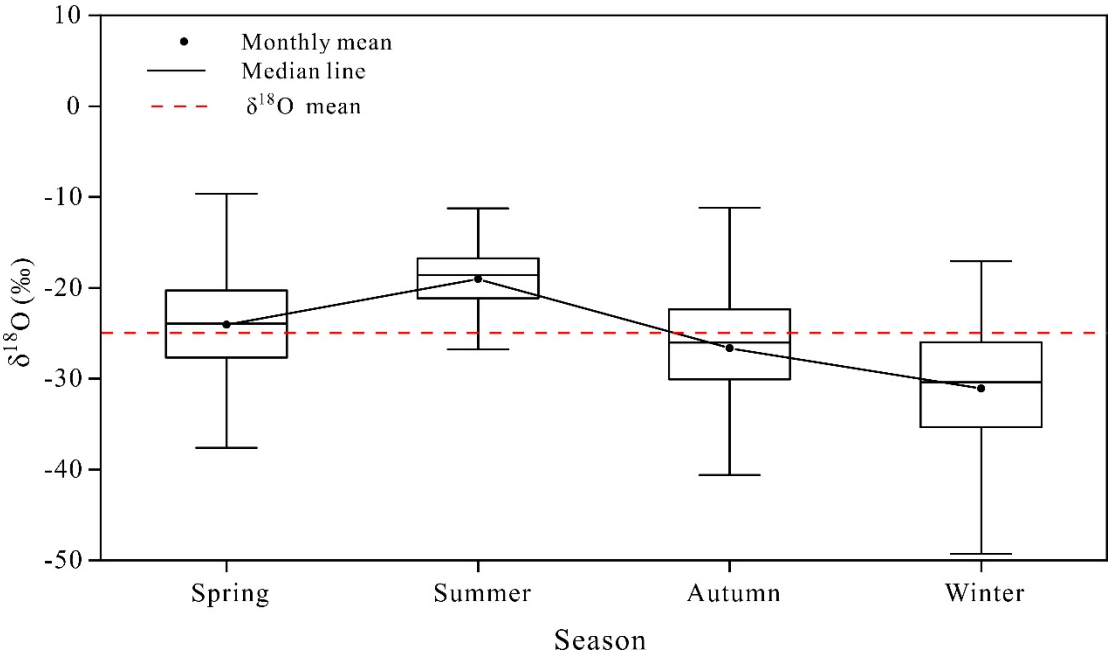


Fig. 3 Temporal distribution characteristics of the precipitation ^{18}O.

On the integrated spatial and temporal scales, this study used the K-means ++ method to classify the data from 10 isotope gauging stations in Germany from 1988-2012 into two categories (Fig. 4). As shown in the figure, different categories of stations have similar precipitation ^{18}O, temperature distributions, and precipitation distributions that differ. It may be because precipitation ^{18}O distribution in Germany is largely determined by temperature. Rainfall has little effect on precipitation ^{18}O distribution when the temperature and water vapor sources are the same [*C. Stumpp et al.*, 2014]. Additionally, the same precipitation ^{18}O may also be obtained for the same rainfall amount and different temperatures within the same category. It may be because the effect of temperature on precipitation ^{18}O is negligible when extreme rainfall conditions are present or when rainfall and regional environmental background influence precipitation more than temperature.

In general, temperature effects are the underlying cause for the different distributions of precipitation ^{18}O in German regions.

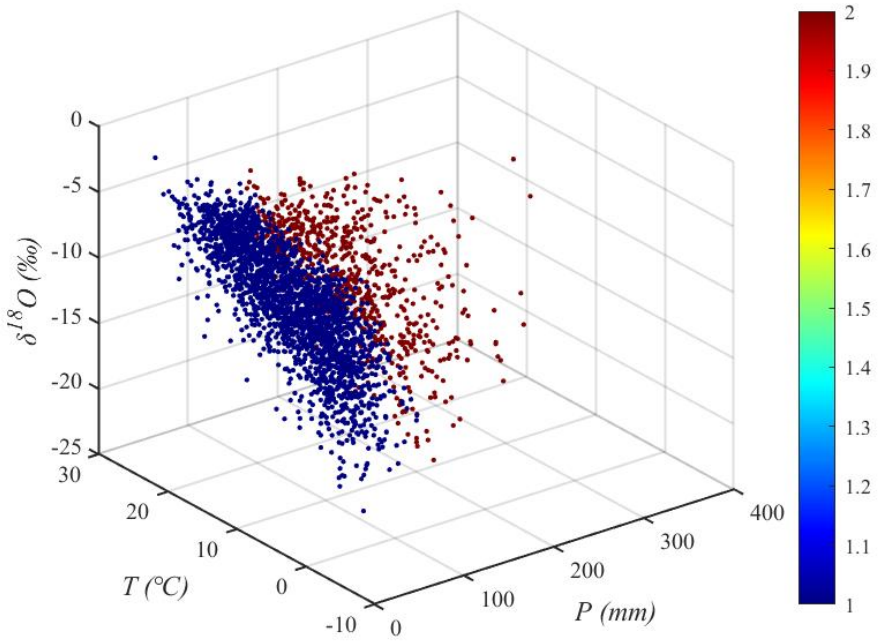


Fig. 4 Spatiotemporal distribution characteristics of the precipitation ^{18}O.

4.2 Overall precipitation ^{18}O forecasting model using deep learning

A comparison of the accuracy of three precipitation ^{18}O forecasting models based on multiple sites (10 sites) for multiple years (1988 to 2012) is shown in Fig. 5. Each 0.1 training ratio in the figure represents precipitation ^{18}O data for one site from 1988 to 2012. For the forecasting of multi-year precipitation ^{18}O at the single-site (i.e., 0.9 training ratio), the predicted RMSE, MAE are arranged in ascending order of Bi-LSTM, CNN-Bi-LSTM, and CNN, while the predicted R^2 and R are arranged in successive order of Bi-LSTM, CNN-Bi-LSTM, and CNN. Compared with CNN and CNN-Bi-LSTM, the RMSE of Bi-LSTM decreased by 37.84% and 7.57%; MAE decreased by 40.9% and 7.51%; R^2 increased by 7.69% and 2.95%; and R grew by 3.77% and 1.45%, respectively. The lower RMSE, MAE and higher R^2, R reflected the lower error and higher fit of Bi-LSTM in single-site, multi-year precipitation ^{18}O forecasting. Therefore, Bi-LSTM is considered as the optimal model for multi-year precipitation ^{18}O forecasting at the next station.

For the forecasting of multi-year precipitation ^{18}O at multi-site, CNN-Bi-LSTM has the lowest RMSE, MAE and highest R^2, R at 4, 5 sites (i.e., 0.6, 0.5 training ratio). Compared with CNN and Bi-LSTM, the RMSE of CNN-Bi-LSTM is reduced by 5.72%~18.97% and 11.59%~12.5%, respectively; MAE decreased by 7.74%~21.89% and 11.77%~14.44%, respectively; R^2 increased by

12.36%~27.29% and 5.95%~20.3%, respectively; R grew by 6.01%~12.82% and 2.92%~9.69, respectively. The forecasting of Bi-LSTM at 3 sites (i.e., 0.7 training ratio) has the lowest RMSE, MAE and the highest R^2, R. At 3 sites (i.e., 0.7 training ratio), Bi-LSTM has the best forecasting. Compared with CNN and CNN-Bi-LSTM, the RMSE of Bi-LSTM decreased by 11.47%, 8.53%; MAE decreased by 11.08%, 7.92%; R^2 increased by 33.1%, 8.48%; R grew by 15.38%, 4.16%, respectively. In addition, CNN-Bi-LSTM still has the lowest RMSE and MAE for forecasting at 2 sites (i.e., 0.8 training ratio). In comparison with CNN and Bi-LSTM, CNN-Bi-LSTM's RMSE is reduced by 84.13%, 6.64%, and MAE is reduced by 112.44%, 8.25%. Comparatively, it has a lower R^2 and R, which is 9.01% lower and 4.42% lower, respectively, than CNN. Combining the four metrics, CNN-Bi-LSTM is the best model for the forecasting of 2 sites. In terms of the total combined error, the CNN-Bi-LSTM reduced the RMSE by 29.68% and 6%, and the MAE by 38.76% and 7%, respectively, when compared with the CNN and Bi-LSTM. In terms of model forecasting stability, the average deviations of RMSE for CNN, Bi-LSTM, and CNN-Bi-LSTM is 0.75, 0.16, and 0.1, respectively; and the average deviations of MAE is 0.77, 0.14, and 0.08, respectively. Thus, CNN-Bi-LSTM is the most suitable model for predicting precipitation ^{18}O at multi-site over multi-year.

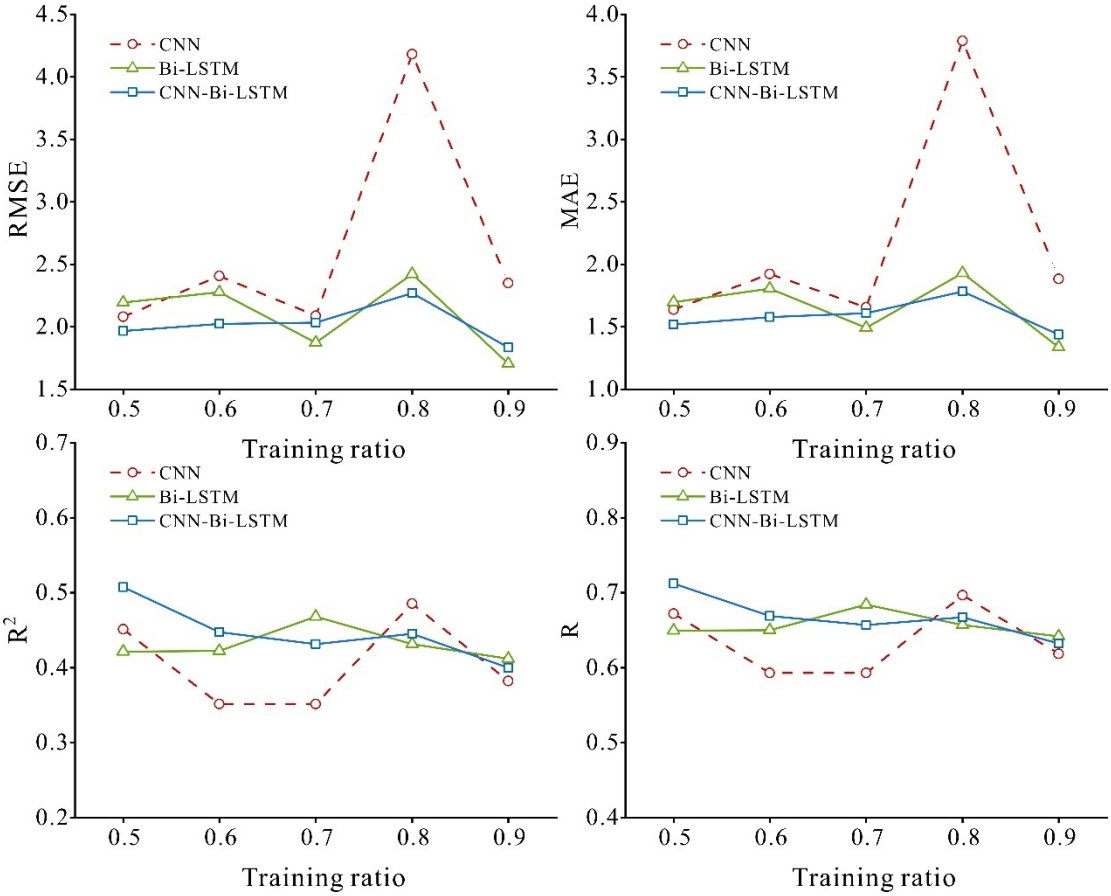


Fig. 5 Overall precipitation ^{18}O forecasting.

4.3 Seasonal forecasting

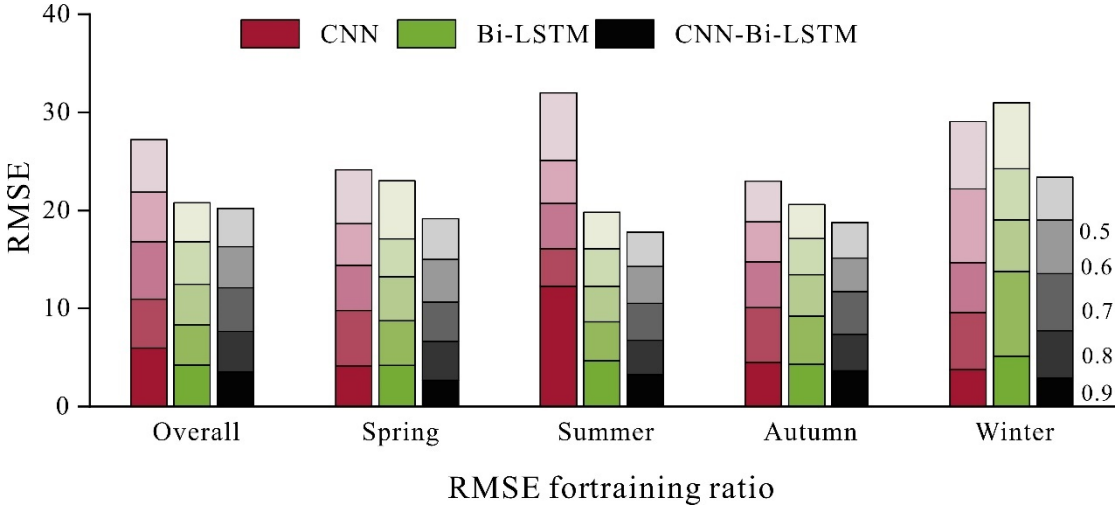
Since the precipitation ^{18}O varies seasonally, all data were divided into four clusters by season. Fig. 6 and Table 2 illustrate the differences between the seasonal and overall forecasting models using the three neural networks. The trend of the seasonal model predictive power is typically like that of the overall predictive model. The overall predictive power of the models is in descending order of CNN-Bi-LSTM, Bi-LSTM, and CNN.

Table 2

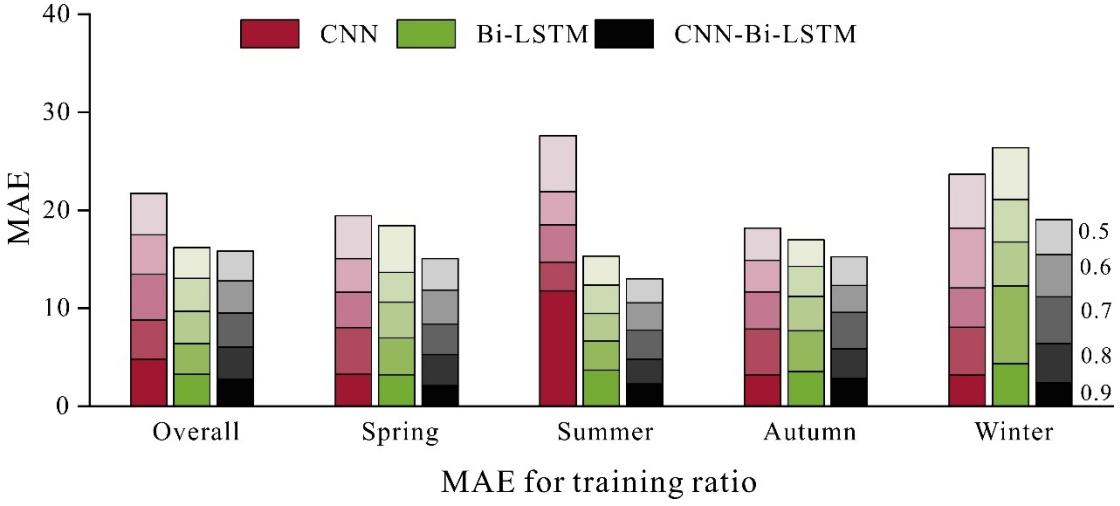
The calculation of model.

Sum	Season all	Spring	Summer	Autumn	Winter
RMSE	68.17	66.29	69.59	62.33	83.37
MAE	53.77	52.95	55.95	50.41	69.07
P	204859.02	44300.58	66858.11	49384.34	44316.00

Sum	Season all	Spring	Summer	Autumn	Winter
T	27088.82	6670.03	12977.99	6754.42	686.38
^{18}O	-25218.01	-6016.53	-4757.87	-6664.33	-7779.29



RMSE fortraining ratio



MAE for training ratio

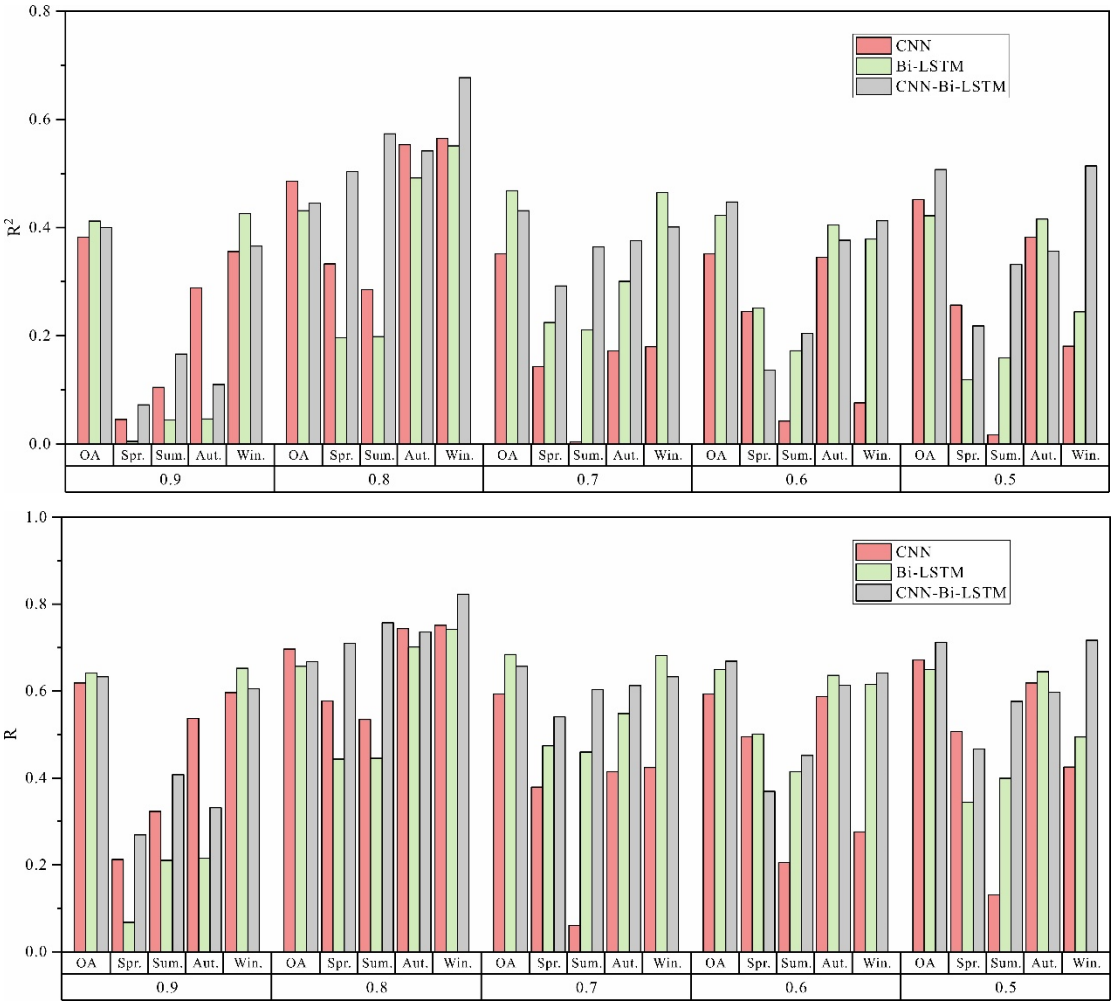


Fig. 6 Seasonal ^{18}O forecasting.

The variation of precipitation ^{18}O in different seasons can have different effects on forecasting accuracy. In terms of forecasting errors, the RMSE ([2.68,5.9]), MAE ([2.1,4.75]) in spring and RMSE ([3.45,5.59]), MAE ([2.7,4.69]) in autumn both lower than the overall level ([3.53,5.95]), ([2.8,4.81]). In contrast, RMSE ([3.24,12.25]), MAE ([2.28,11.79]) in summer and RMSE ([2.88,8.65]), MAE ([2.39,7.95]) in winter are higher than the overall level. Accordingly, seasonal classification may not always be effective in improving forecasting accuracy. The reason for this is that most of the highest temperatures (47.91% of the year) and most precipitation (32.64% of the year) occurred during the summer months. With higher temperatures during rainfall, more ^{18}O will be lost through evaporation. Once the temperature rises to some critical point, precipitation will produce secondary evaporation and isotopic fractionation will

be enhanced, so less ^{18}O will be retained [*Araguás-Araguás et al.*, 1998; *Tian et al.*, 2001]; Winter is the season with the lowest temperatures occured (2.53% of the year) and the lowest precipitation occured (21.63% of the year). With lower temperatures during rainfall, less ^{18}O will be lost through evaporation. Once temperatures drop below 0°C, precipitation will convert into ice and isotope fractionation will be stopped, so more ^{18}O will retained [*Lee et al.*, 2007; *Z Liu et al.*, 2010]. Notably, the R^2 ([0,0.5]), R ([0.07,0.71]) in spring, R^2 ([0,0.57]), R ([0.06,0.76]) in summer, R^2 ([0.05,0.55]), R ([0.22,0.74]) in autumn and R^2 ([0.08,0.68]), R ([0.28,0.82]) in winter are all lower than the overall level ([0.42,0.65]), ([0.65,0.81]). The distribution of forecasting across seasons does not show any specific pattern, and the degree of fit seems to depend only on the degree of dispersion between the data and the strength of the model.

In comparing the forecasting performance of different seasonal models, the CNN-Bi-LSTM has lower RMSE ([2.68,5.82]) and MAE ([2.1,4.76]) than the CNN's RMSE ([3.76,12.25]) and MAE ([3.76,12.25]) and the Bi-LSTM's RMSE ([3.45,8.65]) and MAE ([3.45,8.65]) respectively; however, its R^2 ([0.07,0.68]]) and R([0.27,0.82]) are higher than the CNN's R^2 ([0,0.56]) and R ([0.06,0.75]) and the Bi-LSTM's R^2 ([0,0.55]) and R ([0.07,0.74]) respectively. Overall, the RMSE of CNN-Bi-LSTM is reduced by 36.84% and 19.38%, respectively; the MAE is reduced by 42.44% and 23.56%, respectively, compared with CNN and Bi-LSTM. Further, the average deviations of RMSE for CNN, Bi-LSTM and CNN-Bi-LSTM models is 1.28, 0.87, and 0.57, respectively; the average deviations of MAE are 1.23, 0.83, and 0.53, respectively; the average deviations of R^2 are 0.13, 0.13, and 0.12, respectively; and the average deviations of R is 0.15, 0.14, and 0.12, respectively. The lower average deviations reflect the higher forecasting stability of CNN-Bi-LSTM. In summary, CNN-Bi-LSTM is the best model for predicting precipitation ^{18}O in different seasons.

4.4 Spatial and temporal clustering-based integrated forecasting

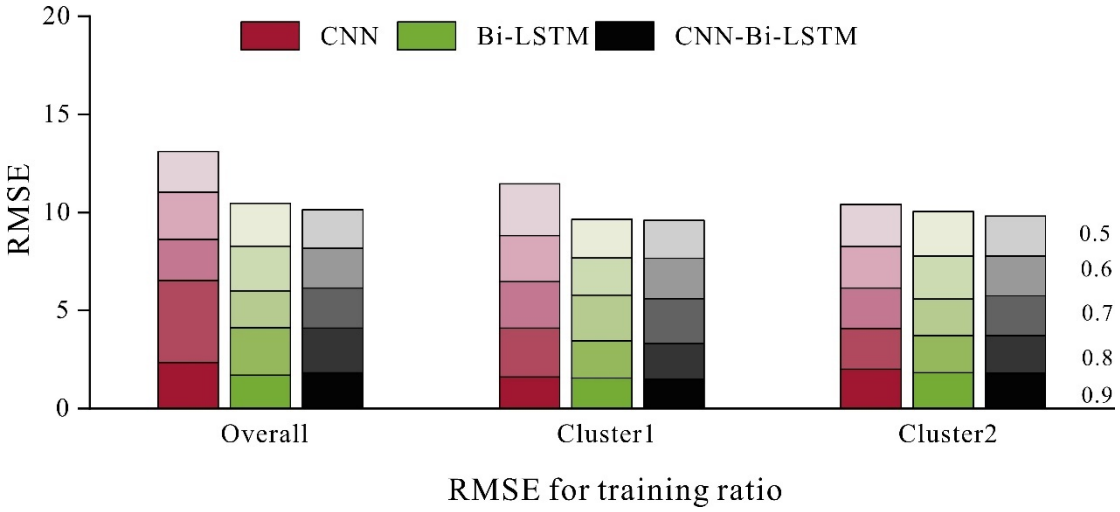
Based on the K-means ++ clustering results, data from 10 stations were divided into two categories (Fig. 5). They have similar precipitation ^{18}O and temperature distributions, but the first group has less precipitation distribution and the second group has more precipitation distribution. On the basis of the clustered data, three models were developed. As seen in Table 3 and Fig. 7, the results are as follows:

Table 3

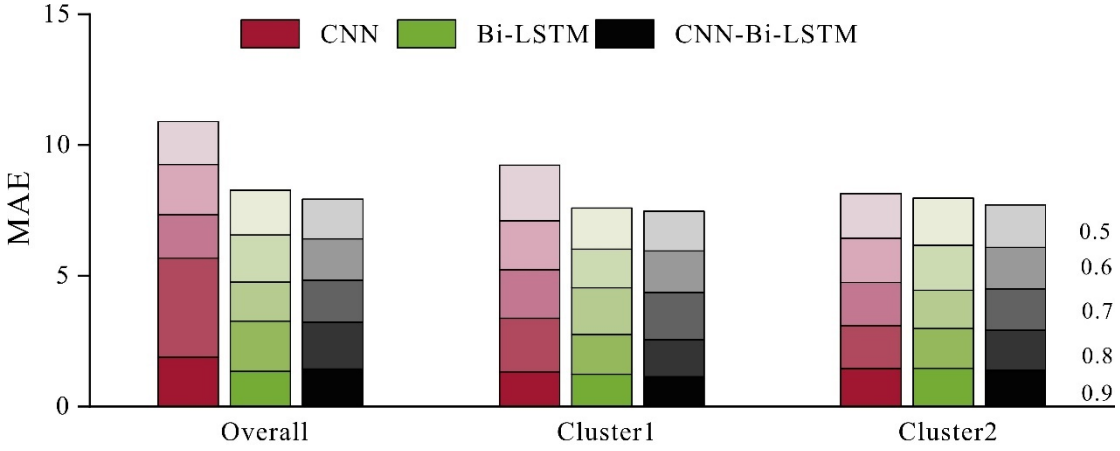
The calculation of model.

Sum	Overall	Cluster 1	Cluster 2
RMSE	33.70	30.71	30.26
MAE	27.08	24.29	23.83
P	204859.02	98473.06	106385.9679
T	27088.82	8043.43	19045.39611

Sum	Overall	Cluster 1	Cluster 2
^{18}O	-25218.01	-6704.56	-18513.4528



RMSE for training ratio



MAE for training ratio

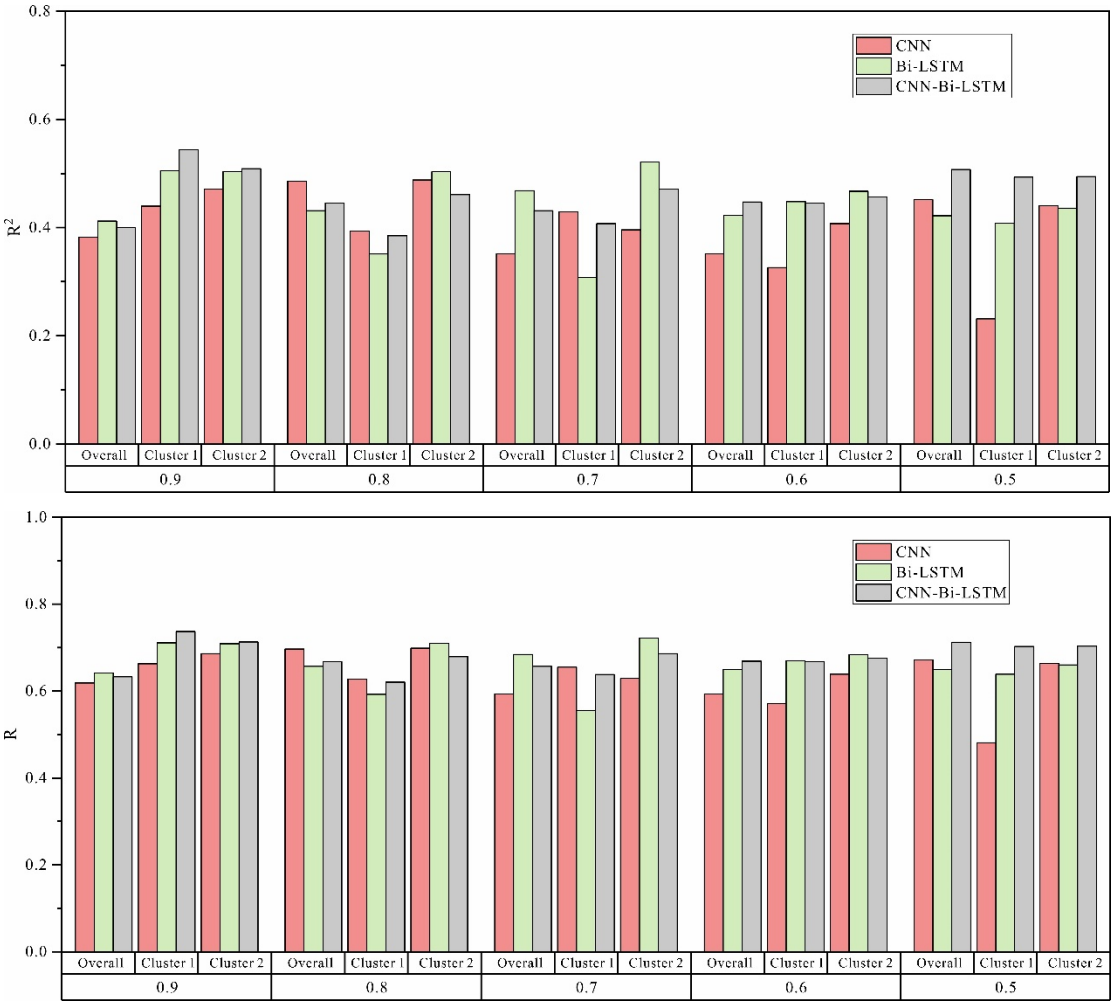


Fig. 7 K-means ++ clustering-based ^{18}O forecasting.

For the same cluster, different forecasting models with different training ratios perform differently. The accuracy of cluster 1 and cluster 2 forecasting is ranked in descending order of CNN-Bi-LSTM, Bi-LSTM, and CNN for the multi-year forecasting aspect of the single-site (i.e., 0.9 training ratio). Compared with CNN and Bi-LSTM, the RMSE of CNN-Bi-LSTM decreased by 9.37%~10.78% and 0.94%~5.77%, respectively; MAE decreased by 3.93%~16.52% and 5.13%~8.06%; R^2 improved by 7.94%~23.68% and 1.11%~7.58%, respectively; R improved by 3.89%~11.21% and 0.56%~3.71%, respectively. For multi-site, multi-year forecasting, cluster 1, cluster 2, and the overall forecasting show similar forecasting patterns. The forecasting accuracies of the models are all in descending order of CNN-Bi-LSTM, Bi-LSTM, CNN. Compared with CNN and Bi-LSTM, the RMSE of CNN-Bi-LSTM was reduced by 19.45% and 0.65%

in cluster 1 and 6.05% and 2.47% in cluster 2, respectively, compared to the overall; the MSE was reduced by 23.61% and 1.69% in cluster 1 and 5.76% and 3.41% in cluster 2, respectively, compared to the overall.

Similarly, the forecasting performance of the same model varies according to the cluster. For single-site, multi-year forecasting, the RMSE and MAE of CNN, Bi-LSTM, and CNN-Bi-LSTM are lower in cluster 1 than in cluster 2 (RMSE decreased by 23.73%, 16.57%, and 22.15%, respectively, and MAE decreased by 9.14%, 19.06%, and 22.36%, respectively); Bi-LSTM and CNN-Bi-LSTM have higher R^2 and R than cluster 2 (R^2 increased by 0.48% and 6.9%, and R grew by 0.24% and 3.38%, respectively), and CNN has lower R^2 and R than cluster 2 (R^2 decreased by 7.19% and R decreased by 3.54%). Forecasting accuracy of all models was ranked in descending order by cluster 1, cluster 2, and overall. For cluster 1, the variability of RMSE, MAE, R^2, and R increased with an increase in the number of predicted sites and time. As compared to cluster 2, the average deviations of RMSE, MAE, R^2, and R for cluster 1 increased by 1.71, 1.46, 1.11, and 1.32, respectively. With the 0.7, 0.8 training ratio forecasting, cluster 2 has a lower RMSE, MAE than cluster 1 and the overall RMSE, MAE; in the 0.6 training ratio forecasting, cluster 1 has a lower RMSE, MAE than cluster 2 and overall; it is evident from the 0.5 training ratio forecasting that the overall RMSE, MAE is lower than that of cluster s 1 and 2 RMSE, MAE, respectively. Thus, in single-site, multi-year forecasting, cluster 1 forecasting demonstrate some advantage in improving forecasting accuracy; cluster 2 forecasting improve the ability to reduce error and increase variance consistency, which are usually better than the cluster 1 forecasting and overall forecasting; once the proportion of forecasting reaches a certain threshold, the accuracy of forecasting is determined only by the amount of input data.

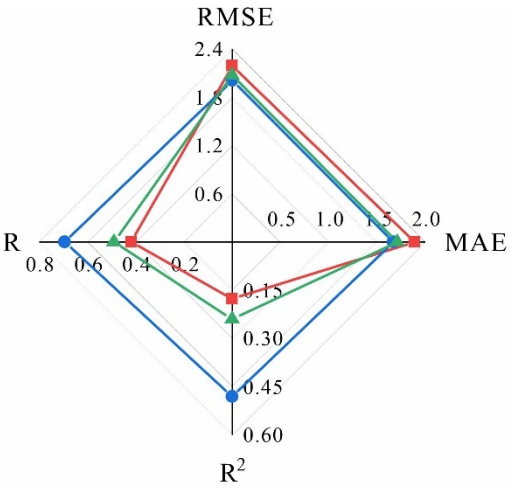
The difference in forecasting accuracy between different models in the same cluster is due to the different predictive abilities of the models themselves. But the reasons for the differences in forecasting accuracy of the same model in different cluster s may be diverse. The range of variation in precipitation in cluster 1 is [0.1,350.1] and in cluster 2 is [0.2,212.6], and the large fluctuating values may cause forecasting models to underestimate highs or overestimate lows. Moreover, the proximity of the additional forecasting sites to the training sites affects the spatial correlation of precipitation ^{18}O, and the worse the correlation, the greater the impact of spatial variability characteristics on the forecasting difficulty of each cluster.

4.5 Optimal model

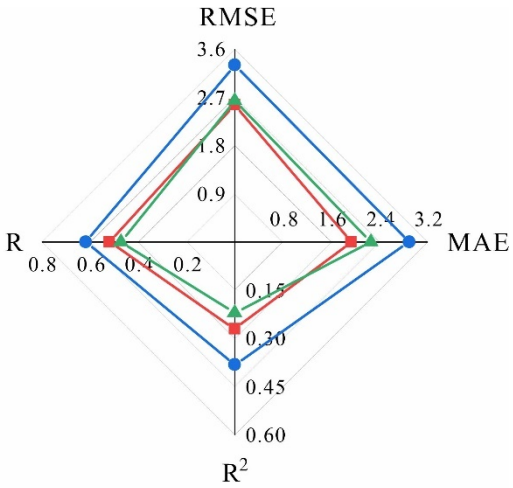
Compared with the overall forecasting, the seasonal forecasting does not show a significant improvement. In the context of single- and multi-site, multi-year precipitation ^{18}O forecasting, K-means ++ classification based on different features is capable of significantly reducing RMSE and MAE. Forecasting based on seasonality or spatiotemporal clustering do not improve the model's R^2, and the degree of fit seems to depend only on the degree of dispersion between the data and the strength of the model. In terms of model types, Bi-LSTM and

CNN-Bi-LSTM are the optimal models for single-site, multi-year forecasting and multi-site, multi-year forecasting, respectively. Consequently, CNN-Bi-LSTM is considered as the optimal forecasting model for all German stations.

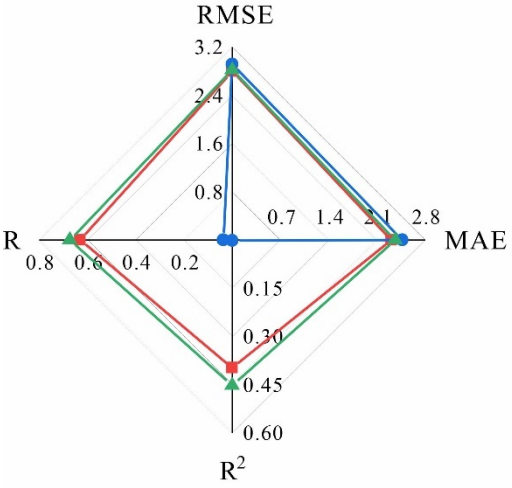
To further illustrate the performance differences between model types, see Fig. 8. Fig. 8 compares the accuracy of one-year forecasting for CNN, Bi-LSTM, and CNN-Bi-LSTM models at site 1002A. For forecasting one year in the future, the Bi-LSTM model has the highest accuracy. Compared with CNN and CNN-Bi-LSTM, Bi-LSTM has 8.79% and 2.82% lower RMSE, 13.13% and 2.33% lower MAE, 172% and 99.38% higher R^2, and 65.13% and 41.21% higher R, respectively. For forecasting two years in the future, CNN has the highest accuracy and the greatest forecasting. Compared with Bi-LSTM and CNN-Bi-LSTM, the RMSE of CNN is reduced by 28.78% and 2.77%, and the MAE is reduced by 49.79% and 17.24%, respectively. For forecasting more than two years in the future, the CNN-Bi-LSTM model provides significantly higher accuracy than both CNN and Bi-LSTM models.



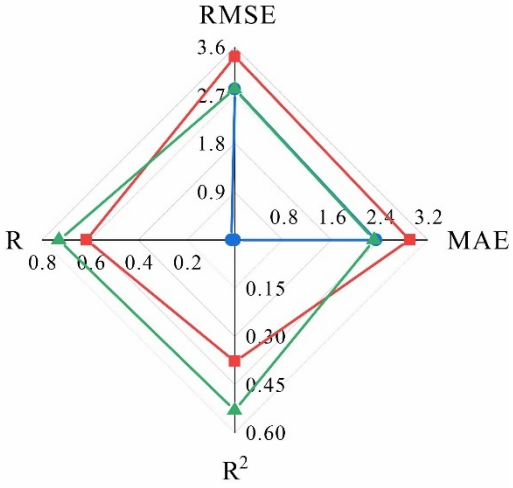
Forecast 1 year



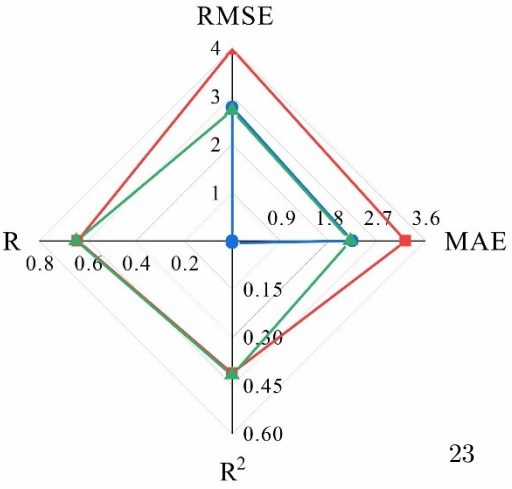
Forecast 2 years



Forecast 3 years



Forecast 4 years



Forecast 5 years

Fig. 8 Comparison of precipitation ^{18}O forecasting models for station 1002A.

5 Discussion

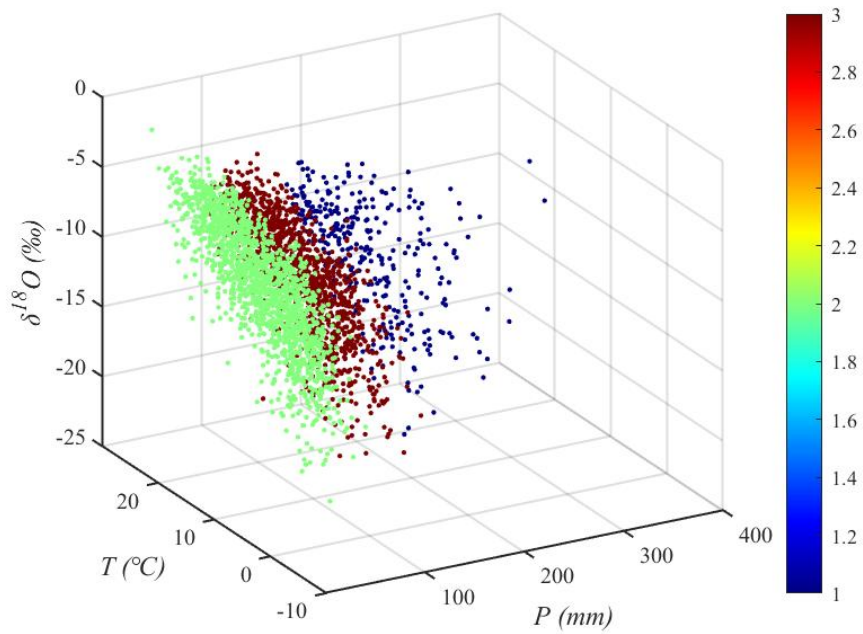
5.1 Enhancing clustering methods

To further validate the optimization of K-means ++ clustering for precipitation ^{18}O forecasting, we categorized the same data from Section 4.4 into three classes (Fig. 9a), built the model, and carried out multi-step forecasting. Fig. 9b illustrates the comparison results of different clusters.

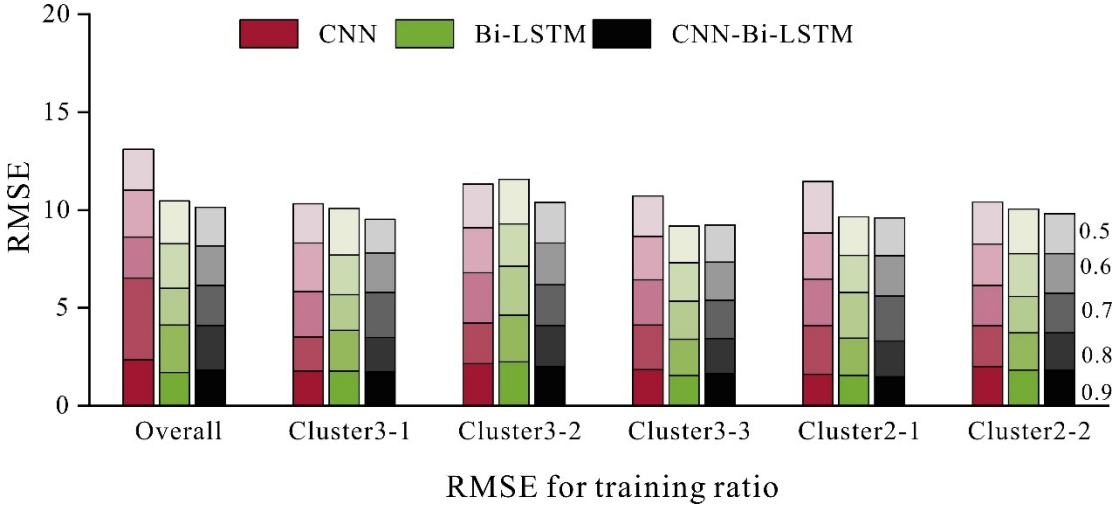
For single-site, multi-year forecasting, clusters 3-1 and 3-3 exhibit similar forecasting patterns. For clusters 3-1 and 3-3, the RMSE ([1.72,1.77]), ([1.54,1.85]) are significantly lower than the overall ([1.7,2.35]); MAE ([1.26,1.38]), ([1.21,1.5]) are lower than the overall ([1.34,1.88]); R^2 ([0.5,0.55]), ([0.36, 0.46]) are higher than the overall ([0.38, 0.41]); R([0.7, 0.74]), ([0.6, 0.68]) are higher than the overall ([0.62, 0.64]). Comparatively, cluster 3-2 has higher RMSE ([2,2.24]), MAE ([1.58,1.78]) and lower R^2 ([0.39,0.46], R ([0.63,0.68]) values than the overall. In terms of forecasting accuracy, clusters 3-1 and 3-2 have CNN-Bi-LSTM, CNN, and Bi-LSTM in order; cluster 3-3 has Bi-LSTM, CNN-Bi-LSTM, and CNN.

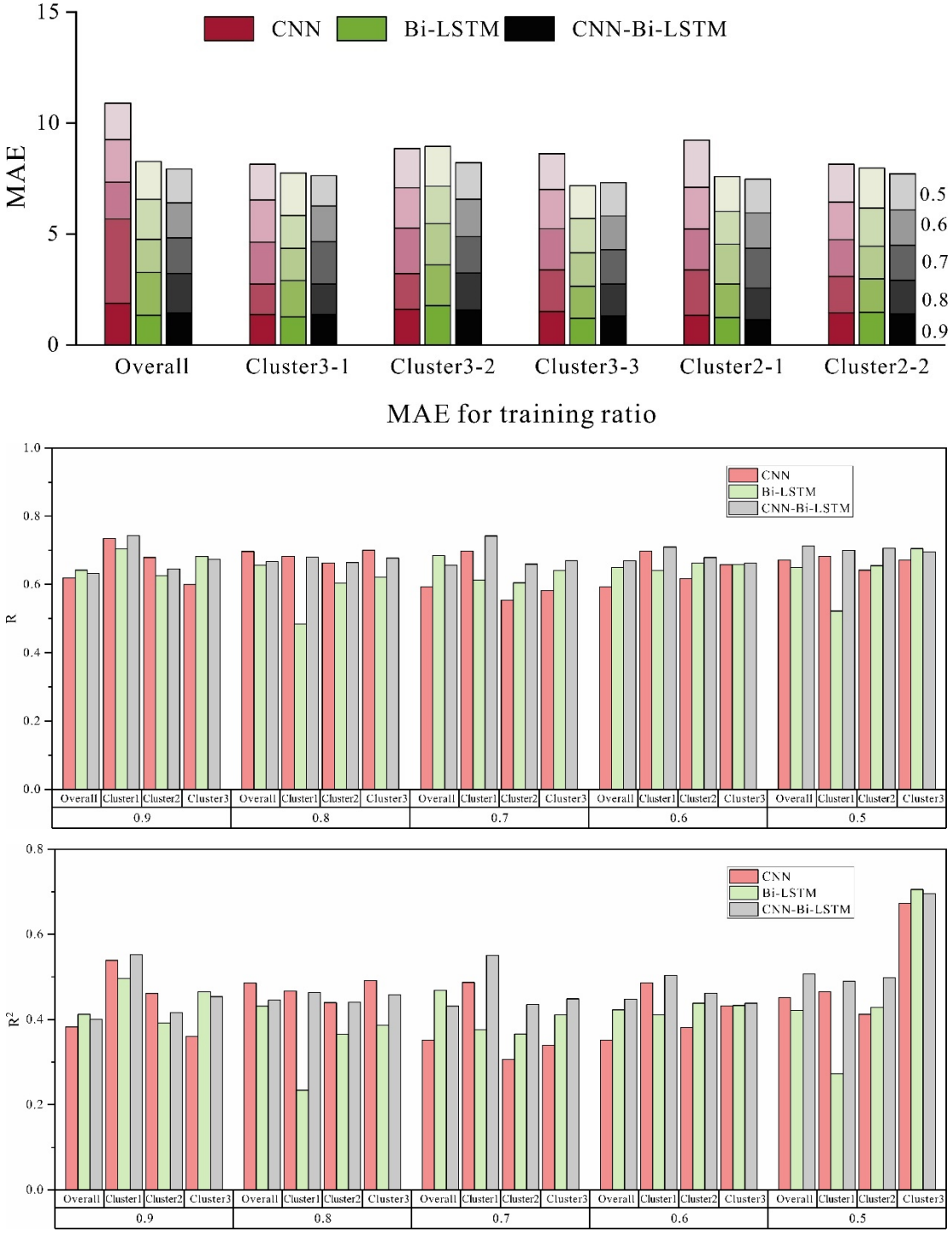
For multi-site, multi-year forecasting, clusters 3-1, 3-2, 3-3, and the overall forecasting have different forecasting patterns according to the training ratio. For cluster 3-1, the most accurate forecasting results were obtained at the 0.8 training ratio; for the overall forecasting, the most accurate results were obtained at the 0.7 training ratio; and for cluster 3-3, the most accurate forecasting results were obtained at the 0.6 and 0.5 training ratios. In addition, compared with the average deviation of 0.96, 0.95 for the overall forecasting RMSE and MAE, the average deviation of 0.32, 0.24 for clustering 3-1 RMSE and MAE, 0.22, 0.15 for clustering 3-2 RMSE and MAE, and 0.15, 0.1 for clustering 3-3 RMSE and MAE are lower, and the clustered model has a higher degree of stability. Regarding the type of model, CNN-Bi-LSTM is the optimal model for clusters 3-1, 3-2, and 3-3 as well as the overall forecasting. The results suggest that clustering does not always improve the accuracy of precipitation ^{18}O models and that optimal clustering results exist for precipitation ^{18}O forecasting models.

In future research, it may be worthwhile to investigate how best to cluster precipitation ^{18}O forecasting models for different regions and how different clusters may affect precipitation ^{18}O forecasting results.



(a) Spatiotemporal distribution characteristics of the precipitation ^{18}O.




CNN
Bi-LSTM
CNN-Bi-LSTM
MAE
15
10
5
0
0.5
0.6
0.7
0.8
0.9
Overall
Cluster3-1
Cluster3-2
Cluster3-3
Cluster2-1
Cluster2-2
MAE for training ratio
R
R^2
1.0
0.8
0.6
0.4
0.2
0.0
Overall
Cluster1
Cluster2
Cluster3
0.9
0.8
0.7
0.6
0.5

(b) K-means ++ clustering-based ^{18}O forecasting.

Fig. 9 Joint forecasting results of K-means ++ clustering and CNN, Bi-LSTM and CNN- Bi-LSTM models.

5.2 Improving precipitation ^{18}O forecasting

In this study, three deep learning methods and one spatiotemporal clustering method (K-means ++) were used to develop and compare precipitation ^{18}O forecasting models. The results show that CNN-Bi-LSTM performs better in multi-site forecasting; seasonal clustering-based forecasting is more suitable for season-specific forecasting; and K-means ++ clustering-based forecasting can significantly improve forecasting accuracy. For a single site, the Bi-LSTM is more suitable for ultra-short-term forecasting, the CNN is more suitable for short-term forecasting, and the CNN- B-LSTM is more suitable for medium- and long-term forecasting. Based on the results of this study, deep learning has proven to be an effective tool for handling big data, especially when the data is geographically and temporally distributed. In some cases, spatial-temporal clustering and models enhance the performance of spatial-temporal data forecasting.

Although clustering based on K-means++ has significantly improved precipitation ^{18}O forecasting accuracy, some factors remain unaccounted for. For example, to ensure the applicability of the model, we only use meteorological and geographical parameters in this study that have the greatest influence on the spatial and temporal distribution of precipitation ^{18}O and are the easiest to obtain as the input variables. However, characteristic parameters such as vegetation cover and slope can also have an impact on the spatial and temporal distribution of precipitation ^{18}O (Clark and Fritz, 1997), and it is unclear whether adding these parameters to the model forecasting would further enhance model accuracy. Furthermore, K-means ++ is only applicable to forecasting models with two main control parameters. How would increasing the control parameters of the model affect the forecasting results, and if there are better clustering methods available. Also, whether there are other models that are suitable for precipitation ^{18}O forecasting in addition to CNN, Bi-LSTM, and CNN-Bi-LSTM. In the following studies, these questions deserve to be explored further.

6 Conclusions

On the basis of a spatiotemporal clustering analysis of precipitation ^{18}O in Germany from 1988 to 2012, CNN, Bi-LSTM and CNN-Bi-LSTM multiyear, single-site, and multisite forecasting models have been developed and their performance compared in this study. The main findings are as follows:

1. On the temporal scale, there were obvious seasonal differences in precipitation ^{18}O, in the order of summer, spring, autumn and winter, with an inverted U-shaped variation on the annual, seasonal, and monthly scales. Four-season forecasting models were developed based on data clustering.

On the integrated spatial and temporal scales, different monitoring sites have similar precipitation ^{18}O and temperature distributions, while precipitation distributions differ. Temperature effects are the underlying cause for the different distributions of precipitation ^{18}O. A spatiotemporal clustering forecasting model based on two clusters was established by the K-means ++ method.

2. Using the entire dataset, the overall forecasting performance in single-site, multi-year forecasting is ranked in descending order of Bi-LSTM, CNN-Bi-LSTM, and CNN. For multi-site, multi-year forecasting, CNN-Bi-LSTM performs best with training ratios of 0.5, 0.6, and 0.8, but Bi-LSTM performs better at 0.7. In light of the small relative errors between CNN-Bi-LSTM and Bi-LSTM prediction values, the integrated CNN-Bi-LSTM-based forecasting model is considered to be the optimal forecasting model for each German region by combining the overall errors and forecast stability.

3. In comparison with the overall forecasting, the seasonal forecasting does not demonstrate a significant improvement in performance. The distribution of forecasting across seasons does not show any specific pattern, and the degree of fit seems to depend only on the degree of dispersion between the data and the strength of the model. Overall, the seasonal forecasting method is better suited for improving forecasting for a particular season. As a matter of model type, CNN-Bi-LSTM model tends to perform better than CNN and Bi-LSTM models.

4. After using K-means ++ clustering, the forecasting of cluster 1 shows some advantages in improving the prediction accuracy at the training ratio of 0.6 to 0.9, while the forecasting of cluster 2 improves the ability to reduce errors and improve variance consistency. And at the training ratio of 0.5, the overall forecasting outperformed cluster 1 and cluster 2. Thus, once the proportion of forecasting reaches a certain threshold, the accuracy of forecasting is determined only by the amount of input data. In terms of model type, CNN-Bi-LSTM model generally performs better than CNN and Bi-LSTM models.

Combining the above findings and additional studies, we found that the K-means ++ clustering followed by CNN-Bi-LSTM forecasting is the best model for forecasting precipitation ^{18}O in Germany. It is important to note that the number and type of clusters may have a significant impact on the prediction results.

Acknowledgments

This study is supported by the Natural Science Foundation of Hubei Province of China (2020CFB750), the National Natural Science Foundation of China (41602246) and Science and Technology Project of Hubei Geological Bureau (KJ2021-8). The authors would like to thank all the members of the project group for their help with laboratory work.

Open Research

1. Data Availability Statement:

Data archived in a repository: The [Precipitation isotopes, temperature, and precipitation in Germany] data used in the study [4 Results] can be [freely accessible] in [Datasets-Group-GNIP-Monthly] by registering an account at [https://nucleus.iaea.org/wiser/index.aspx].

The [CNN, Bi-LSTM, and CNN-Bi-LSTM model codes] used in the study can be [freely accessed] in [Precipitation-isotope-prediction-model] through the repository search after registering an account at [Github].

2. Software Availability Statement:

The [Matlab 2021a] used to run the [4 Results prediction model code] is available through the [Products section after registering an account at https://ww2.mathworks.cn/?s_tid=gn_logo].

References

Alzubaidi, L., J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaria, M. A. Fadhel, M. Al-Amidie, &L. Farhan (2021), Review of deep learning: concepts, CNN architectures, challenges, applications, future directions, *J Big Data*, *8*(1), 53. doi: 10.1186/s40537-021-00444-8.Araguás-Araguás, L., K. Froehlich, &K. Rozanski (1998), Stable isotope composition of precipitation over southeast Asia, *Journal of Geophysical Research: Atmospheres*, *103*(D22), 28721-28742. doi: 10.1029/98jd02582.Barnes, C. J., &G. B. Allison (1983), The distribution of deuterium and 18O in dry soils, *Journal of Hydrology*, *60*(1-4), 141-156. doi: 10.1016/0022-1694(83)90018-5 Barzegar, R., M. T. Aalami, &J. Adamowski (2020), Short-term water quality variable prediction using a hybrid CNN–LSTM deep learning model, *Stochastic Environmental Research and Risk Assessment*, *34*(2), 415-433. doi: 10.1007/s00477-020-01776-2.Belayneh, A., J. Adamowski, B. Khalil, &B. Ozga-Zielinski (2014), Long-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet neural network and wavelet support vector regression models, *Journal of Hydrology*, *508*, 418-429. doi: 10.1016/j.jhydrol.2013.10.052.Bowen, G. J., &B. Wilkinson (2002), Spatial distribution of 18O in meteoric precipitation, *Geology*, *30*(4). doi: 10.1130/0091-7613(2002)030<0315:Sdooim>2.0.Co;2.Bowen, G. J., &J. Revenaugh (2003), Interpolating the isotopic composition of modern meteoric precipitation, *Water Resour Res*, *39*(10). doi: 10.1029/2003wr002086.Chun, I. Y., S. Noh, D. J. Love, T. M. Talavage, S. Beckley, &S. J. Kisner (2016), Mean Squared Error (MSE)-Based Excitation Pattern Design for Parallel Transmit and Receive SENSE MRI Image Reconstruction, *IEEE Transactions on Computational Imaging*, 10.1109/tci.2016.2610141, 1-1. doi: 10.1109/tci.2016.2610141.Craig, H. (1961), Isotopic Variations in Meteoric Waters, *Science*, *133*(3465), 1702-1703. doi: 10.1126/science.133.3465.1702.Cui, J., J. Lang, T. Chen, S. Mao, S. Cheng, Z. Wang, &N. Cheng (2019), A framework for investigating the air quality variation characteristics based on the moni-

toring data: Case study for Beijing during 2013-2016, *J Environ Sci (China)*, *81*, 225-237. doi: 10.1016/j.jes.2019.01.009.Gaines, K. P., J. W. Stanley, F. C. Meinzer, K. A. McCulloh, D. R. Woodruff, W. Chen, T. S. Adams, H. Lin, &D. M. Eissenstat (2016), Reliance on shallow soil water in a mixed-hardwood forest in central Pennsylvania, *Tree Physiol*, *36*(4), 444-458. doi: 10.1093/treephys/tpv113.Gonfiantini, R., L. I. Wassenaar, L. Araguas-Araguas, &P. K. Aggarwal (2018), A unified Craig-Gordon isotope model of stable hydrogen and oxygen isotope fractionation during fresh or saltwater evaporation, *Geochim Cosmochim Ac*, *235*, 224-236. doi: 10.1016/j.gca.2018.05.020.Gower, J. C. (1967), A Comparison of Some Methods of Cluster Analysis, *Biometrics*, *23*(4). doi: 10.2307/2528417.Guo, H., G. Zhu, Y. He, J. Zhou, H. Pan, X. Ma, Y. Zhang, M. Huang, &J. Xiang (2019), Dynamic characteristics and influencing factors of precipitation 18O, China, *Theoretical and Applied Climatology*, *138*(1-2), 899-910. doi: 10.1007/s00704-019-02867-z.Hernández-Pérez, E., G. Levresse, J. Carrera-Hernández, &R. García-Martínez (2020), Short term evaporation estimation in a natural semiarid environment: New perspective of the Craig – Gordon isotopic model, *Journal of Hydrology*, *587*. doi: 10.1016/j.jhydrol.2020.124926.Hochreiter, S. (2011), The Vanishing Gradient Problem During Learning Recurrent Neural Nets and Problem Solutions, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *06*(02), 107-116. doi: 10.1142/s0218488598000094.Kattenborn, T., J. Leitloff, F. Schiefer, &S. Hinz (2021), Review on Convolutional Neural Networks (CNN) in vegetation remote sensing, *ISPRS Journal of Photogrammetry and Remote Sensing*, *173*, 24-49. doi: 10.1016/j.isprsjprs.2020.12.010.Klaus, J., &J. J. McDonnell (2013), Hydrograph separation using stable isotopes: Review and evaluation, *Journal of Hydrology*, *505*, 47-64. doi: 10.1016/j.jhydrol.2013.09.006.Klaus, J., K. P. Chun, &C. Stumpp (2015), Temporal trends in 18O composition of precipitation in Germany: insights from time series modelling and trend analysis, *Hydrol Process*, *29*(12), 2668-2680. doi: 10.1002/hyp.10395.Klaus, J., J. J. McDonnell, C. R. Jackson, E. Du, &N. A. Griffiths (2015), Where does streamwater come from in low-relief forested watersheds? A dual-isotope approach, *Hydrology and Earth System Sciences*, *19*(1), 125-135. doi: 10.5194/hess-19-125-2015.Koeniger, P., C. Leibundgut, T. Link, &J. D. Marshall (2010), Stable isotopes applied as water tracers in column and field studies, *Org Geochem*, *41*(1), 31-40. doi: 10.1016/j.orggeochem.2009.07.006.Koeniger, P., M. Gaj, M. Beyer, &T. Himmelsbach (2016), Review on soil water isotope-based groundwater recharge estimations, *Hydrological Processes*, *30*(16), 2817-2834. doi: 10.1002/hyp.10775.Kratzert, F., D. Klotz, C. Brenner, K. Schulz, &M. Herrnegger (2018), Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks, *Hydrology and Earth System Sciences*, *22*(11), 6005-6022. doi: 10.5194/hess-22-6005-2018.Latifoglu, L. (2022), A novel combined model for prediction of daily precipitation data using instantaneous frequency feature and bidirectional long short time memory networks, *Environ Sci Pollut Res Int*, *29*(28), 42899-42912. doi: 10.1007/s11356-022-18874-z.LeCun, Y., Y. Bengio, &G. Hinton (2015), Deep learning, *Nature*, *521*(7553), 436-444. doi: 10.1038/nature14539.Lee, J.-E., I. Fung, D. J. DePaolo, &C. C.

Henning (2007), Analysis of the global distribution of water isotopes using the NCAR atmospheric general circulation model, *Journal of Geophysical Research*, *112*(D16). doi: 10.1029/2006jd007657.Li, J., F. Li, Q. Liu, &Y. Suzuki (2014), Nitrate pollution and its transfer in surface water and groundwater in irrigated areas: a case study of the Piedmont of South Taihang Mountains, China, *Environ Sci Process Impacts*, *16*(12), 2764-2773. doi: 10.1039/c4em00200h.Li, Y. Q., Y. C. Meng, S. Song, C. H. Du, &Q. Y. Xiang (2021), [Distribution of hydrogen and oxygen stable isotope of water in soil-plant-atmosphere continuum (SPAC)system of a typical forest area], *Ying Yong Sheng Tai Xue Bao*, *32*(6), 1928-1934. doi: 10.13287/j.1001-9332.202106.020.Liu, H., H.-q. Tian, D.-f. Pan, &Y.-f. Li (2013), Forecasting models for wind speed using wavelet, wavelet packet, time series and Artificial Neural Networks, *Applied Energy*, *107*, 191-208. doi: 10.1016/j.apenergy.2013.02.002.Liu, X., Z. Rao, X. Zhang, W. Huang, J. Chen, &F. Chen (2015), Variations in the oxygen isotopic composition of precipitation in the Tianshan Mountains region and their significance for the Westerly circulation, *Journal of Geographical Sciences*, *25*(7), 801-816. doi: 10.1007/s11442-015-1203-x.Liu, Z., G. J. Bowen, &J. M. Welker (2010), Atmospheric circulation is reflected in precipitation isotope gradients over the conterminous United States, *Journal of Geophysical Research*, *115*(D22). doi: 10.1029/2010jd014175.Mohammadi, B., &S. Mehdizadeh (2020), Modeling daily reference evapotranspiration via a novel approach based on support vector regression coupled with whale optimization algorithm, *Agricultural Water Management*, *237*. doi: 10.1016/j.agwat.2020.106145.Panahi, M., N. Sadhasivam, H. R. Pourghasemi, F. Rezaie, &S. Lee (2020), Spatial prediction of groundwater potential mapping based on convolutional neural network (CNN) and support vector regression (SVR), *Journal of Hydrology*, *588*. doi: 10.1016/j.jhydrol.2020.125033.Shahid, F., A. Zameer, &M. Muneeb (2020), Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM, *Chaos Solitons Fractals*, *140*, 110212. doi: 10.1016/j.chaos.2020.110212.Shi, Y., S. Wang, M. Zhang, A. A. Argiriou, R. Guo, Y. Song, &X. Zhu (2020), Isoscape of 18O in Precipitation of the Qinghai-Tibet Plateau: Assessment and Improvement, *Water-Sui*, *12*(12). doi: 10.3390/w12123392.Stumpp, C., &P. Maloszewski (2010), Quantification of preferential flow and flow heterogeneities in an unsaturated soil planted with different crops using the environmental isotope 18O, *Journal of Hydrology*, *394*(3-4), 407-415. doi: 10.1016/j.jhydrol.2010.09.014.Stumpp, C., J. Klaus, &W. Stichler (2014), Analysis of long-term stable isotopic composition in German precipitation, *Journal of Hydrology*, *517*, 351-361. doi: 10.1016/j.jhydrol.2014.05.034.Takashi Onoda, M. S., Seiji Yamada (2010), Careful Seeding Based on Independent Component Analysis for k-Means Clustering, paper presented at 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology.Tian, L., T. Yao, W. Sun, M. Stievenard, &J. Jouzel (2001), Relationship between D and 18O in precipitation on north and south of the Tibetan Plateau and moisture recycling, *Science in China Series D: Earth Sciences*, *44*(9), 789-796. doi: 10.1007/bf02907091.White, P., &M. Carty (2010), Reducing bias through process inventory dataset normalization, *The International*

Journal of Life Cycle Assessment, *15*(9), 994-1013. doi: 10.1007/s11367-010-0215-0.Wu, Y., L. Lin, J. Wang, &S. Wu (2020), [Application of semantic segmentation based on convolutional neural network in medical images], *Sheng Wu Yi Xue Gong Cheng Xue Za Zhi*, *37*(3), 533-540. doi: 10.7507/1001-5515.201906067.Xiang, Z., J. Yan, &I. Demir (2020), A Rainfall-Runoff Model With LSTM-Based Sequence-to-Sequence Learning, *Water Resour Res*, *56*(1). doi: 10.1029/2019wr025326.Yan, R., J. Liao, J. Yang, W. Sun, M. Nong, &F. Li (2021), Multi-hour and multi-site air quality index forecasting in Beijing using CNN, LSTM, CNN-LSTM, and spatiotemporal clustering, *Expert Systems with Applications*, *169*. doi: 10.1016/j.eswa.2020.114513.Yang, Q. C., H. K. Mu, H. Wang, X. Y. Ye, H. Y. Ma, &J. D. Martin (2018), Quantitative evaluation of groundwater recharge and evaporation intensity with stable oxygen and hydrogen isotopes in a semi-arid region, Northwest China, *Hydrological Processes*, *32*(9), 1130-1136. doi: 10.1002/hyp.11474.Zhao, P., X. Y. Tang, P. Zhao, C. Wang, &J. L. Tang (2013), Identifying the water source for subsurface flow with deuterium and oxygen-18 isotopes of soil water collected from tension lysimeters and cores, *Journal of Hydrology*, *503*, 1-10. doi: 10.1016/j.jhydrol.2013.08.033.Zhu, Q., &T. Burzykowski (2013), A Markov-chain-based regression model with random effects for the analysis of18O-labelled mass spectra, *Journal of Statistical Computation and Simulation*, *83*(1), 145-157. doi: 10.1080/00949655.2011.620610.