Multi-Model Coupling Water Demand Prediction Optimization Method for Megacities Based on Time Series Decomposition

Xin Liu (✉ xinliudoc@outlook.com)

North China University of Water Resources and Electric Power   https://orcid.org/0000-0003-0902-6277

Xuefeng Sang

China Institute of Water Resources and Hydropower Research

Jiaxuan Chang

China Institute of Water Resources and Hydropower Research

Yang Zheng

China Institute of Water Resources and Hydropower Research

Multi-Model Coupling Water Demand Prediction Optimization Method for Megacities Based on Time Series Decomposition

Xin Liu[1,2] · **Xuefeng Sang**[2] · **Jiaxuan Chang**[2] · **Yang Zheng**[2]

Abstract

The fluctuation of water supply is affected by the living habits and population mobility, so the daily water supply is significantly non-stationarity, which presents a great challenge to the water demand prediction based on data-driven model. To solve this problem, the Hodrick-Prescott (HP) and wavelet transform (WT) time series decomposition methods, and ensemble learning (EL) were introduced, coupling model bidirectional long short term memory (BLSTM), seasonal autoregressive integrated moving average (SARIMA) and Gaussian radial basis function neural network (GRBFNN) were developed, and interval prediction was carried out based on student's t-test (T-test). This research method was applied to the daily water demand prediction in Shenzhen and cross-validation was performed. It is found that the decomposed subseries has obvious law, and WT is superior to HP decomposition method. However, the maximum decomposition level (MDL) of WT should not be set too high, otherwise the trend characteristics of subseries will be weakened. The results show that the potential characteristics and quantitative relationships of historical data can be learned accurately based on WT and coupling model. Although the corona virus disease 2019 (COVID-19) outbreak in 2020 caused a variation in water supply law, this variation is still within the interval prediction. The WT and

Xuefeng Sang
sangxf@iwhr.com

Extended author information available on the last page of the article

coupling model satisfactorily predicted water demand and provided the lowest mean square error (0.17%), mean relative error (0.1) and mean absolute error (3.32%) and the highest Nash-Sutcliffe efficiency (97.21%) and correlation coefficient (0.99) in testing set.

1 Introduction

The water demand prediction has always been a hot topic for scholars. The megacities population and buildings are dense, and the megacities have a high modernization degree. Therefore, the local water resources are relatively short, and the water supply mainly depends on water diversion. For megacities, the water demand prediction is very important for the rational allocation of water resources and the water diversion plan. Because the accurate water demand prediction is conducive to know water supply and demand balance, so as to discover the shortage of water resources in advance. Therefore, the water demand prediction is particularly important. In recent years, data-driven models, such as regression (Huang et al. 2020; Li and Willems 2020; Akbari et al. 2019), artificial neural network (ANN) (Amaranto et al. 2019; Baek et al. 2019; Hadi and Tombul 2019), time series (Kajewska-Szkudlarek 2020; Mehdizadeh 2020; Zarei and Mahmoudi 2020) and deep learning (Ma et al. 2020; Zhong et al. 2019; Li et al. 2016) models, have been widely used in hydrological prediction. This is because the data-driven models will not be affected by the external physical environment. It is based on mining the potential correlation between data to establish the quantitative relationship (Fu et al. 2019) between input and output features, the modeling speed is fast and the prediction accuracy

is high. In the fields of precipitation prediction (Pan et al. 2019), flood prediction (Khan et al. 2018), water demand prediction (Al-Zahrani et al. 2015; Haque et al. 2017) and water quality prediction (Mahmoudi et al. 2016), many good results have been achieved.

However, as the units of time series get shorter, the generalization ability of data-driven models becomes worse. Based on a large number of experiments and analysis, the non-stationarity of time series (Yang et al. 2020; Serinaldi et al. 2018) has a great influence on the predicted results. The stationarity of annual data or monthly data is better than the daily data, so it is easier to construct the model using such data. For the daily water supply data in megacities, the stationarity of the data is poor. If the data-driven model is directly constructed, the generalization ability of the model may seriously deteriorate, which brings new challenges to the data-driven model. In addition, it is found that the multi-collinearity (Bassiouni et al. 2016; Huang et al. 2015) among variables leads to the distortion of the multiple regression model. The ANN model is prone to overfitting (Sari et al. 2017; Piotrowski and Napiorkowski 2013) due to its limited learning ability, and the global parameter adjustment is prone to local optimal solution (Sun et al. 2016). Meanwhile, the poor randomness of the model is also a key factor affecting the generalization ability of the model.

Although the stationarity of daily water supply data in megacities is not good, the data has the potential law. If the appropriate method is used to decompose the time series, and the prediction and reconstruction of the subseries can improve the prediction accuracy. Hodrick-Prescott (HP) (Chen et al. 2020) decomposition method is commonly used at present, which decomposes time series into trend subseries (TS) and period subseries (PS). However, the experiment results show that the stationarity of PS generated by HP is not good, and the

significance level of the period law is affected by the irregular values. Sometimes, PS contain many irregular values, very similar to Gaussian noise (Pathiraja et al. 2018). Therefore, in this study, wavelet transform (WT) (Poul et al. 2019; Wright et al. 2017) is used to decompose the time series into TS, PS and noise subseries (NS). The coupling model bidirectional long short term memory (BLSTM), seasonal autoregressive integrated moving average (SARIMA) and Gaussian radial basis function neural network (GRBFNN) were developed for subseries prediction, and compared with the predicted results of HP. For different subseries, different data-driven models are used for prediction, and the interval prediction is carried out. It is expected that through the decomposition, prediction and reconstruction of time series, the predicted value with less error from the measured value can be obtained.

2 Material and Methods

2.1 Study Area and Dataset

Shenzhen is located between longitudes of 113°43' and 114°38' E, and latitudes of 22°24' and 22°52' N, adjacent to Hong Kong. It is a sub-provincial city of Guangdong Province and the first city in China to be fully urbanized. Shenzhen is a port city with the largest number of ports, the largest number of entry-exit people and the largest traffic flow in China, but Shenzhen is facing a serious water resources shortage problem. Due to the influence of salt tide, the groundwater supply is very small, and the water supply dispatching mainly depends on the water diversion. With the rapid development of cities around the water diversion project, the water sources have been polluted to varying degrees, which further exacerbates the crisis of Shenzhen's water resources. Therefore, if the daily water demand in Shenzhen can be accurately

predicted, the allocation of water resources can be effectively and scientifically carried out. Shenzhen can make full use of the rainfall in the flood season and find out the potential water supply shortage in the future in advance, which is of great significance to the sustainable development of the society. The data in this study are from the daily measured data without vacancy of Shenzhen Digital Water System from January 1, 2015 to December 31, 2020.

2.2 Time Series Decomposition

2.2.1 Hodrick-Prescott (HP) Decomposition

There are two kinds of time series decomposition method: additive and multiplicative decomposition methods, while additive decomposition method is more commonly used, such as HP method (Eqs. 1-2). It is a widely used decomposition method, which decomposes time series into TS and PS. The two subseries are predicted separately, and the final predicted value can be obtained by adding the predicted values of the two subseries. The TS is solved and the PS is obtained by subtracting the TS from the original series.

$$\min \text{Loss} = \min\{\sum_{t=1}^{n}(Y_t - T_t)^2 + \lambda \sum_{t=3}^{n}[(T_t - T_{t-1}) - (T_{t-1} - T_{t-2})]^2\} \qquad (1)$$

$$P = Y - T \qquad (2)$$

where Y, T and P are the original series, the TS and the PS, respectively; t is the time, n is the length of the original series, and λ is the smoothing factor.

2.2.2 Wavelet Transform (WT) Decomposition

In general, the irregular values in time series are noise. In this study, the time series is decomposed into three subseries using Daubechies wavelet (Parmar and Bhardwaj 2015)

transform (Eq. 3) including TS, PS and NS. The time series is decomposed based on the WT, and TS is obtained by filtering the wavelet coefficients through the soft threshold method. The high-frequency component is mostly noise, so after subtracting the TS, the non-common part of the high-frequency component is filtered to get the PS, and then the NS can be obtained.

$$WT_f(s,\tau) = \frac{1}{\sqrt{s}} \int_{-\infty}^{+\infty} f(t)\, \psi\left(\frac{t-\tau}{s}\right) dt \tag{3}$$

$$N = Y - T - P \tag{4}$$

where WT_f, s, t, τ and ψ are the wavelet transform coefficient, the scale, time, deviation and wavelet base, respectively; N is the NS.

2.3 The Coupled Model

2.3.1 Seasonal Autoregressive Integrated Moving Average (SARIMA)

Autoregressive integrated moving average (ARIMA) (Tencaliec et al. 2015; Nguyen 2020) model (Eq. 5) is a statistical machine learning model, which shows good performance in the prediction of time series. However, the model has high requirements for the data stationarity, and the ARIMA model has few parameters, including only autoregressive order, difference order and moving order. As for the time series data with obvious period laws, it is difficult to effectively mine the period relationship between the data. Therefore, SARIMA (Elganainy and Eldwer 2018) model is developed in this study to predict PS. On the basis of ARIMA, SARIMA model adds seasonal regression order, seasonal difference order and seasonal moving order, which has a strong ability to learn period laws, and maximum likelihood estimation (Eq. 6) is applied to solve model parameters.

$$x_t = \theta_1 x_{t-1} + \theta_2 x_{t-2} + \cdots + \theta_p x_{t-p} + \mu_t + \alpha_1 \mu_{t-1} + \alpha_2 \mu_{t-2} + \alpha_q \mu_{t-q} \tag{5}$$

$$L(\lambda) = \prod_{i=1}^{n} f(x_i|\lambda) \quad , lnL(\lambda) = \prod_{i=1}^{n} lnf(x_i|\lambda) \tag{6}$$

where x, θ, μ and α are the values at different time, the autoregressive coefficient, the noise at different time and the moving regression coefficient, respectively; L and λ are likelihood function and parameters, respectively.

2.3.2 Bidirectional Ensemble Learning Long Short Term Memory (BELLSTM)

Long short term memory (LSTM) (Xiang et al. 2020) model (Eqs. 9-10) is an improved RNN model, is a time series model with better performance. The model can store memory that is constantly attenuated (Eqs. 7-8). The output of the network is not entirely dependent on the input of the current time, but is also affected by the output of the previous time and the attenuation memory, and the weight of the recent memory is larger. Considering the non-linear problem of the time series, s*igmoid* and *tanh* (Eqs. 11-12) activation functions are developed to add the non-linear factor. To avoid the problem of abnormally large parameters passing between hidden layers, the hidden layer normalization (Eq. 14) is applied to rectify the output so that the output is rectified to between 0 and 1 before passing to the next layer. However, LSTM is a unidirectional model, and in order to allow the model to learn bidirectionally (Zhang et al. 2018; Chen et al. 2014), bidirectional long short term memory (BLSTM) was developed to improve the learning ability of the model. At present, data-driven models are usually single model or single coupling model. The limitation of this model limits the improvement of model generalization ability. In this study, ensemble learning (EL) (Abbaszadeh and Moradkhani 2019; Qi et al. 2019) was introduced. During each round of training, the connections between some neurons were cut off at random with a probability of 0.1. In this way, the models of each round

were equivalent to different models, and these models were eventually assembled into the final EL model. Meanwhile, genetic algorithm (GA) (Zhang et al. 2019) was developed to optimize the solution process of the model to avoid the occurrence of local optimal solution as much as possible. The model is solved by the least square (LS) method (Eq. 13).

$$\widetilde{C}_t = Tanh(w_c \times [h_{t-1}, x_t] + b_c) \tag{7}$$

$$C_t = f_{t-1} \times C_{t-1} + f_t \times \widetilde{C}_t \tag{8}$$

$$O_t = Sigmoid(w_o \times [h_{t-1}, x_t] + b_o) \tag{9}$$

$$h_t = O_t \times Tanh(C_t) \tag{10}$$

$$Sigmoid(x) = \frac{1}{1+e^{-x}} \tag{11}$$

$$Tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{12}$$

$$minValue_{LS}(\theta) = sum|Y - P|^2 \tag{13}$$

$$norm = \frac{Y_i - Y_{min}}{Y_{max} - Y_{min}} \tag{14}$$

where f, h, x, w, b, $\widetilde{C}_t$, C and O are attenuation factor, the output of the hidden layer, the input at different time, the weight, the bias, the memory of the current time, attenuation memory and the output, respectively; Y, P and θ are measured value vector, predicted value vector and parameters, respectively.

2.3.3 Gaussian Radial Basis Function Ensemble Learning Neural Network (GRBFELNN)

Compared with ANN, radial basis function neural network (RBFNN) (Gholami et al. 2019) is a local parameter adjustment neural network, so the convergence speed of the model is faster. It has been proved that RBFNN is a kind of neural network with simple training and good performance (Bizzi et al. 2018). In this study, Gaussian (Xiang et al. 2020) function was applied

as radial basis function to develop GRBFNN (Eqs. 15-16). Meanwhile, normalization, non-linear activation function, EL and GA were also introduced into the modeling process. The model was also solved by LS.

$$G(\text{x}) = ae^{\frac{-(x-b)^2}{2c^2}} \tag{15}$$

$$f(x) = \sum_{j=1}^{n} w_j G_j(x) + b \tag{16}$$

where G and c are the Gauss function and the standard deviation, respectively; a and b are constants.

2.4 Model Evaluation Standard and Optimization

In this study, six standards were used to validate the performance of the model to avoid the contingency of a single evaluation standard. Mean square error (MSE), Nash-Sutcliffe efficiency (NSE), mean relative error (MRE), mean absolute error (MAE), Pearson correlation coefficient (r) and relative error (RE) (Eqs. 17-22), which respectively represent the actual error, fitting degree, relative deviation degree, absolute deviation degree, correlation degree between measured and predicted values and stability of method. Among them, RE is compared by figure. Considering the high time complexity of the element-by-element iterative calculation, deep learning Tensor was introduced as the data structure in this study, a one-dimensional Tensor is equivalent to a vector, and a multidimensional Tensor is equivalent to a matrix.

$$MSE = \frac{\sum(Y-P)^2}{n} \tag{17}$$

$$NSE = \left\{1 - \left[\frac{MSE(Y,P)}{Var(Y)}\right]\right\} \times 100\% \tag{18}$$

$$MRE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - P_i}{Y_i}\right| \tag{19}$$

$$MAE = \frac{1}{n}\sum|\boldsymbol{Y} - \boldsymbol{P}| \tag{20}$$

$$r = \frac{\sum(\boldsymbol{P}-\bar{\boldsymbol{P}})(\boldsymbol{Y}-\bar{\boldsymbol{Y}})}{\sqrt{\sum(\boldsymbol{Y}-\bar{\boldsymbol{Y}})^2\sum(\boldsymbol{P}-\bar{\boldsymbol{P}})^2}} \tag{21}$$

$$RE = \left|\frac{\boldsymbol{Y}-\boldsymbol{P}}{\boldsymbol{Y}}\right| \tag{22}$$

where *var* and *n* are the variance and the length of data, respectively; $\bar{\boldsymbol{P}}$ and $\bar{\boldsymbol{Y}}$ are the mean of $\boldsymbol{P}$ and the mean of $\boldsymbol{Y}$, respectively.

3 Results

3.1 Time Series Exploration and Decomposition

Auto correlation function (ACF) (Dariane et al. 2018) (Eq. 23) is an effective method to know stationarity of time series, which is helpful for mining the potential characteristics of data. Figure 1 shows the autocorrelation plot of data based on ACF.

$$ACF(k) = \frac{Cov(y_t, y_{t-k})}{Var(y_t)} \tag{23}$$

where *y*, *t*, *k*, *var* and *cov* are the measured value, the time, the number of lag time, the variance and the correlation coefficient.

As can be seen from Fig. 1, the daily water supply series in Shenzhen did not fall into the confidence interval for 50 lag time, indicating that the stationarity of this series was not good enough. Therefore, the time series needs to be decomposed, and the TS, PS and NS are respectively constructed the data-driven model by BELLSTM, SARIMA and GRBFELNN.

3.2 Model Validation

In order to validate the effectiveness of the models, the data in 2020 were taken as the testing set, and the data from January 1, 2015 to December 31, 2019 were divided into the training set and the validation set in a ratio of 8:2.

Table 1 displays the results of the models on the training set, and all the five models have good convergence. In the training stage, the TS gave the better MSE, NSE, MRE, MAE and r. According to the five evaluation standards of PS, the results of WT-PS are obviously superior to that of HP-PS. The evaluation standards of WT-NS are not as good as that of WT-TS and WT-PS due to the acute fluctuation of NS. Table 2 shows the results of the models on the validation set, the standards of the WT-PS are optimal, the evaluation standards of TS are also very good. The validation results of the HP-PS are significantly worse than training results, while the validation results of WT-PS are still excellent. The validation results of WT-TS and WT-NS reveal that BELLSTM and GRBFELNN models have strong learning ability. The validation results of WT-PS show SARIMA model has a strong learning ability for period law.
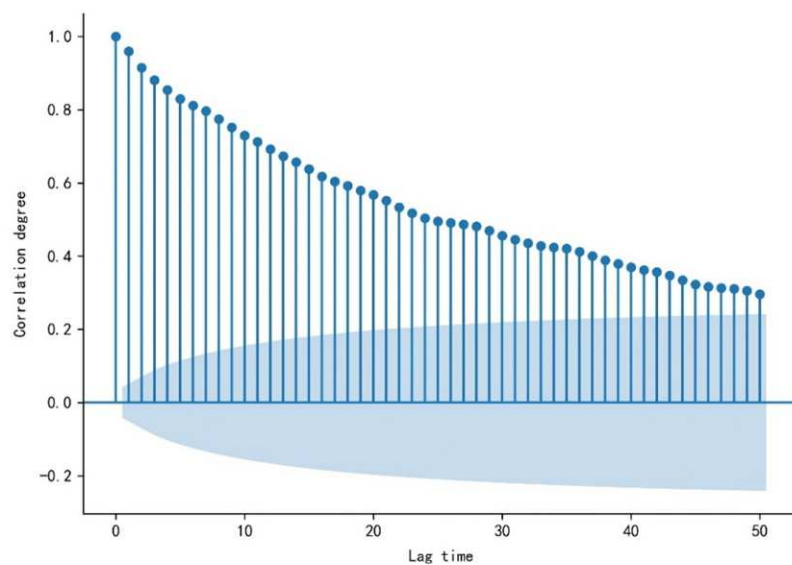
Fig. 1 ACF plot of daily water supply series in Shenzhen, the shaded area corresponds to the 95% confidence interval

Table 1 The training results

Method	MSE/%	NSE/%	MRE	MAE/%	r
BELLSTM -HP-TS	0.04	98.7	0.03	1.48	0.99
SARIMA-HP-PS	0.94	20.58	0.15	7.14	0.61
BELLSTM -WT-TS	0.06	98.07	0.06	1.79	0.99
SARIMA-WT-PS	0.14	99.10	0.01	0.22	0.99
GRBFELNN-WT-NS	0.38	33.09	0.08	4.67	0.57

Table 2 The validation results

Method	MSE/%	NSE/%	MRE	MAE/%	r
BELLSTM -HP-TS	0.05	59.95	0.03	2.29	0.99
SARIMA-HP-PS	2.22	-121.26	0.27	13.09	-0.17
BELLSTM -WT-TS	0.06	80.52	0.03	2.23	0.97
SARIMA-WT-PS	0	100	0	0	0.99
GRBFELNN-WT-NS	0.28	35.69	0.06	3.64	0.59

3.3 The Predicted results

In Table 3, there are certain differences in the predicted results of the subseries. According to the two PS, the predicted results of HP-PS are seriously distorted, the MSE is large and NSE becomes negative, so predicted results are completely unreliable. However, the evaluation standards of WT-PS are the best, which reveals that WT-PS is superior to HP-PS. This is because WT-PS has an obvious period law (Fig. 3), while HP-PS does not show an obvious period law, but shows acute fluctuation so that PS is more like NS (Fig. 2). According to the five evaluation standards of the two TS, the results of the BELLSTM model in the testing set is not different from that of the training set. According to Fig. 2 and 3, the error and deviation degree between the predicted value and the measured value of the TS are small, and the fitting degree and correlation degree are high. So the BELLSTM model shows a strong generalization ability, and the model is close to unbiased prediction. The MSE, NSE, MAE and r of the two TS are similar, but the MRE of WT-TS is 0.79, which is 81% lower than that of HP-TS, so the relative deviation degree of HP-TS is larger that of WS-TS. The results reveal that WS-TS is superior to HP-TS.

Table 3 The predicted results

Method	MSE/%	NSE/%	MRE	MAE/%	r
BELLSTM -HP-TS	0.14	97.93	4.16	2.74	0.99
SARIMA-HP-PS	1.46	-15.56	0.24	9.02	0.06
BELLSTM -WT-TS	0.17	97.39	0.79	2.98	0.99
SARIMA-WT-PS	0	100	0	0	1.0
GRBFELNN-WT-NS	0.6	25.00	0.09	5.1	0.5
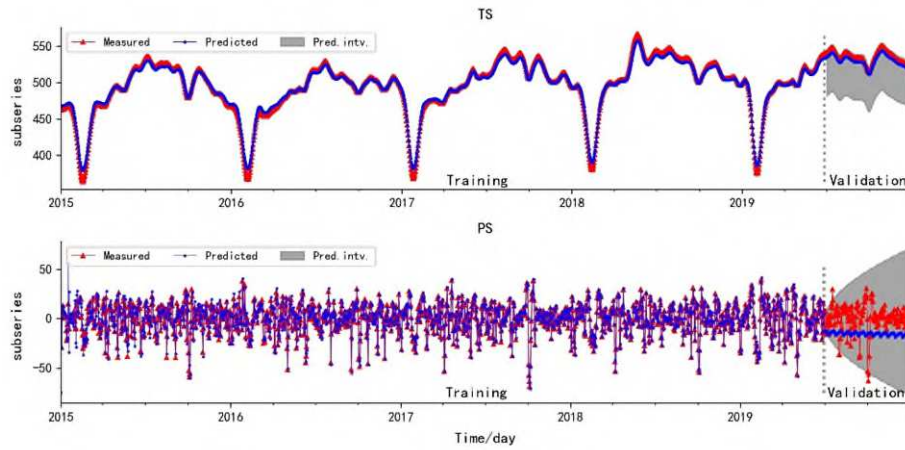
Fig. 2 The training and validation results of HP-TS and HP-PS, the grey area is the prediction interval (pred. intv.)
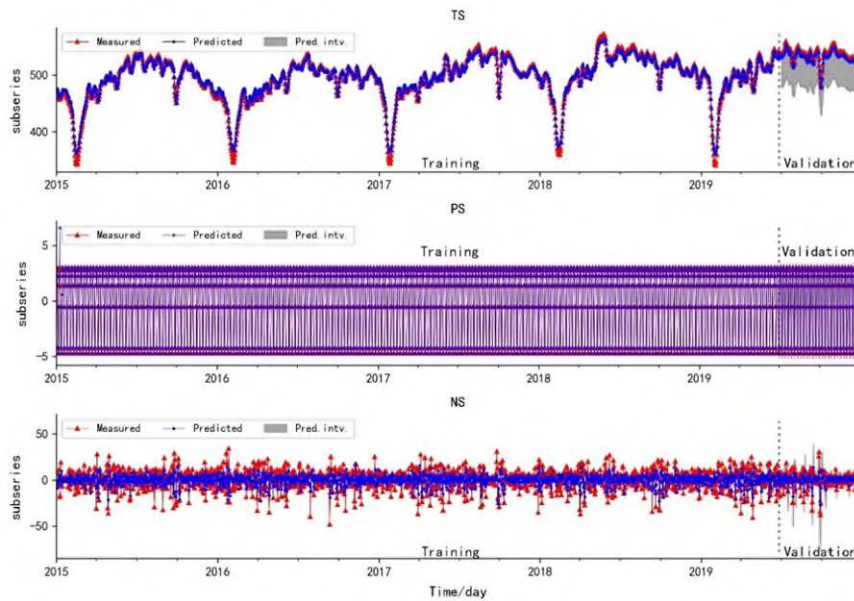
Fig. 3 The training and validation results of WT-TS , WT-PS and WT-NS

For WT-NS, although the NSE and *r* of GRBFELNN model are low, the MSE, MRE and MAE are small, and the error and deviation degree between the predicted value and the measured value is not large (Fig. 3).

After the subseries was reconstructed, the predicted value of WT is better than the predicted value of HP (Fig. 4), the predicted value of WT gave better MSE, NSE, MRE, MAE and r (0.27%, 95.73%, 0.33, 4.01% and 0.98, respectively) (Table 4). The predicted results are close to unbiased prediction and this shows that WT is superior to HP decomposition method. However, February 2020 is the period of corona virus disease 2019 (COVID-19) outbreak, and the floating population cannot return to Shenzhen at the end of the Spring Festival holiday. Shenzhen has only permanent residents, so the variation of water supply law leads to the deviation between the predicted value and the measured value (Fig. 5). Therefore, the prediction interval was estimated in this study based on the student's t-test (T-test) (Hu et al. 2017) (Eq. 24).

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \tag{24}$$

where $\bar{x}$, μ, s and n are the mean value of the measured value, mean value of the predicted value, standard deviation of the measured value and length of measured value, respectively.

According to the historical law of water supply in Shenzhen, the water supply will gradually increase after the Spring Festival holiday and return to the normal level in the Lantern Festival. This variation reveals that although data-driven models can learn potential quantitative relationships from data, they cannot predict variation due to emergency. Therefore, this study performed interval prediction based on T-test.
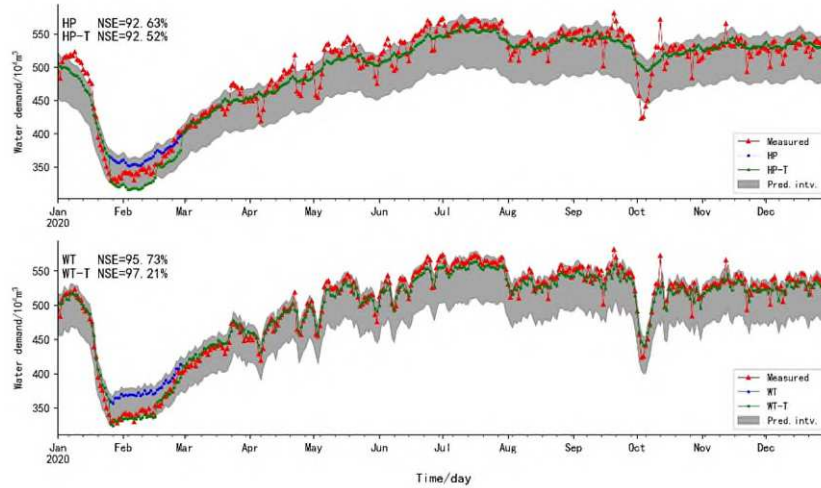
Table 4 Predicted results after reconstruction and rectified results based on T-test

Method	MSE/%	NSE/%	MRE	MAE/%	r
HP	0.46	92.63	0.29	5.25	0.96
HP-T	0.47	92.52	0.19	5.31	0.96
WT	0.27	95.73	0.33	4.01	0.98
WT-T	0.17	97.21	0.1	3.32	0.99

HP-T: Rectified results of HP based on T-test, WT-T: Rectified results of WT based on T-test
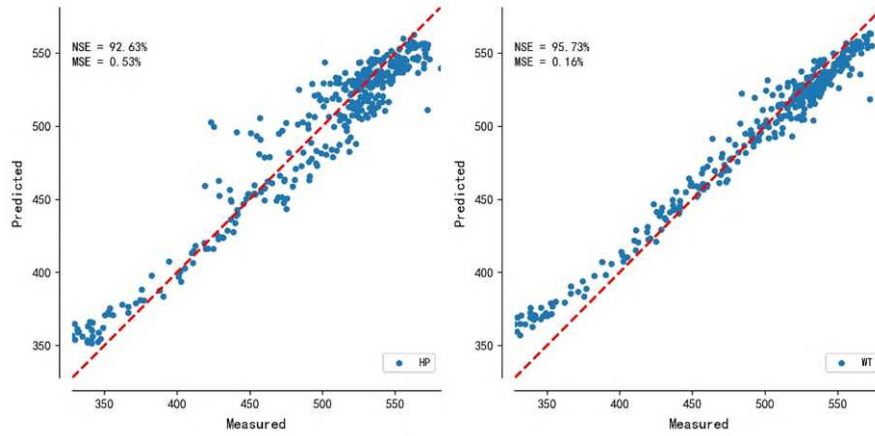
Fig. 4   The distribution plot of predicted value and measured values, the dots on the red line indicate that the predicted value is equal to the measured value, and the closer the dots are to the red line, the smaller the error between the predicted value and the measured value

Fig. 5   Fitting plot of predicted value after reconstruction and measured values, the blue line is the predicted value of the data-driven model, the green line is rectified results based on T-test

For HP, some the measured values did not fall into the predicted interval, and the fluctuation law of water supply was not fully fitted. However the predicted value of WT well fitted the measured value and most measured values fell into the predicted interval. This reveals that although the law of water supply in Shenzhen is affected by the floating population, the fluctuation range of water supply still conforms to the prediction interval. When the country issued the quarantine policy during the Spring Festival holiday, the water supply variation caused by the large number of floating population is predictable, so the lower envelope of the

predicted interval should be used to rectify the predicted value (Fig. 5). The predicted value of WT-T fits well with the measured value, and it gives the best MSE, NSE, MRE, MAE and r (0.17%, 97.21%, 0.1, 3.32% and 0.99, respectively) (Table 4). When the dispatching personnel knows that the water demand will decrease in the future, the water diversion can be reduced and more water diversion quotas can be used in the summer peak to provide support for water supply dispatching. However, the error between the HP-T and the measured value is still large, indicating that HP decomposition method is inferior to WT decomposition method. Finally, the violin plots of different methods are shown to compare the RE distributions (Table 5 and Fig. 6).

Compared with the violin parameters of HP, the confidence interval range and the interquartile range (IQR) of WT are small, and the RE data distribution is relatively dense, indicating that WT is more stable than HP decomposition method. The violin parameters of HP-T are inferior to that of HP, while the violin parameters of WT-T are superior to that of WT. Compared with the violin parameters of WT, the violin parameters of WT-T have smaller confidence interval upper limit (CIUL), upper quartile (UQ), median and lower quartile (LQ). The WT-T gives the best CIUL, UQ, median, LQ and confidence interval lower limit (CILL) (4.34, 2.23, 1.46, 0.82 and 0, respectively) (Table 5). The violin plots of HP and HP-T are almost the same, while the violin plot of WT-T is significantly better than that of WT. The RE data distribution of WT-T is denser (Fig. 6).

4 Discussion

In order to further validate the effectiveness of this research method, the entire data set was

divided into training set and test set according to 7:3 for cross-validation (Table 6).

In Table 6, the predicted results of cross-validation are similar to that of Table 3. The predicted results of WT-PS are superior to that of HP-PS. Although the predicted results of the two TS are still good, but MRE of HP-TS is higher. Compared with the MRE of HP-TS, the MRE of WT-TS is reduced by 80.43%, which shows that the relative deviation degree between the predicted value of WT-TS and the measured value is smaller. Because the period law of HP-PS is not obvious, the predicted result of SARIMA model is poor. According to the predicted results of WT-NS, the error and deviation degree between the predicted value and the measured value are small, which shows that GRBFELNN model has strong generalization ability.

Similar to the results in Fig. 5, the fitting degree of HP is still not as good as that of WT (Fig. 7). However, compared with Fig. 5, the predicted interval of HP in cross-validation is better. This shows that the prediction interval of HP decomposition method is affected by data length, while the prediction interval of WT decomposition method is not affected by data length, so WT is more reliable than HP.

Table 5 The violin parameters

Method	CIUL	UQ	Median	LQ	CILL
HP	7.97	3.85	2.19	1.10	0
HP-T	8.33	3.99	2.25	1.10	0
WT	5.22	2.62	1.55	0.88	0
WT-T	4.34	2.23	1.46	0.82	0

Table 6 The cross-validation results

| Method	MSE/%	NSE/%	MRE	MAE/%	r
BELLSTM -HP-TS	0.1	97.78	2.35	2.37	0.99
SARIMA-HP-PS	1.42	-13.97	0.24	8.45	0.01
BELLSTM -WT-TS	0.12	97.25	0.46	2.55	0.99
SARIMA-WT-PS	0	100	0	0	1.0
GRBFELNN-WT-NS	0.49	30.97	0.08	4.76	0.55
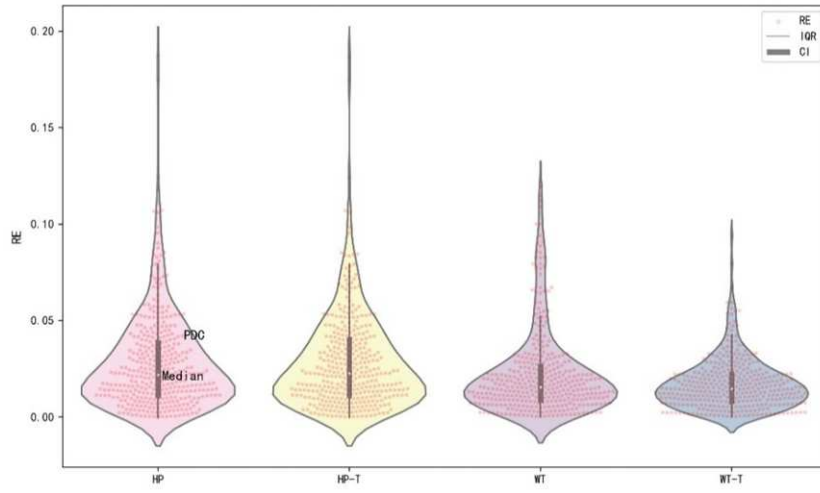
Fig. 6 The violin plots of different methods, the outer curve is the probability density curve (PDC), the red dots are RE data, the thin line is the 95% confidence interval (CI), the thick line is IQR, the white dot is the median
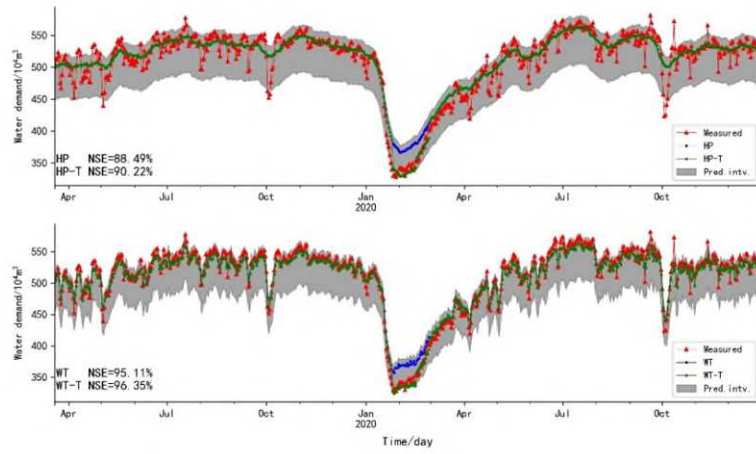
Fig. 7 The fitting plot of predicted value after reconstruction and measured values in cross-validation

As can be seen from box plots (Fig. 8), the box height of HP-T is slightly decreased compared to the box height of HP, but the RE distribution is almost same. This is because the NSE between predicted and measured values of HP is not good enough. Compared with box plot of WT, the box height of WT-T decreases obviously, the maximum value of RE decreases obviously, and the RE distribution is denser, with most dots distributed in the interval [0,0.05]. In conclusion, WT decomposition method is superior to HP decomposition method.

The error of HP is larger than that of WT mainly because of the acute fluctuation of HP-PS, which makes it look more like noise. The above studies have shown that the GRBFELNN model
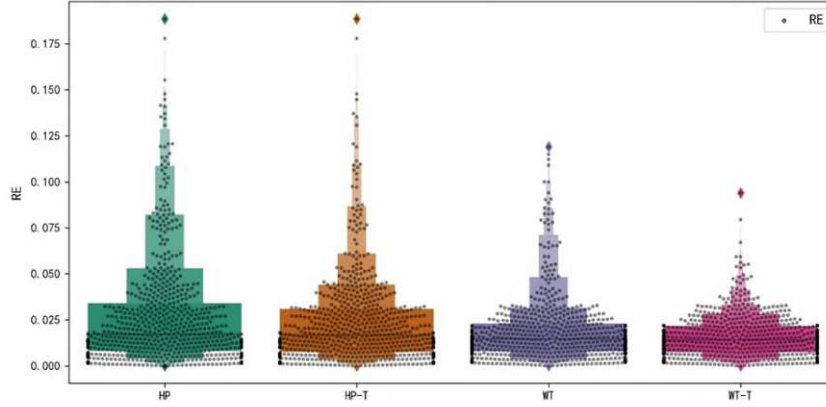
Fig. 8 The box plots of the methods, the box number in the shaded area represents the density of RE data, the smaller the box height, the denser the RE data,

has a strong modeling ability for NS. So the GRBFELNN model was used to construct data-driven model for HP-PS (Table 7).

Table 7 sums up the reconstruction results of different HP-PS modeling. The five evaluation standards of BELLSTM-GRBFELNN-T-test (B-G-T) are all better than BELLSTM-SARIMA-T-test (B-S-T), which shows that GRBFELNN model has better learning ability than SARIMA model for acutely fluctuating sequence. However, compared with the predicted results of WT-T (Table 4), the prediction accuracy of B-G-T is inferior to that of WT-T, indicating that WT decomposition method is superior to HP.

In addition, through a large number of experiments and analysis, it is found that the WT-PS is relatively stable, while increasing the maximum decomposition level (MDL) of WT has little effect on the WT-PS, but has a large effect on the WT-TS. In this study, the MDL was increased to decompose time series again (WT2) and compared with the results of the original MDL (WT1).

The trend of the WT2-TS is too smooth, the wave peak and wave trough are weakened to a certain extent, while the WT2-NS fluctuates more acutely and is stretched along the vertical

axis, and the maximum value of the WT2-NS is expanded (Fig. 9). Although the prediction accuracy of the WT2-TS is improved, the prediction accuracy of the WT2-NS is significantly decreased, resulting in the increase of the error after reconstruction (Table 8). As the predicted results of WT1 and WT2 are not much different, this error bar plot was applied to compare their predicted results (Fig. 10).

Compared with error bar of WT1, the error bar of WT2 becomes longer, and the error bar increases obviously at some moments. Therefore, WT decomposition should not set MDL too high. Based on repeated experiments, it is found that the optimal MDL is 3 or 4, so as to ensure the obvious trend of the TS and reduce the acute fluctuation of the NS.

Table 7 The reconstruction results of different HP-PS modeling

Method	MSE/%	NSE/%	MRE	MAE/%	r
B-S-T	0.47	92.52	0.19	5.31	0.96
B-G-T	0.28	95.53	0.11	4.12	0.98

Table 8 The comparison of reconstruction results of WT1-T and WT2-T

Method	MSE/%	NSE/%	MRE	MAE/%	r
WT1-T	0.17	97.21	0.1	3.32	0.98
WT2-T	0.33	94.73	0.12	4.46	0.97

WT1-T: Rectified results of WT1 based on T-test, WT2-T: Rectified results of WT2 based on T-test
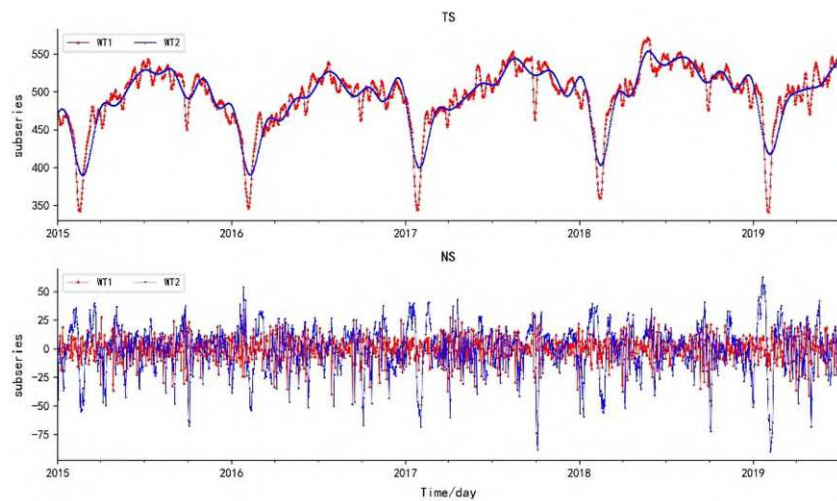
Fig. 9 The comparison of TS and NS of WT1 and WT2
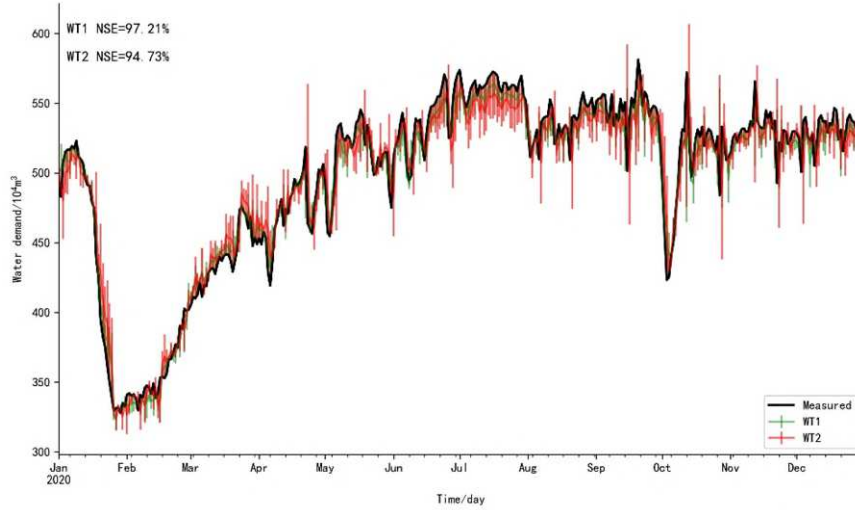
Fig. 10   The error bar plot of predicted value of WT1 and WT2

5 Conclusions

In this paper, through a lot of research and experiment analysis, it is found that the non-stationarity of time series is a major bottleneck for data-driven model. If we want to effectively carry out modeling of time series, the potential laws and stationarity of time series need to be known. ACF is an effective method to mine the potential laws of time series, which can provide some support for modeling. Based on the results of ACF, the time series decomposition method based on HP and WT was proposed in the study, and the BELLSTM, SARIMA and GRBFELNN models were developed for modeling of TS, PS and NS, respectively. Non-linear activation functions were used to increase the non-linear factors, and normalization between hidden layers was applied to rectify the output. At the same time, EL was introduced to increase the generalization ability of the models, and GA was developed to optimize the solution process of the model, and interval prediction based on T-test was carried out. The research methods were applied to daily water demand prediction in Shenzhen and cross-validation was also performed. Multiple statistical figures such as violin plot, box plot and error bar plot were used

to display RE data distribution and compare their results.

The results show that WT decomposition method is superior to HP decomposition method. Although there was a variation in the water supply law in February 2020, its distribution still conforms to the prediction interval of WT. WT-T gives the best MSE, NSE, MRE, MAE and r (0.17%, 97.21%, 0.1, 3.32% and 0.98, respectively). The fitting degree and correlation degree between the predicted value of WT-T and the measured value are the highest, and the error and deviation degree are the lowest, which is close to the unbiased prediction. According to violin plot, it is found that the RE distribution of WT is better, WT-T gives the best CIUL, UQ, median, LQ and CILL (4.34, 2.23, 1.46, 0.82 and 0, respectively), so WT is more stable than HP decomposition method. In cross-validation, the results of WT-T are still the best, indicating that WT was more reliable than HP.

For the results of the subseries, BELLSTM model can predict the TS with high accuracy. SARIMA model has good predicted results for WT-PS, but poor predicted results for HP-PS. GRBFELNN model shows a strong generalization ability in the predicted results of WT-NS, and B-G-T is superior to B-S-T, which reveals that GRBFELNN model has a stronger learning ability for fluctuation sequence. MDL has less effect on the WT-PS, but more effect on the WT-TS. By comparing the influence of MDL on the predicted results, it is found that MDL is the best when set to 3 or 4. Too high MDL will weaken the trend characteristics of TS, aggravate the fluctuation degree of NS, and the prediction accuracy will be reduced.

References

Abbaszadeh, P., Moradkhani, H., & Zhan, X. (2019). Downscaling SMAP radiometer soil

moisture over the CONUS using an ensemble learning method. Water Resources Research, 55(1), 324-344. https://doi.org/10.1029/2018WR023354

Akbari, M., Salmasi, F., Arvanaghi, H., Karbasi, M., & Farsadizadeh, D. (2019). Application of Gaussian process regression model to predict discharge coefficient of Gated Piano Key Weir. Water Resources Management, 33(11), 3929-3947. https://doi.org/10.1007/s11269-019-02343-3

Al-Zahrani, M. A., & Abo-Monasar, A. (2015). Urban residential water demand prediction based on artificial neural networks and time series models. Water resources management, 29(10), 3651-3662. https://doi.org/10.1007/s11269-015-1021-z

Amaranto, A., Munoz-Arriola, F., Solomatine, D. P., & Corzo, G. (2019). A spatially enhanced data-driven multimodel to improve semiseasonal groundwater forecasts in the High Plains aquifer, USA. Water Resources Research, 55(7), 5941-5961. https://doi.org/10.1029/2018WR024301

Baek, D., Seo, I. W., Kim, J. S., & Nelson, J. M. (2019). UAV-based measurements of spatio-temporal concentration distributions of fluorescent tracers in open channel flows. Advances in Water Resources, 127, 76-88. https://doi.org/10.1016/j.advwatres.2019.03.007

Bassiouni, M., Vogel, R. M., & Archfield, S. A. (2016). Panel regressions to estimate low‐flow response to rainfall variability in ungaged basins. Water Resources Research, 52(12), 9470-9494. https://doi.org/10.1002/2016WR018718

Bizzi, S., Cominola, A., Mason, E., Castelletti, A., & Paik, K. (2018). Multicriteria Optimization Model to Generate on‐DEM Optimal Channel Networks. Water Resources

Research, 54(8), 5727-5740. https://doi.org/10.1029/2018WR022977

Chen, K., Wang, G., Chen, J., Yuan, S., & Wei, G. (2020). Impact of climate changes on manufacturing: Hodrick-Prescott filtering and a partial least squares regression model. International Journal of Computational Science and Engineering, 22(2-3), 211-220. https://doi.org/10.1504/IJCSE.2020.107343

Chen, L. J., Feng, Q., Li, F. R., & Li, C. S. (2014). A bidirectional model for simulating soil water flow and salt transport under mulched drip irrigation with saline water. Agricultural Water Management, 146, 24-33. https://doi.org/10.1016/j.agwat.2014.07.021

Dariane, A. B., Farhani, M., & Azimi, S. (2018). Long term streamflow forecasting using a hybrid entropy model. Water resources management, 32(4), 1439-1451. https://doi.org/10.1007/s11269-017-1878-0

Elganainy, M. A., & Eldwer, A. E. (2018). Stochastic forecasting models of the monthly streamflow for the Blue Nile at Eldiem Station. Water Resources, 45(3), 326-337. https://doi.org/10.1134/S0097807818030041

Fu, J., Zhong, P. A., Chen, J., Xu, B., Zhu, F., & Zhang, Y. (2019). Water Resources Allocation in Transboundary River Basins Based on a Game Model Considering Inflow Forecasting Errors. Water Resources Management, 33(8), 2809-2825. https://doi.org/10.1007/s11269-019-02259-y

Gholami, A., Bonakdari, H., Zaji, A. H., & Akhtari, A. A. (2019). An efficient classified radial basis neural network for prediction of flow variables in sharp open-channel bends. Applied Water Science, 9(6), 1-17. https://doi.org/10.1007/s13201-019-1020-y

Hadi, S. J., & Tombul, M. (2018). Streamflow forecasting using four wavelet transformation

combinations approaches with data-driven models: a comparative study. Water Resources Management, 32(14), 4661-4679. https://doi.org/10.1007/s11269-018-2077-3

Haque, M. M., de Souza, A., & Rahman, A. (2017). Water demand modelling using independent component regression technique. Water Resources Management, 31(1), 299-312. https://doi.org/10.1007/s11269-016-1525-1

Hu, H., Jiang, C., Ma, H., Ding, L., Geng, J., Xu, K., ... & Ren, H. (2017). Removal characteristics of DON in pharmaceutical wastewater and its influence on the N-nitrosodimethylamine formation potential and acute toxicity of DOM. Water research, 109, 114-121. https://doi.org/10.1016/j.watres.2016.10.010

Huang, C., Wang, G., Zheng, X., Yu, J., & Xu, X. (2015). Simple linear modeling approach for linking hydrological model parameters to the physical features of a river basin. Water Resources Management, 29(9), 3265-3289. https://doi.org/10.1007/s11269-015-0996-9

Huang, J., Zhang, Y., Arhonditsis, G. B., Gao, J., Chen, Q., & Peng, J. (2020). The magnitude and drivers of harmful algal blooms in China's lakes and reservoirs: A national-scale characterization. Water Research, 181, 115902. https://doi.org/10.1016/j.watres.2020.115902

Kajewska-Szkudlarek, J. (2020). Clustering approach to urban rainfall time series prediction with support vector regression model. Urban Water Journal, 17(3), 235-246. https://doi.org/10.1080/1573062X.2020.1760319

Khan, U. T., He, J., & Valeo, C. (2018). River flood prediction using fuzzy neural networks: an investigation on automated network architecture. Water Science and Technology, 2017(1), 238-247. https://doi.org/10.2166/wst.2018.107

Li, C., Bai, Y., & Zeng, B. (2016). Deep feature learning architectures for daily reservoir inflow forecasting. Water Resources Management, 30(14), 5145-5161. https://doi.org/10.1007/s11269-016-1474-8

Li, X., & Willems, P. (2020). A hybrid model for fast and probabilistic urban pluvial flood prediction. Water Resources Research, 56(6), e2019WR025128. https://doi.org/10.1029/2019WR025128

Ma, J., Ding, Y., Cheng, J. C., Jiang, F., & Xu, Z. (2020). Soft detection of 5-day BOD with sparse matrix in city harbor water using deep learning techniques. Water research, 170, 115350. https://doi.org/10.1016/j.watres.2019.115350

Mahmoudi, N., Orouji, H., & Fallah-Mehdipour, E. (2016). Integration of shuffled frog leaping algorithm and support vector regression for prediction of water quality parameters. Water resources management, 30(7), 2195-2211. https://doi.org/10.1007/s11269-016-1280-3

Mehdizadeh, S. (2020). Using AR, MA, and ARMA time series models to improve the performance of MARS and KNN approaches in monthly precipitation modeling under limited climatic data. Water Resources Management, 34(1), 263-282. https://doi.org/10.1007/s11269-019-02442-1

Nguyen, X. H. (2020). Combining statistical machine learning models with ARIMA for water level forecasting: The case of the Red river. Advances in Water Resources, 142, 103656. https://doi.org/10.1016/j.advwatres.2020.103656

Pan, B., Hsu, K., AghaKouchak, A., & Sorooshian, S. (2019). Improving precipitation estimation using convolutional neural network. Water Resources Research, 55(3), 2301-2321. https://doi.org/10.1029/2018WR024090

Parmar, K. S., & Bhardwaj, R. (2015). River water prediction modeling using neural networks, fuzzy and wavelet coupled model. Water resources management, 29(1), 17-33. https://doi.org/10.1007/s11269-014-0824-7

Pathiraja, S., Moradkhani, H., Marshall, L., Sharma, A., & Geenens, G. (2018). Data‐driven model uncertainty estimation in hydrologic data assimilation. Water resources research, 54(2), 1252-1280. https://doi.org/10.1002/2018WR022627

Piotrowski, A. P., & Napiorkowski, J. J. (2013). A comparison of methods to avoid overfitting in neural networks training in the case of catchment runoff modelling. Journal of Hydrology, 476, 97-111. https://doi.org/10.1016/j.jhydrol.2012.10.019

Poul, A. K., Shourian, M., & Ebrahimi, H. (2019). A comparative study of MLR, KNN, ANN and ANFIS models with wavelet transform in monthly stream flow prediction. Water Resources Management, 33(8), 2907-2923. https://doi.org/10.1007/s11269-019-02273-0

Qi, Y., Zhou, Z., Yang, L., Quan, Y., & Miao, Q. (2019). A decomposition-ensemble learning model based on LSTM neural network for daily reservoir inflow forecasting. Water Resources Management, 33(12), 4123-4139. https://doi.org/10.1007/s11269-019-02345-1

Sari, V., dos Reis Castro, N. M., & Pedrollo, O. C. (2017). Estimate of suspended sediment concentration from monitored data of turbidity and water level using artificial neural networks. Water resources management, 31(15), 4909-4923. https://doi.org/10.1007/s11269-017-1785-4

Serinaldi, F., Kilsby, C. G., & Lombardo, F. (2018). Untenable nonstationarity: An assessment of the fitness for purpose of trend tests in hydrology. Advances in Water Resources, 111, 132-155. https://doi.org/10.1016/j.advwatres.2017.10.015

Sun, P., Jiang, Z. Q., Wang, T. T., & Zhang, Y. K. (2016). Research and application of parallel normal cloud mutation shuffled frog leaping algorithm in cascade reservoirs optimal operation. Water resources management, 30(3), 1019-1035. https://doi.org/10.1007/s11269-015-1208-3

Tencaliec, P., Favre, A. C., Prieur, C., & Mathevet, T. (2015). Reconstruction of missing daily streamflow data using dynamic regression models. Water Resources Research, 51(12), 9447-9463. https://doi.org/10.1002/2015WR017399

Wright, A. J., Walker, J. P., & Pauwels, V. R. (2017). Estimating rainfall time series and model parameter distributions using model data reduction and inversion techniques. Water Resources Research, 53(8), 6407-6424. https://doi.org/10.1002/2017WR020442

Xiang, Z., Yan, J., & Demir, I. (2020). A rainfall‐runoff model with LSTM‐based sequence‐to‐sequence learning. Water resources research, 56(1), e2019WR025326. https://doi.org/10.1029/2019WR025326

Xiang, Z., Yan, J., & Demir, I. (2020). A rainfall‐runoff model with LSTM‐based sequence‐to‐sequence learning. Water resources research, 56(1), e2019WR025326. https://doi.org/10.1029/2019WR025326

Yang, L., Li, J., Kang, A., Li, S., & Feng, P. (2020). The Effect of Nonstationarity in Rainfall on Urban Flooding Based on Coupling SWMM and MIKE21. Water Resources Management, 1-17. https://doi.org/10.1007/s11269-020-02522-7

Zarei, A. R., & Mahmoudi, M. R. (2020). Ability Assessment of the Stationary and Cyclostationary Time Series Models to Predict Drought Indices. Water Resources Management, 34(15), 5009-5029. https://doi.org/10.1007/s11269-020-02710-5

Zhang, X., Bao, W., Liang, W., & Shen, D. (2018). A variable parameter bidirectional stage routing model for tidal rivers with lateral inflow. Journal of Hydrology, 564, 1036-1047. https://doi.org/10.1016/j.jhydrol.2018.07.065

Zhang, Y., Gao, X., Smith, K., Inial, G., Liu, S., Conil, L. B., & Pan, B. (2019). Integrating water quality and operation into prediction of water production in drinking water treatment plants by genetic algorithm enhanced artificial neural network. Water research, 164, 114888. https://doi.org/10.1016/j.watres.2019.114888

Zhong, Z., Sun, A. Y., & Jeong, H. (2019). Predicting co2 plume migration in heterogeneous formations using conditional deep convolutional generative adversarial network. Water Resources Research, 55(7), 5830-5851. https://doi.org/10.1029/2018WR024592

Affiliations

Xin Liu[1,2] · Xuefeng Sang[2] · Jiaxuan Chang[2] · Yang Zheng[2]

Xin Liu

xinliudoc@outlook.com

1 School of Water Conservancy, North China University of Water Resources and Electric Power, Zhengzhou 450046, China

2 Research Office for Water Resources Management, China Institute of Water Resources and Hydropower Research, Beijing 100038, China

Figures
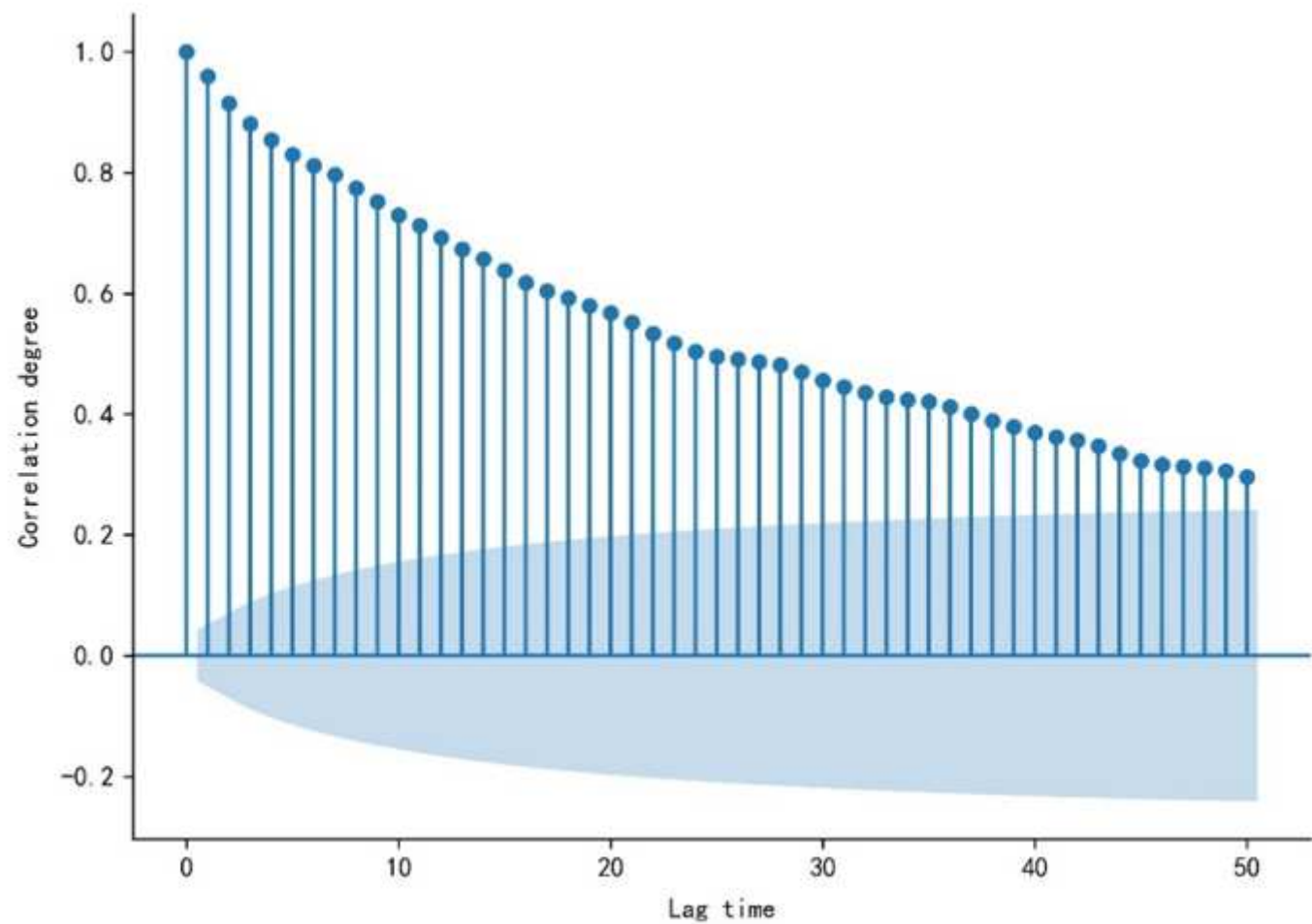
Figure 1

ACF plot of daily water supply series in Shenzhen, the shaded area corresponds to the 95% confidence interval
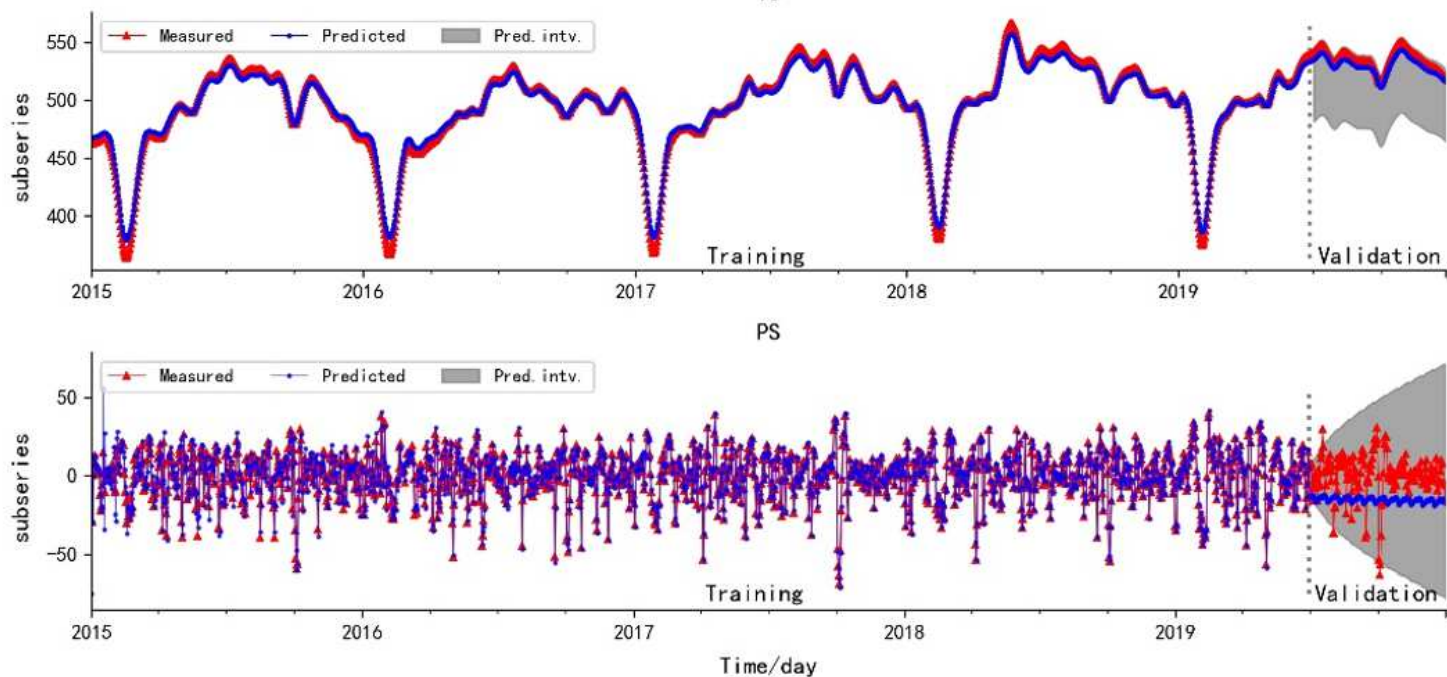
Figure 2

The training and validation results of HP-TS and HP-PS, the grey area is the prediction interval (pred. intv.)
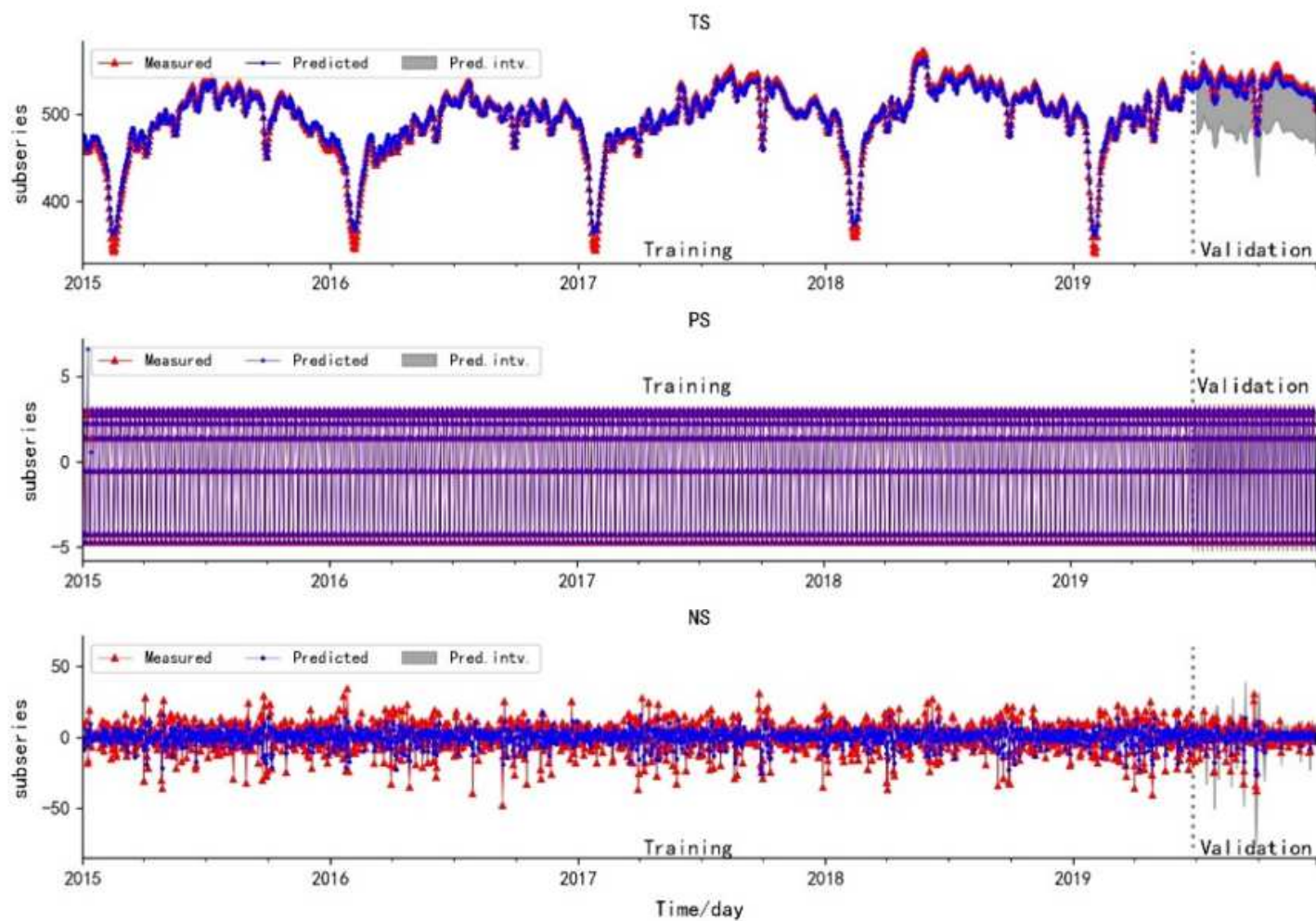
Figure 3

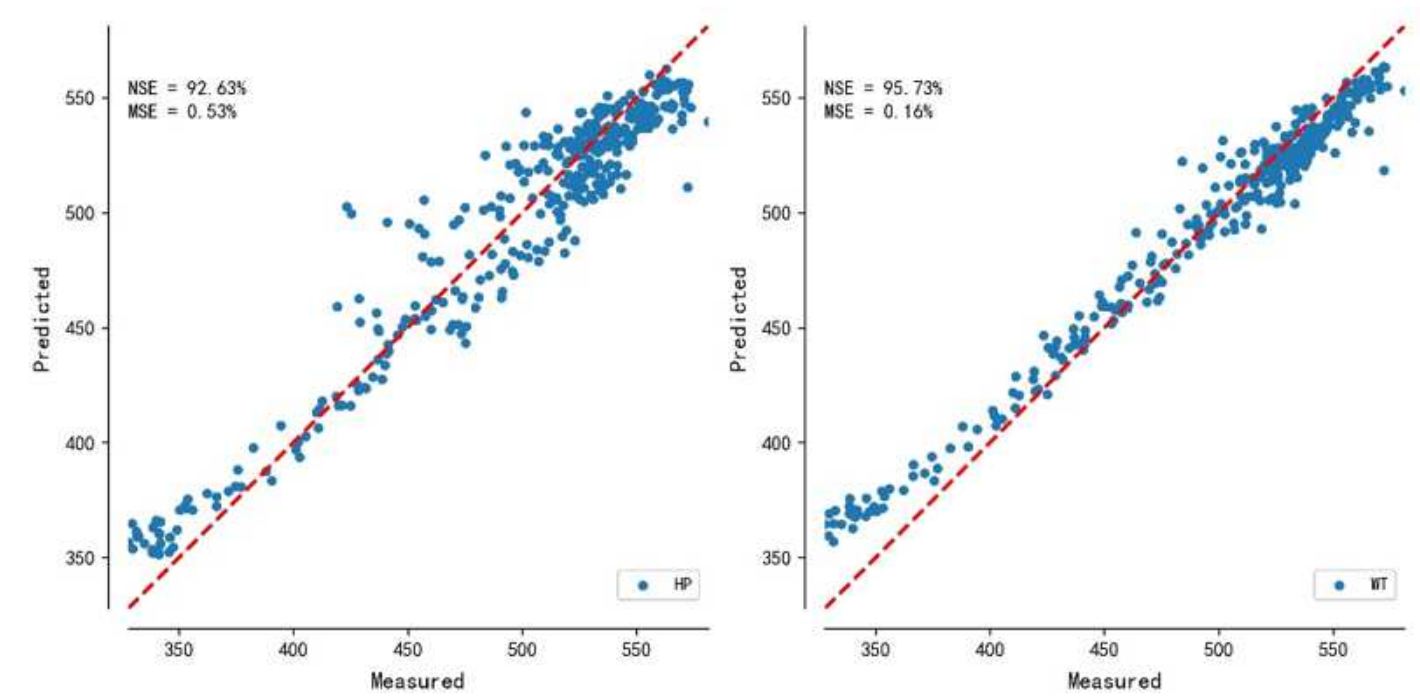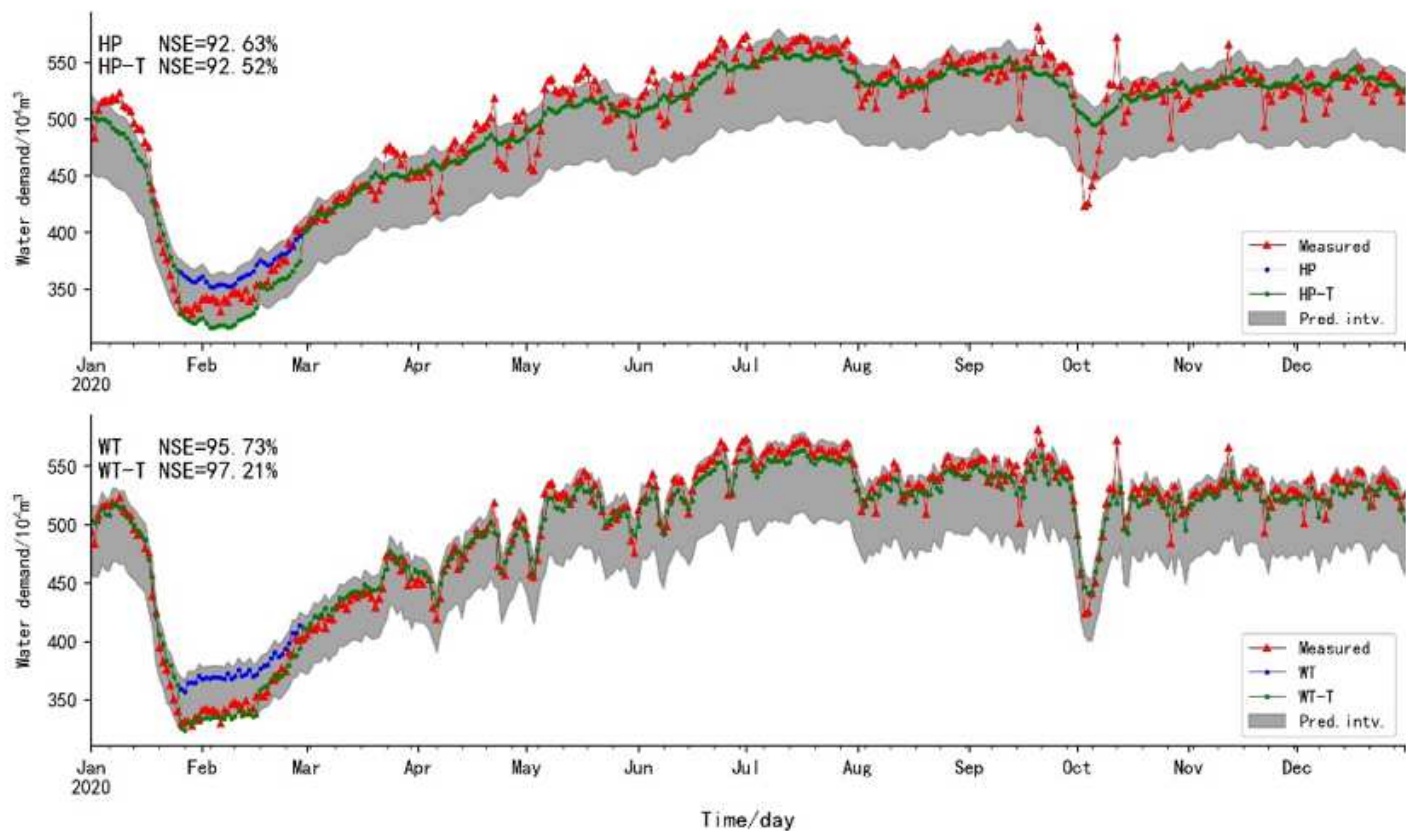The training and validation results of WT-TS , WT-PS and WT-NS

Figure 4

The distribution plot of predicted value and measured values, the dots on the red line indicate that the predicted value is equal to the measured value, and the closer the dots are to the red line, the smaller the error between the predicted value and the measured value

Figure 5

Fitting plot of predicted value after reconstruction and measured values, the blue line is the predicted value of the data-driven model, the green line is rectified results based on T-test
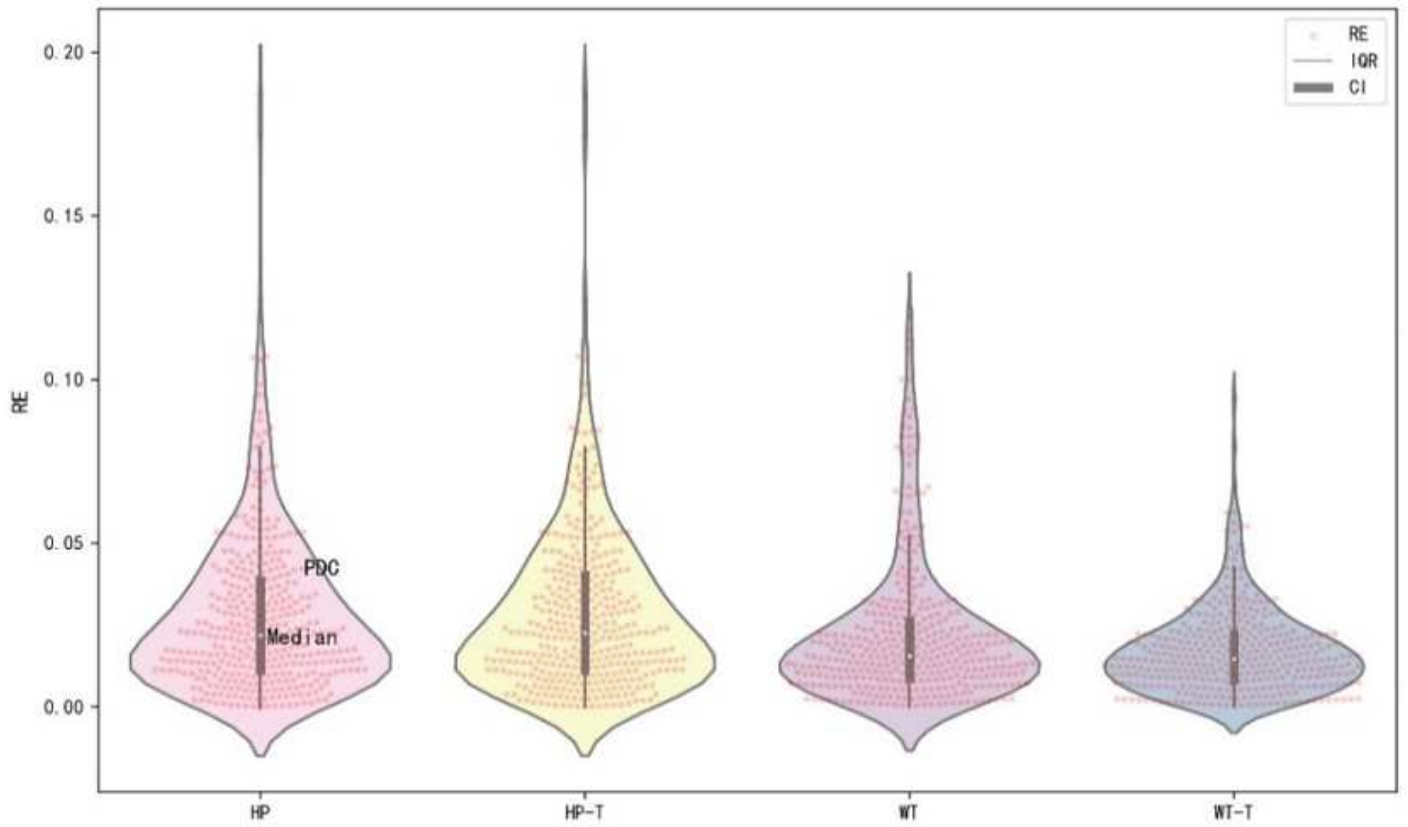
Figure 6

The violin plots of different methods, the outer curve is the probability density curve (PDC), the red dots are RE data, the thin line is the 95% confidence interval (CI), the thick line is IQR, the white dot is the median
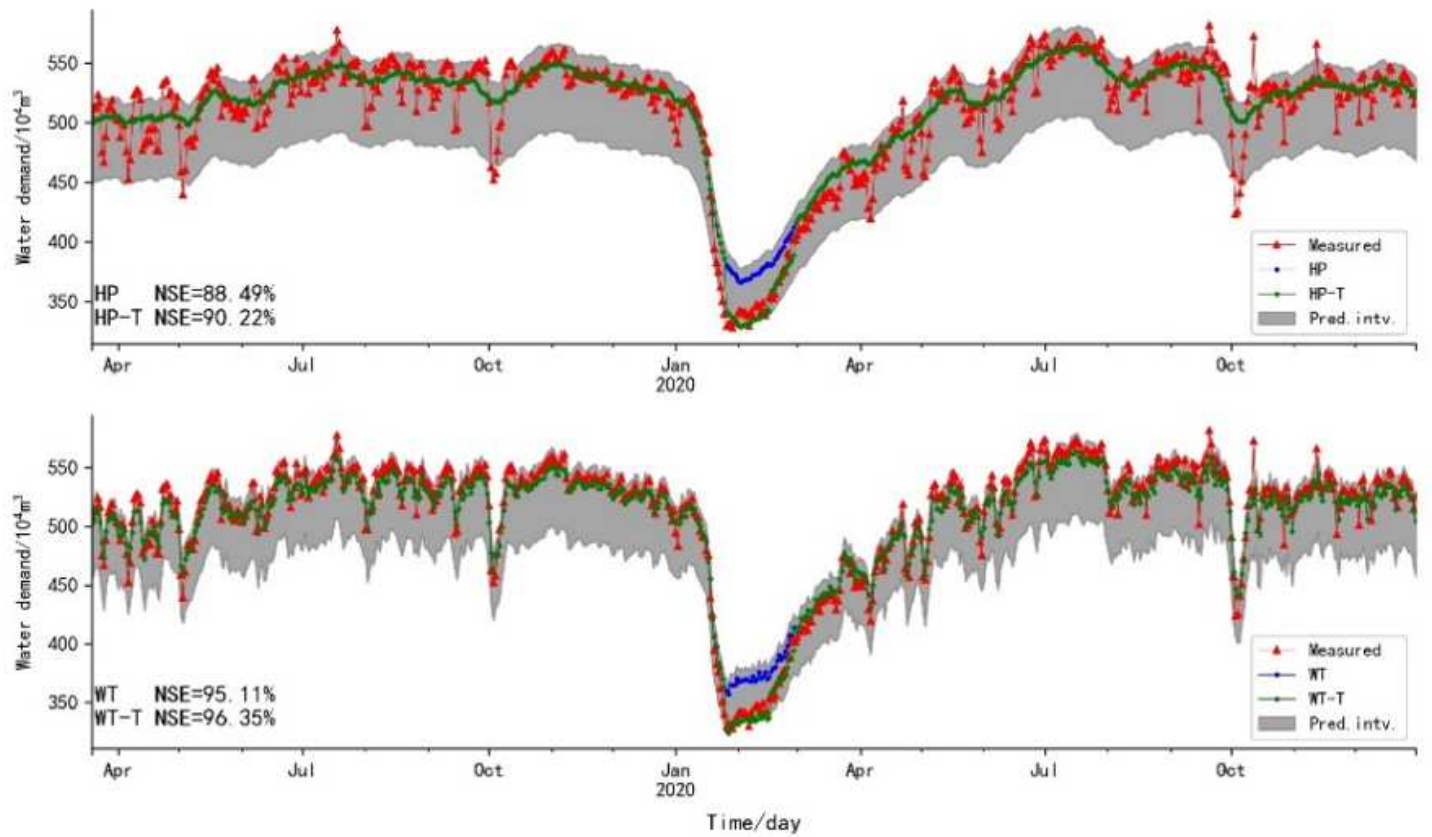
Figure 7

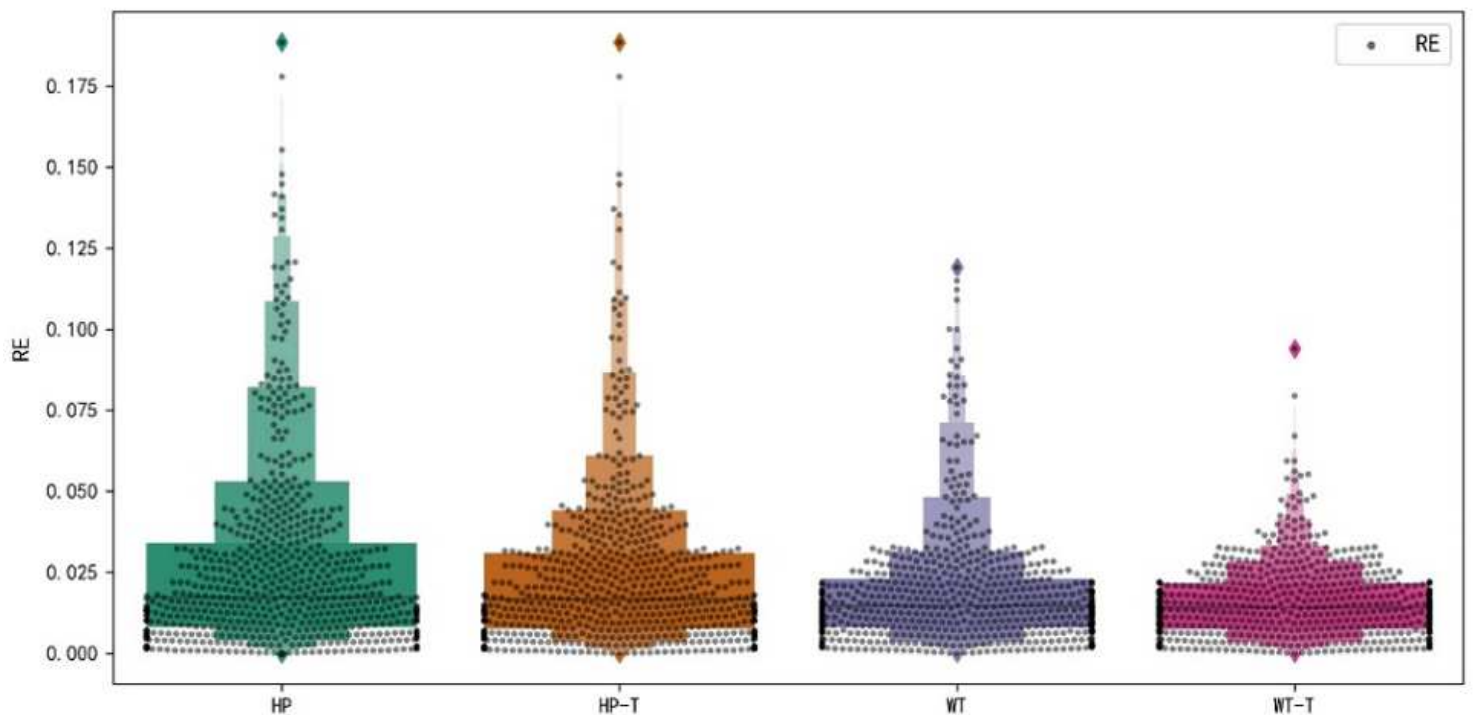The fitting plot of predicted value after reconstruction and measured values in cross-validation

Figure 8

The box plots of the methods, the box number in the shaded area represents the density of RE data, the smaller the box height, the denser the RE data,
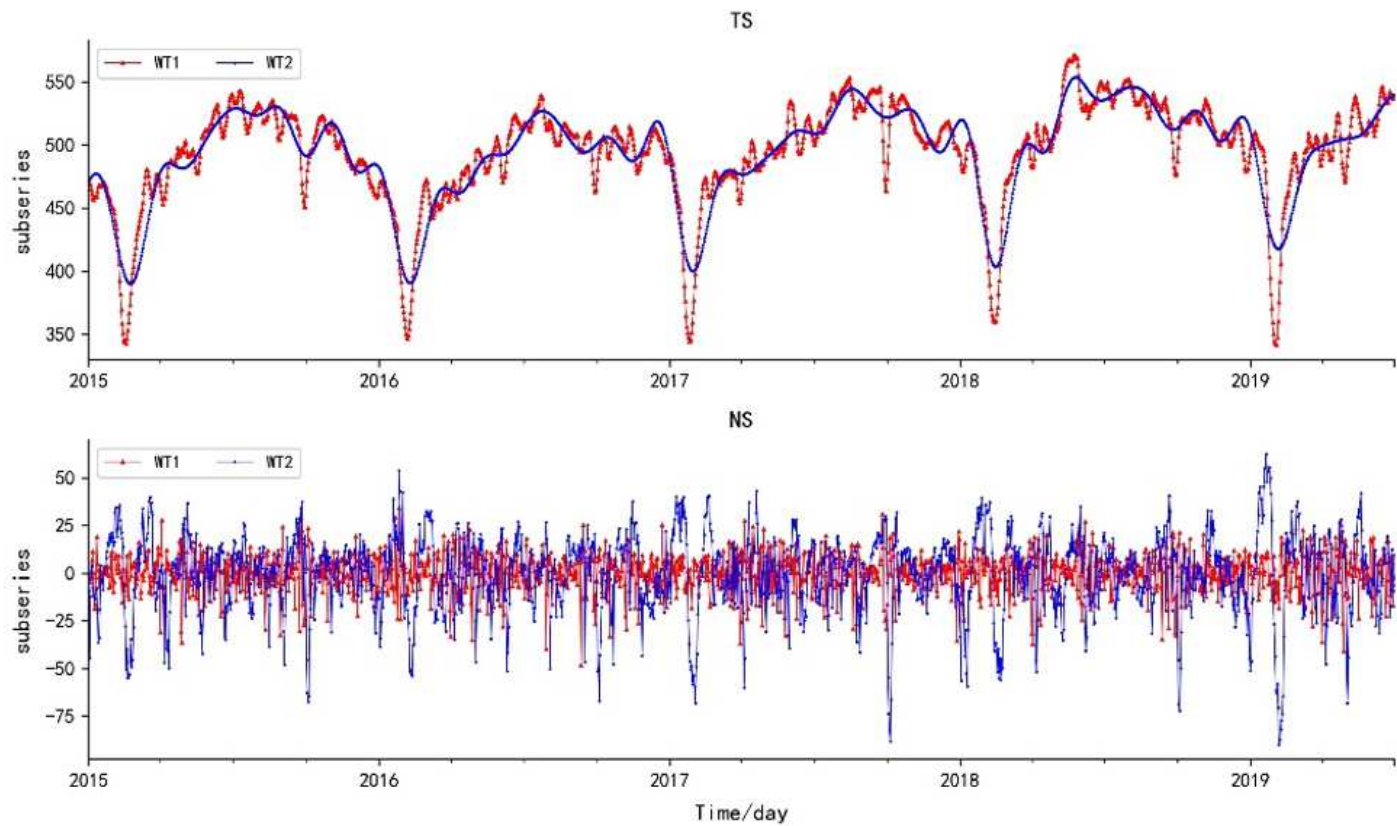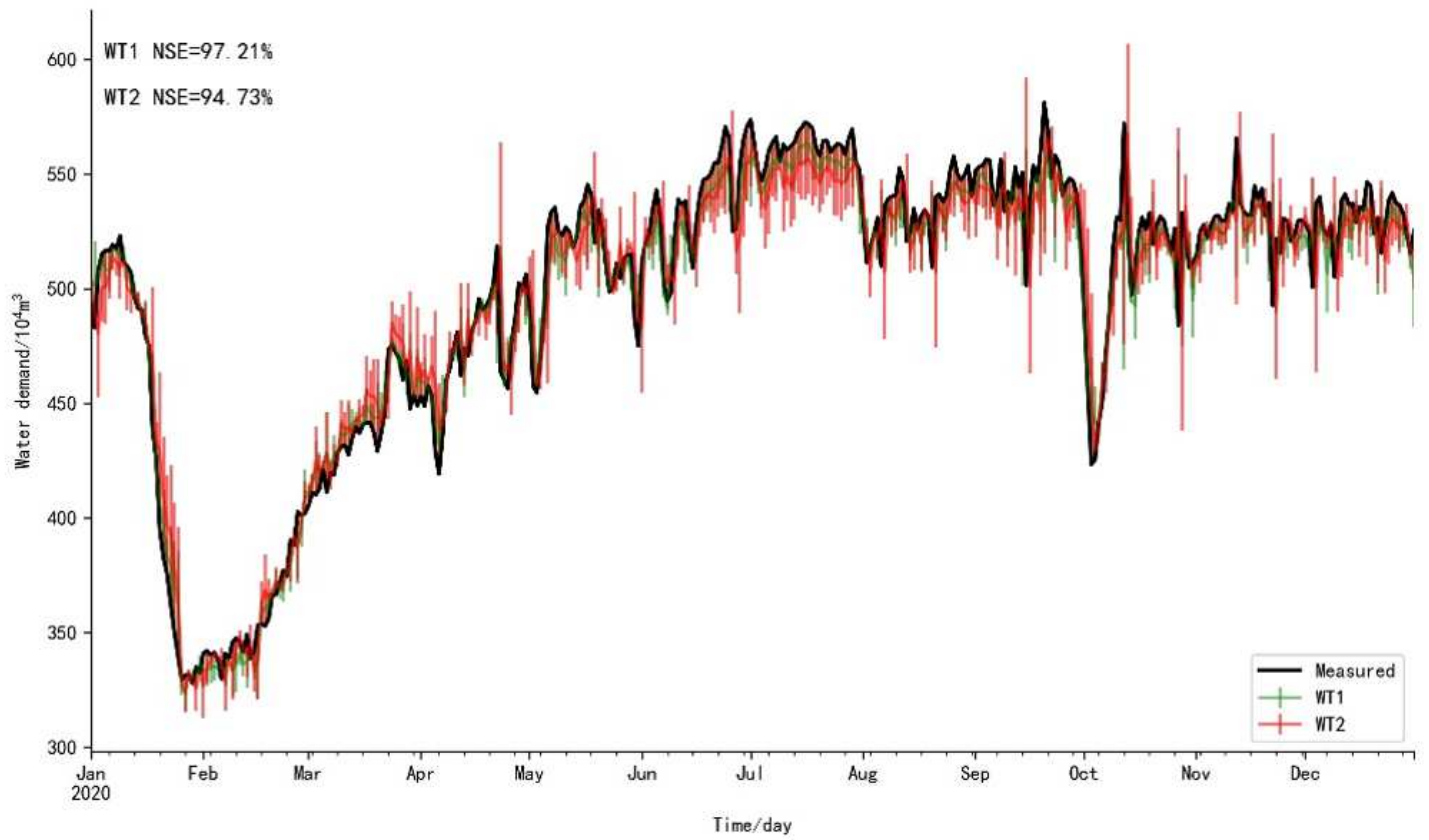

Figure 9

The comparison of TS and NS of WT1 and WT2

Figure 10

The error bar plot of predicted value of WT1 and WT2