IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received March 3, 2020, accepted March 17, 2020, date of publication March 23, 2020, date of current version April 8, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2982433

Water Level Prediction Model Based on GRU and CNN

MINGYANG PAN[1], HAINAN ZHOU[1], JIAYI CAO[1], YISAI LIU[1], JIANGLING HAO[1], SHAOXI LI[1], AND CHI-HUA CHEN[2], (Senior Member, IEEE)

[1]Navigation College, Dalian Maritime University, Dalian 116026, China
[2]College of Mathematics and Computer Sciences, Fuzhou University, Fuzhou 350108, China

Corresponding author: Mingyang Pan (panmingyang@ dlmu.edu.cn)

This work was supported in part by the Fundamental Research Funds for the Central Universities under Grant 3132019400, and in part by the National Natural Science Foundation of China under Grant 61906043.

ABSTRACT Massive amount of water level data has been collected by using Internet of Things (IoT) techniques in the Yangtze River and other rivers. In this paper, utilizing these data to construct deep neural network models for water level prediction is focused. To achieve higher accuracy, both the factors of time and locations of data collection sensors are considered to perform prediction. And the network structures of gated recurrent unit (GRU) and convolutional neural network (CNN) are combined to build a CNN-GRU model in which the GRU part learns the changing trend of water level, and the CNN part learns the spatial correlation among water level data observed from adjacent water stations. The CNN-GRU model that using data from multiple locations to predict the water level of the middle location has higher accuracy than the model only based on GRU and other state-of-the-art methods including autoregressive integrated moving average model (ARIMA), wavelet-based artificial neural network (WANN) and long-short term memory model (LSTM), because of its ability to decrease the affections of abnormal value and data randomness of a single water station to some extent. The results are verified on an experiment dataset that including 30-year observed data of water level at several collection stations in the Yangtze River. For forecasting the 8-o'clock water levels of future 5 days, accuracy of the CNN-GRU model is better than that of ARIMA, WANN and LSTM models with three evaluation factors including Nash-Sutcliffe efficiency coefficient (NSE), average relative error (MRE) and root mean square error (RMSE).

INDEX TERMS CNN, GRU, water level prediction.

I. INTRODUCTION

The water level of the inland waterway is an important factor in guiding the navigation of vessels and their reasonable loading. Accurate prediction of the middle-term trend of water level is helpful to waterway maintenance and to improve the traffic safety and capacity.

By setting IoT sensors based on auto-telemetry technology at observation locations along an inland river, the dynamics water level can be periodically observed and be used to monitor the status of waterway. After perception, a large amount of historical data of water level has been collected. How to mine the value of these historical data of water level, especially to catch the water level trend for accurate prediction, are still hot topics that many scholars are constantly researching and exploring.

It is too complex and difficult to build a mathematical model which considers factors (e.g. rainfall, temperature, riverbed, water conservancy projects and other physical factors) to catch and analyse the trend of water level. Therefore, for water level prediction, the more mainstream methods have used the historical data of water level to build models by statistics and machine learning techniques.

Behzad *et al.* used the support vector machine (SVM) and artificial neural network (ANN) to predict groundwater level in different weathers and periods, and the results showed that SVM has more advantages than ANN in medium and long-term water level prediction, especially in the case of a small amount of data [1]. Guo *et al.* proposed an improved least squares support vector machine (LSSVM) model by including an extra bias error control term in the objective

The associate editor coordinating the review of this manuscript and approving it for publication was Junhua Li.

function for intelligent prediction of the daily water level in the Yangtze River, and more accurate forecasts were obtained although the improvement is regarded as moderate [2]. Adnan *et al.* proposed a water level prediction model for river flooding period based on back propagation (BP) neural network with a Kalman filter added at the output to improve prediction accuracy [3]. Galavi *et al.* proposed an ARIMA model and an adaptive network-based fuzzy inference system (ANFIS) to predict the water level of the Klang River, and the experimental results showed that the optimized ARIMA model achieved better results than the ANFIS model [4]. The ANN and ARIMA were combined by A. Wibowo *et al.* to establish a water level prediction model to achieve a higher prediction accuracy [5]. Wang *et al.* proposed an EEMD-ARIMA model coupling the ARIMA and ensemble empirical mode decomposition (EEMD) to forecast annual runoff time series, and the results showed that EEMD-ARIMA model can significantly improve ARIMA time series approaches for annual runoff time series forecasting [6]. Lee *et al.* proposed a short-term water level prediction model by combining neural networks and genetic algorithms (GA) for 15 water level locations in four major rivers in Korea, and the experimental results showed that the model has strong accuracy and adaptability [7]. Yao *et al.* proposed a GA-Elman method for river water level prediction using GA to optimize the Elman neural network, and the experimental results showed that the model converges quickly and has high precision [8]. Fan *et al.* analysed the influence of downstream return water and upstream water for forecasting the water level of a middle location, and proposed a short-term water level prediction model based on multiple linear regression (MLR) to obtain high accurate prediction in the Jianli water observation station of the Yangtze River [9]. Zhang *et al.* proposed a groundwater level prediction model based on principal components analysis (PCA) and multivariate time series controlled auto-regressive (CAR) according to the hysteresis and randomness of water level trend to obtain better results [10]. Yang *et al.* proposed a time-series forecasting model based on Random Forest to forecast the Taiwan Shimen reservoir's water level, and the experimental results indicate that the Random Forest forecasting model when applied to variable selection with full variables has better forecasting performance than the other models [11]. Adamowski and Chan proposed a WA-ANN model based on discrete wavelet transform (DWT) and ANN to perform monthly prediction of groundwater level, and the results show that the WA-ANN model can provide more accurate monthly average groundwater level prediction than ANN and ARIMA [12]. Seo *et al.* applied wavelet-based artificial neural network (WANN) and wavelet-based adaptive neural fuzzy inference system (WANFIS) to forecast daily water level, and the results indicated that the conjunction of wavelet decomposition and artificial intelligence models can be a useful tool for accurate forecasting daily water level and can yield better efficiency than the conventional forecasting models [13]. Wang *et al.* proposed a hybrid approach WD-RSPA based on wavelet de-noise (WD) and rank set pair analysis (RSPA) to improve forecasts of hydro-meteorological time series [14]. Anh *et al.* proposed a wavelet-artificial neural network (WAANN) model to addresses daily water level forecasting with short time, in which wavelet analysis (WA) was used to remove high-frequency random noise of time series data and ANN was then used to make the short-term prediction, the results of WAANN of water level forecasting showed better performance than ANN [15].

As mentioned above, various statistics and machine learning models such as SVM, ANFIS, ANN, GA, ARIMA, WA and their hybrid methods have been used in many related studies. However, utilizing deep neural networks to predict water level has rarely been discussed, which have won big successes in many other fields. For numerical prediction, it has been proven that the recurrent neural network (RNN) is good at building models from time series data. Since the water level data is normally collected location by location, and the several years dataset of each location is a typical time series data. Therefore, there is a big chance to get higher accuracy with the RNN method. Furthermore, as the water stations are adjacent one by one, a CNN could be a useful tool to capture the spatial relationship among them, and find the relationship of their water levels. Therefore, the contributions of this study is highlighted as follows.

1) This study proposes a CNN-GRU model to analyse the relationship of spatial-temporal data for water level prediction.
2) The GRU layers in the proposed method can be used to learn the time series wave shape of water level changes in each observation station.
3) The convolutional layers in the proposed method can be used to extract the spatial data features of water level changes from several observation stations.
4) A 30-years practical dataset of water level from several observation stations along the Yangtze River was applied to evaluate the proposed model.

The remainder of the paper is organized as follows. Section II discusses the techniques and applications of RNN, GRU, and CNN in previous studies. The dataset of water level from several observation stations along the Yangtze River is presented in Section III. Section IV presents the GRU-based prediction model to analyse the time series wave shape of water level changes in each observation station, and Section V illustrates the proposed CNN-GRU prediction model to analyse the spatial-temporal features of water level changes from several observation stations for the improvement of water level prediction. Sections VI and VII shows the practical experimental results and comparisons for the evaluation of the proposed model. Finally, the conclusions and future work are summarized in Section VIII.

II. LITERATURE REVIEWS

In recent years, GRU-based RNN has been successfully applied to spatial-temporal data and been quite popular among many scholars in other fields. In 2016, Fu *et al.* applied LSTM and GRU neural network methods to predict short-term traffic flow [16]. In 2017, Liu *et al.* applied GRU neural network to build predictive models to forecast Chinese primary energy consumption in 2021 [17]. In 2018, Zhao *et al.* proposed a local feature-based gated recurrent unit (LFGRU) networks to monitor machine health [18]. And, Zhang *et al.* proposed a GRU-based deep learning approach to predict urban traffic flow with combining weather condition data [19]. In 2019, Deng *et al.* proposed a sequence-to-sequence deep learning architecture based on the bidirectional gated recurrent unit (BiGRU) for type recognition and time location of combined power quality disturbance [20]. Li *et al.* proposed a multi-GRU prediction system based on GRU models to predict the future electricity generation [21]. Moreover, Le *et al.* proposed an approach by using multi-layer GRU to identify electron transport proteins [22].

As far as the combination of CNN and GRU, there is also some excellent works. For instance, a classifier by using the CNN with GRU was proposed to strengthen the relationship between words and words, text and text according to the input feature matrix, and high accurate text classification was obtained by the classifier [23]. In 2019, Tao *et al.* proposed a deep learning model based on 1D convnets and bidirectional GRU neural networks to forecast air pollution [24]. And Li *et al.* proposed a method by integrating CNN and GRU networks with vibration and acoustic emission signals to solve the gear pitting fault diagnosis problem [25]. Furthermore, Zhang *et al.* proposed a BiGRU-FCN network combining CNN and bidirectional GRU to investigate time series classification [26].

Similarly, in this paper, GRU is used to analyze the trends of water level of different stations, and CNN is further combined to consider the spatial relationship between water stations

III. DATASET OF WATER LEVEL

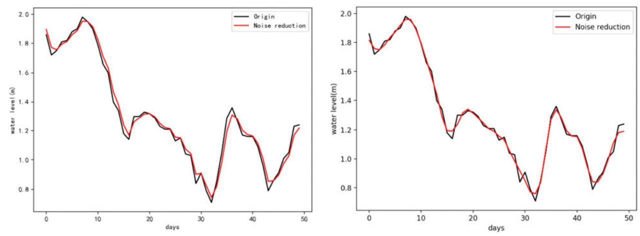
With the construction of digital waterway systems, several water level observation stations have been deployed along the Yangtze River. A large amount of water level data has been collected by IoT techniques such as auto-telemetry systems. According to the demand of vessels to develop voyage plans, the prediction of water level is normally on a daily basis. Therefore, the 8 o'clock observation values of every day of some water level stations along the Yangtze River are picked up to construct the dataset. The 8 o'clock observation values of each station in about 30 years form a time series dataset, where water level fluctuates periodically with daily changes.

Table 1 shows the dataset information with record count and the values of mean, std (standard deviation), min, max and quantiles (25%, 50% and 75%) of water level.

For a time series data, the data quality has a great influence on the prediction accuracy of the result models [27], [28]. Therefore, according to characteristics of the water level dataset, a two-steps preprocess is performed before training.

TABLE 1. Description of water level dataset.

Water station	Count	Mean	Std	Min	25%	50%	75%	Max
Jiujiang	11908	6.33	3.52	0.13	3.23	6.15	9.11	15.91
Anqing	11908	5.93	3.16	0.28	3.17	5.77	8.43	14.31
Wuhu	11908	4.04	2.20	0.25	2.11	3.80	5.75	10.08
Nanjing	11543	3.32	1.68	0.30	1.86	3.17	4.59	8.09



(a) EWA method (b) SG filtering

FIGURE 1. Noise reduction of the water level dataset.

1) There are a few abnormal outliers in water level dataset, probably caused by some systematic errors. The box plot is a standardized way of displaying the distribution of data based on the five number summary: minimum, first quartile (Q1), median, third quartile (Q3), and maximum. Values that are either more above an upper limit or more below a lower limit can be defined as outliers or suspected outliers [29]. So, to detect those outliers which deviate obviously from the general trend of the water level sequence, the box plot method is used with the upper limit of Q3 + 1.5 * IQR (the gap between Q1 and Q3) and the lower limit of Q1 − 1.5 * IQR. Then, the detected outliers will be replaced by the average values of the four points around them.

2) Denoising is useful to reduce over-fitting and improve the applicability of the final prediction model for a time series data. Savitzky-Golay (SG) filtering is used commonly in signal processing to filter out noise and eliminate effectively the randomness without changing the shape and width of the original signal obviously [30], [31]. In this case, a SG filtering is designed to perform the second procedure of noise reduction for the water level dataset, Figure 1 shows the denoising results of part of water level dataset.

Figure 1(a) shows the smoothed result of a classical exponential weighted averaging (EWA) method. A slight right shift of the result can be observed, and the mean absolute error (MAE) between the smoothed result and the original data is about 2.518. Figure 1(b) shows the smoothed result of the SG filtering, in which the frame length is set to 9 and the polynomial order is set to 5. With the same dataset, the MAE is only about 0.22 which is much better than that of EWA method, and that is why it was chosen to perform the denoising.

IV. GRU-BASED PREDICTION MODEL

This section applies the GRU-based prediction model to analyze the time series wave shape of water level changes.

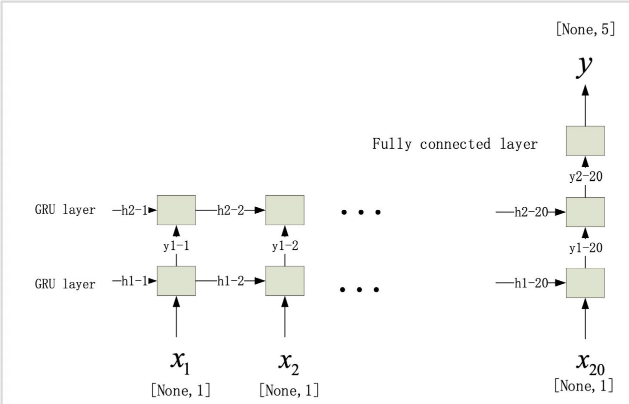


FIGURE 2. GRU-based prediction model of single water station.

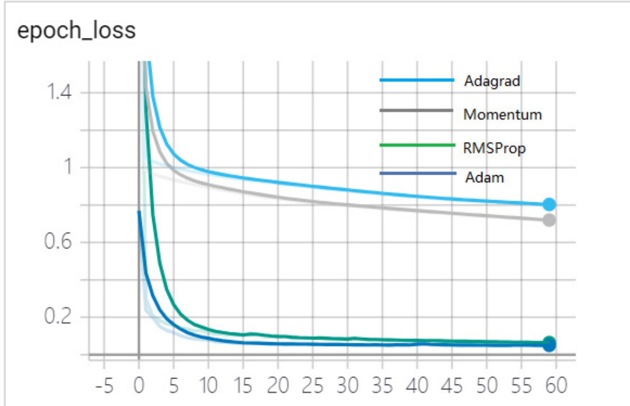


FIGURE 3. The loss curves with different optimizers.

In Subsection IV.A considers the data from the single water station and presents the principle of GRU-based model. Furthermore, the GRU-based model is used to analyze the data from the multiple water stations in Subsection IV.B.

A. SINGLE WATER STATION

GRU is a special kind of RNN, and compared with the LSTM architecture, it has some advantages of fast convergence and less parameters [32]. In this paper, a GRU-based prediction model of single water station is firstly constructed, its network is shown in Figure 2.

It is a 3-layer structure, both hidden layers are GRU, and the output layer is a common fully connected layer. The shape of the input layer is [None,20,1], 'None' is a placeholder that represents the batch size of training such as 128; '20' is the time step witch indicates the data of previous 20 days will be considered; '1' is the number of input features, that is, the 8 o'clock water level value of every day. The shape of the output layer is [None, 5], 'None' also represents batch size, and '5' is the number of output features, that is, the five predicted water level values of the future five days. So, the number of parameters of the model is 2369541, which is 784896 less than the model based LSTM, which could also lead to faster training speed.

The activation functions of the first and second GRU layers are set to 'tanh' and 'relu' respectively, and the output layer does not have an activation function. To prevent overfitting, add L2 normalization to each layer. So, the cost function of the model is:

$$J(\theta) = \frac{1}{2m}\sum_{i=1}^{m} L(\hat{y}^{(i)}, y^{(i)}) + \frac{\lambda}{2m}||\theta||_2^2 \tag{1}$$

where, θ is the parameters of model, m represents the number of samples, $L(\hat{y}^{(i)}, y^{(i)})$ represents the square error between sample value and predicted value, and is an empirical value, λ and is set to 0.001 here. And, to make the $J(\theta)$ decrease to converge, an optimizer is used.

Adaptive optimizers include Momentum, Adagrad, RMSProp and Adam (which is a combination of Momentum and RMSProp to some extent [33]) were compared, and their loss curves are shown in Figure 3 (with the dataset of Wuhu water station).

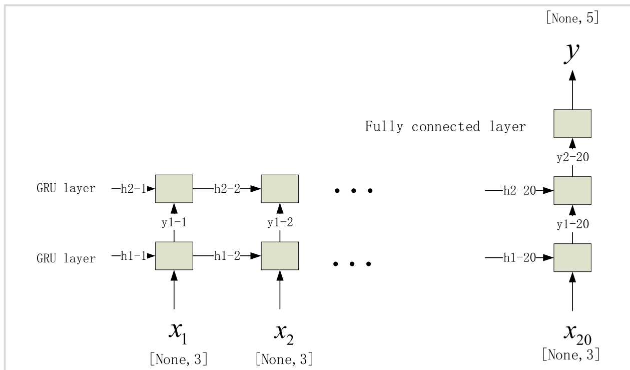


FIGURE 4. GRU-based prediction model with multiple water stations.

Obviously, the Adam optimizer has the best optimization effect with the fastest convergence and the lowest loss value. So the Adam is selected as the optimizer for the GRU-based model and the following other models.

In the actual prediction procedure, the shapes of input and output will be [1, 20, 5] and [1, 5] respectively, that is using the water level values of the last 20 days to predict the water level of next 5 days.

B. MULTIPLE WATER STATIONS

According to experiences of waterway maintenance and data analysis, the water level values of stations close to each other have great correlation. In theory, for a water station, utilizing not only the data of itself but also the data from around stations will help to improve the prediction model. Therefore, a prediction model based on the data from multiple water stations is further investigated. Its network structure is shown in Figure 4.

The structure is basically consistent with that of the single water station. The difference is mainly in the shape of the input layer. It is [None, 20, 3], here, '3' represents that it includes three water level value from three adjacent water stations. The shape of output layer is still [None, 5], that is, it only outputs the prediction values of next five days

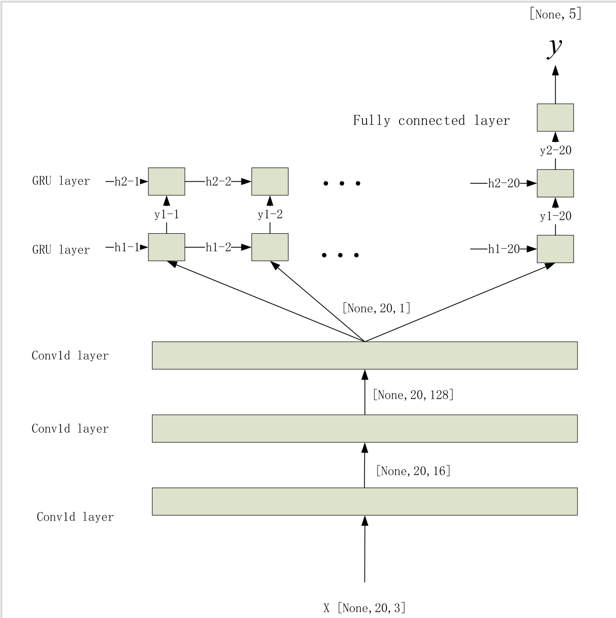


FIGURE 5. **CNN-GRU-Based prediction model with multiple water stations.**

of one station which is the middle station here. So, in the model, the water level values from an upstream station and a downstream station are utilized to help to predict the water level values of the middle station.

V. CNN-GRU-BASED PREDICTION MODEL

In the above-mentioned GRU-based prediction model, the correlation of adjacent water stations was considered in some extent. However, it is difficult to reflect the spatial relationship between the water stations only by the simple structural design of the input layer. The data correlative degrees of neighboring stations with different distances are not considered carefully.

As is well known, in image and audio recognition, and other fields, CNN has excellent performance because of its excellent ability at capturing spatially related features [34]. In order to better reflect the spatial distribution characteristics of water stations, CNN is further combined with the GRU network to build the water level prediction model. Its structure is shown in Figure 5.

The network consists of three convolution layers and three GRU layers. The shape of the input layer is [None, 20, 3], '3' represents three water level values of the three adjacent water stations; the shape of the output layer is [None, 5], '5' represents the predicted water level value of the next 5 days in the middle water station.

The three convolutional layers, namely C1, C2, and C3, all are one-dimensional convolutional layer, they can be expressed as:

$$C = f(wx + b) \tag{2}$$

f is the activation function, and all three layers are set to the 'relu' function. C1 has 16 convolution kernels, C2 has 128 kernels, and C3 has one kernel. So the final shape of the output of convolutional layers is [None, 20, 1]. It means the three values from three water stations are abstracted into one value according to their correlative degrees to the middle station. These values will have better smoothness than the original water level value of the middle water station, they then enter the first GRU layer. The rest part of GRUs is consistent with the one for single water station.

(a) The group with Anqing in middle (b) The group with Wuhu in middle

FIGURE 6. **The spatial relationship among water stations.**

VI. PRACTICAL EXPERIMENTAL RESULTS

The above proposed three prediction models are implemented by Python and the deep learning framework of TensorFlow-gpu2.0.0 in a workstation with two graphics cards of NVIDIA GeForce GTX 1080. And they are trained and tested in the water level dataset of the Yangtze River. The 30-year data is divided into training set and test set with the ratio of 8:2. Furthermore, according to comparison analysis of the historical data and experience, the dataset is further classified into three periods, the dry season (December-March), the middle water season (April, November) and the flood season (May-October)[35]. In different period the water level has different trends, so the classification will help to improve the prediction accuracy.

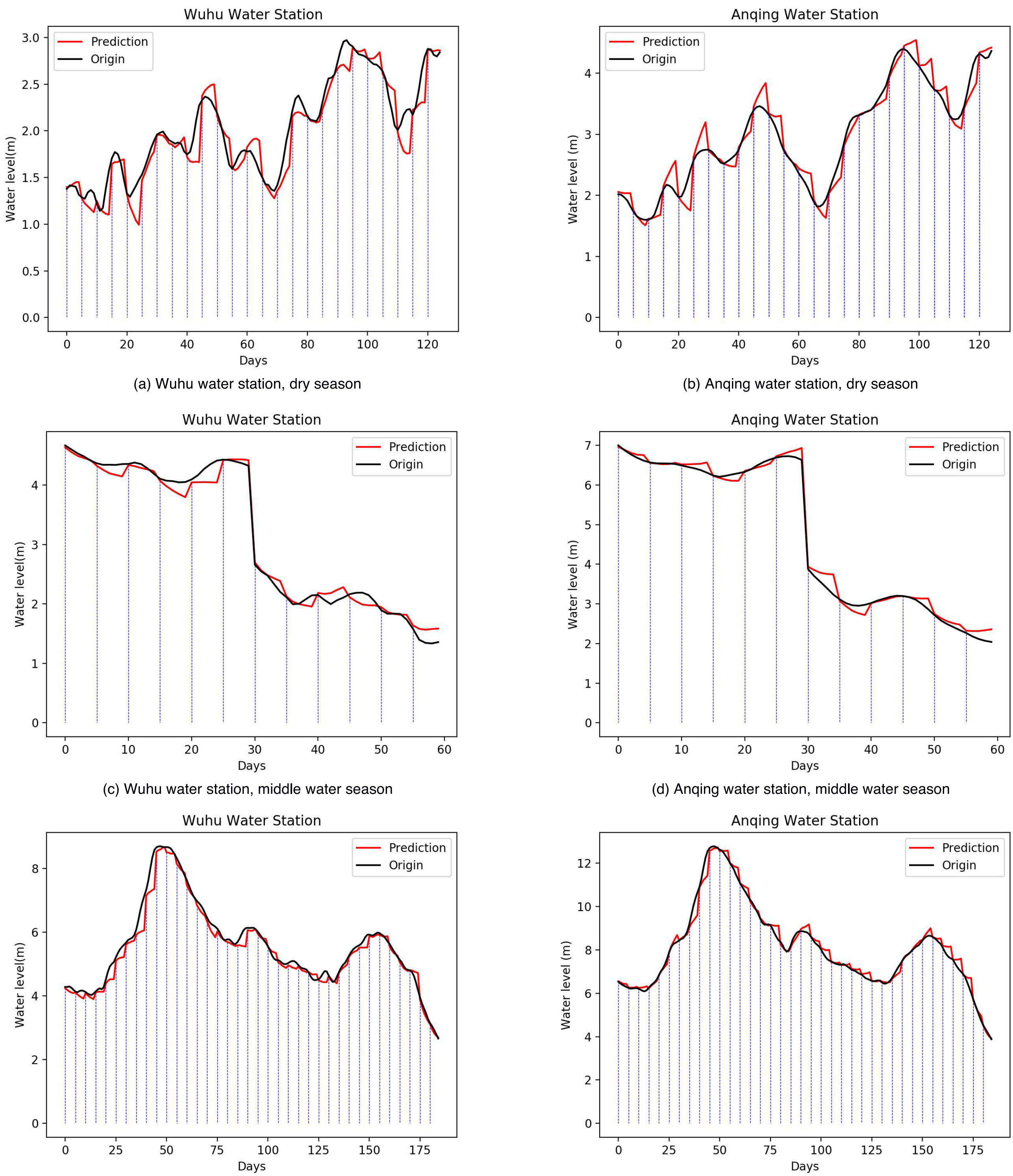
For testing the models with multiple water stations, two groups of three water stations are selected. As shown in Figure 6, the three stations located at Jiujiang, Anqing and Wuhu form one group with the Wuhu station in middle, and the three stations located at Anqing, Wuhu and Nanjing form the other group with the Anqing station in middle.

Figure 7 shows the prediction test results of two water stations, located at the Wuhu and Anqing city respectively, with GRU-based prediction model of single water station. Figure 8 shows the test results of the GRU-based prediction model with three water stations. And Figure 9 shows the test results of the CNN-GRU-based prediction model with three water stations. In the figures, interval of the vertical dotted line is 5, indicating the water level was predicted 5 days by 5 days (from the data of previous 20 days).

VII. COMPARATIVE ANALYSES OF PRACTICAL EXPERIMENTAL RESULTS

A. THE EVALUATION FACTORS

In order to test the performance of different water level prediction models, three factors are selected to evaluate



(a) Wuhu water station, dry season

(b) Anqing water station, dry season

(c) Wuhu water station, middle water season

(d) Anqing water station, middle water season

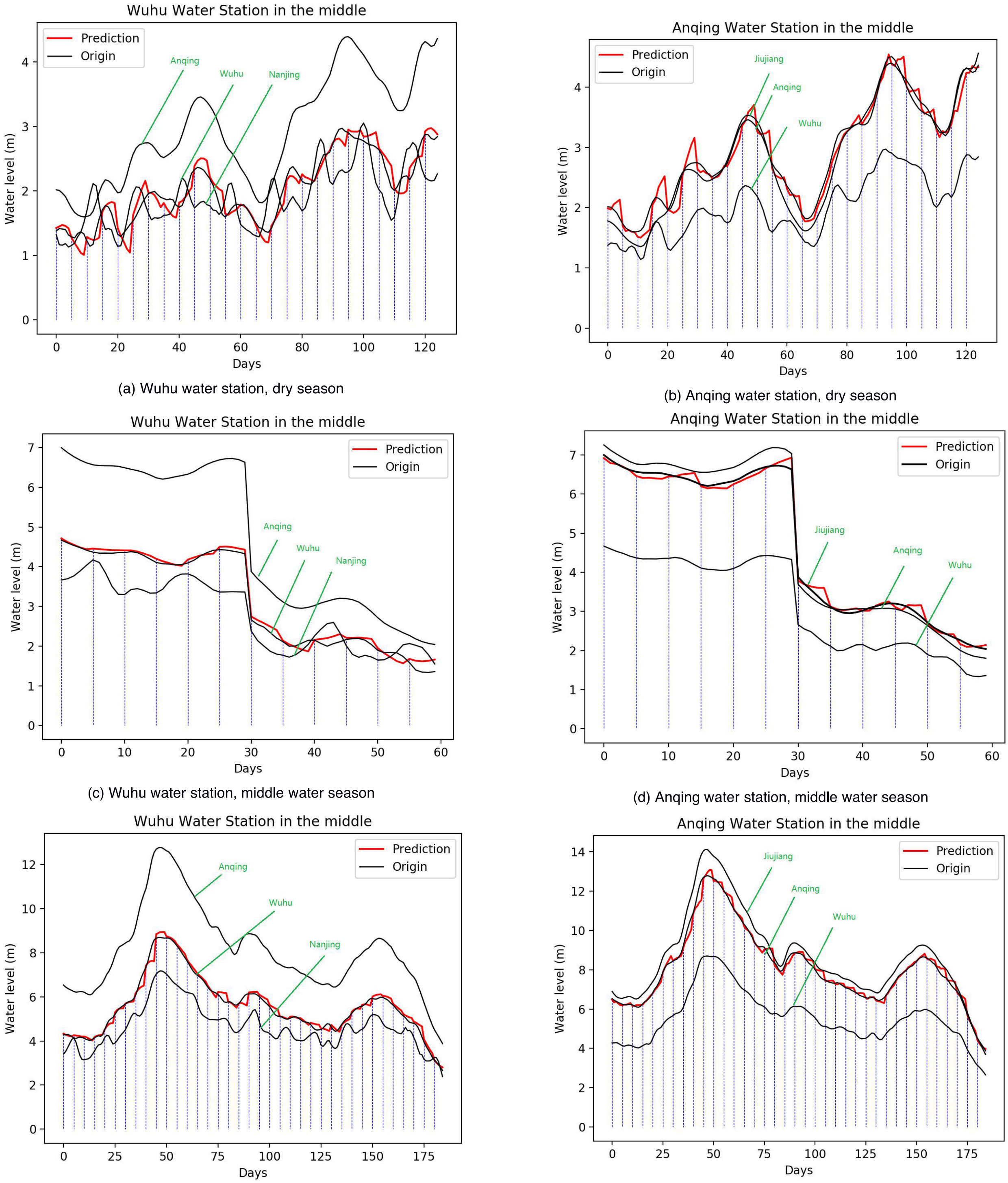
(e) Wuhu water station, flood season

(f) Anqing water station, flood season

FIGURE 7. **Test results of the prediction model of single water station.**

the accuracy including Nash-Sutcliffe efficiency coefficient (NSE), average relative error (MRE) and root mean square error (RMSE) which can be computed as the following formulas where y^i is the observed value, $\hat{y}^i$ is the predicted value and $\bar{y}$ is the observed average.

$$NSE = 1 - \frac{\sum_{i=1}^{n}(y^{(i)} - \hat{y}^{(i)})^2}{\sum_{i=1}^{n}(y^{(i)} - \bar{y})^2} \tag{3}$$



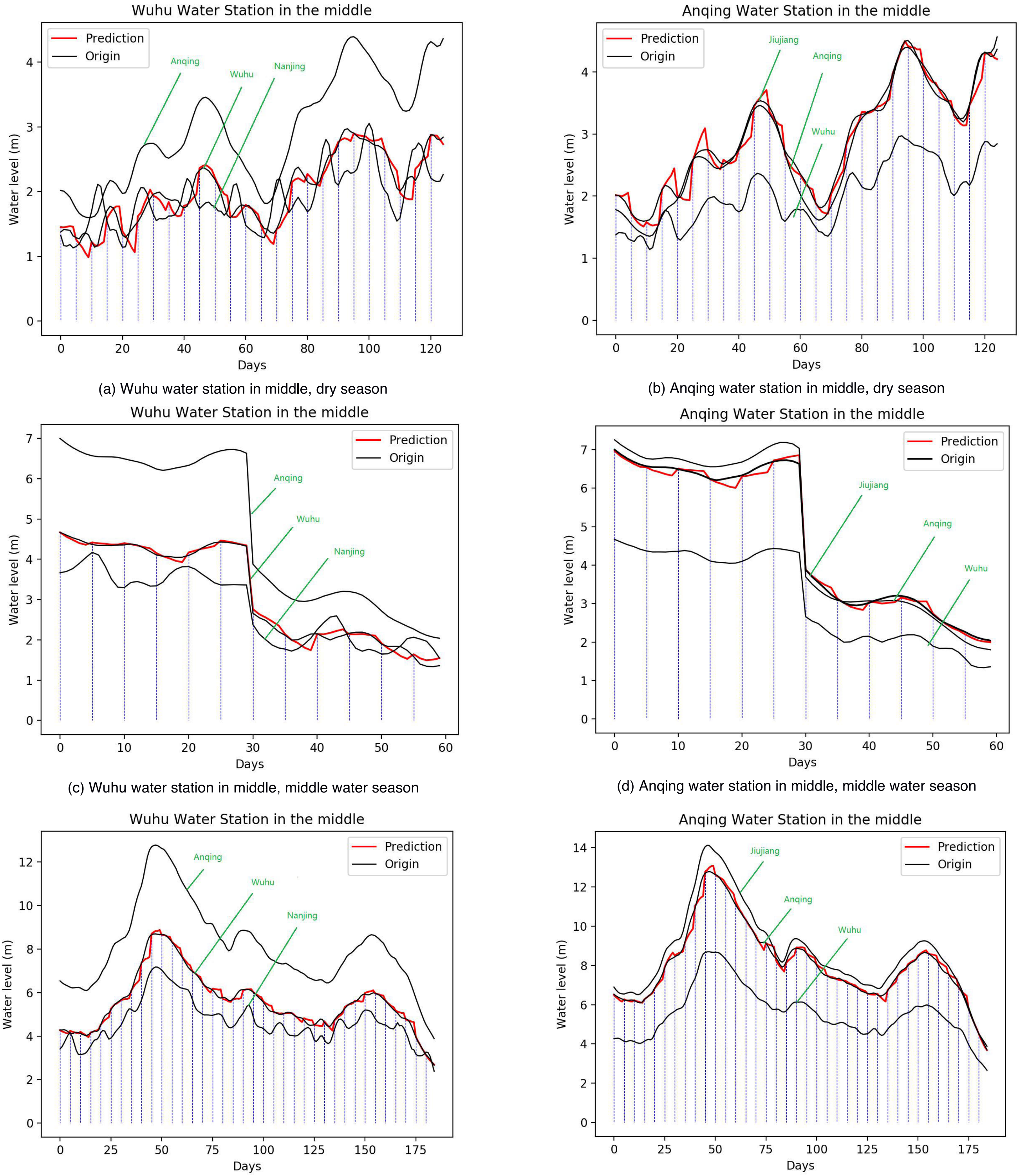
(a) Wuhu water station, dry season
(b) Anqing water station, dry season
(c) Wuhu water station, middle water season
(d) Anqing water station, middle water season
(e) Wuhu water station, flood season
(f) Anqing water station, flood season

FIGURE 8. Test results of GRU-based prediction model with three water stations.

$$MRE = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{y^{(i)} - \hat{y}^{(i)}}{y^{(i)}} \right| \tag{4}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y^{(i)} - \hat{y}^{(i)})^2} \tag{5}$$

The NSE is a factor usually used to evaluate the predictive power of hydrological models [36]. The value range of NSE is $(-\infty, 1]$. The closer the NSE is to 1, the better the prediction result, the closer the NSE is to 0, the closer the simulation result is to the average value of the measured values. When the NSE is less than 0, it indicates that the simulation results



(a) Wuhu water station in middle, dry season

(b) Anqing water station in middle, dry season

(c) Wuhu water station in middle, middle water season

(d) Anqing water station in middle, middle water season

(e) Wuhu water station in middle, flood season

(f) Anqing water station in middle, flood season

FIGURE 9. **Test results of CNN-GRU-based prediction model with three water stations.**

of the model are unreliable (not even the average of the measurements).

MRE and RMSE can reflect the difference between the observed value and the predicted value, their value range is $[0, +\infty)$. The perfect fit between the observed value and the predicted value will make RMSE and MRE 0, and the larger the value, the worse the prediction effect.

TABLE 2. Comparison of NSE (5-day average) of different models.

Dataset		ARIMA	WANN	LSTM	Model1	Model2	**Model3**
Dry season	Wuhu	0.9645	0.9573	0.9770	0.9668	0.9807	**0.9834**
	Anqing	0.9854	0.9771	0.9826	0.9829	0.9890	**0.9894**
Middle water season	Wuhu	0.9769	0.9792	0.9863	0.9874	0.9896	**0.9917**
	Anqing	0.9869	0.9937	0.9951	0.9935	0.9961	**0.9966**
Flood season	Wuhu	0.7771	0.8538	0.8418	0.8713	0.9112	**0.9196**
	Anqing	0.9401	0.9378	0.9314	0.9429	0.9659	**0.9676**
Average value		0.9385	0.9498	0.9524	0.9575	0.9721	**0.9747**

TABLE 3. Comparison of MRE (5-day average) of different models.

Dataset		ARIMA	WANN	LSTM	Model1	Model2	**Model3**
Dry season	Wuhu	9.637%	7.034%	7.373%	6.533%	6.392%	**5.910%**
	Anqing	5.684%	5.398%	5.714%	5.082%	4.160%	**3.918%**
Middle water season	Wuhu	4.059%	5.752%	4.167%	4.040%	4.324%	**3.569%**
	Anqing	2.943%	2.979%	2.073%	3.018%	2.169%	**1.921%**
Flood season	Wuhu	3.158%	3.555%	2.184%	2.730%	2.276%	**1.903%**
	Anqing	2.477%	2.182%	1.960%	1.919%	1.582%	**1.565%**
Average value		4.660%	4.483%	3.912%	3.887%	3.484%	**3.131%**

TABLE 4. Comparison of RMSE (5-day average) of different models.

Dataset		ARIMA	WANN	LSTM	Model1	Model2	**Model3**
Dry season	Wuhu	0.2269	0.1940	0.1937	0.1821	0.1512	**0.1439**
	Anqing	0.1860	0.1933	0.1932	0.1814	0.1493	**0.1411**
Middle water season	Wuhu	0.1649	0.1721	0.1284	0.1336	0.1218	**0.1085**
	Anqing	0.1609	0.1472	0.1254	0.1328	0.1158	**0.1076**
Flood season	Wuhu	0.2402	0.2716	0.2211	0.2394	0.1824	**0.1694**
	Anqing	0.2279	0.2709	0.2158	0.2313	0.1819	**0.1683**
Average value		0.2011	0.2082	0.1796	0.1834	0.1504	**0.1398**

B. COMPARATIVE ANALYSES

Table 2, Table 3 and Table 4 compare respectively the NSE, MRE and RMSE factors of the above prediction models and other state-of-the-art methods including ARIMA, WANN and LSTM models.

In the tables, "Model1" represents the GRU-based prediction model of single water station, "Model2" represents the GRU-based prediction model with three water stations, and "Model3" represents the CNN-GRU-based prediction model with three water stations. Figure 10 further shows the comparison of RMSE (5-day average) in the form of a histogram.

From the above comparisons, the following results can be observed:

1) In most cases, all three evaluation factors of the models based on GRU and LSTM are better than that of the classical ARIMA and WANN model, reflecting the advantages of deep learning.

2) In most cases, all three evaluation factors of "Model1" which based on GRU are nearly the same as the model based LSTM. However, with fewer parameters and faster convergence speed, GRU is chosen to construct the prediction models rather than LSTM.

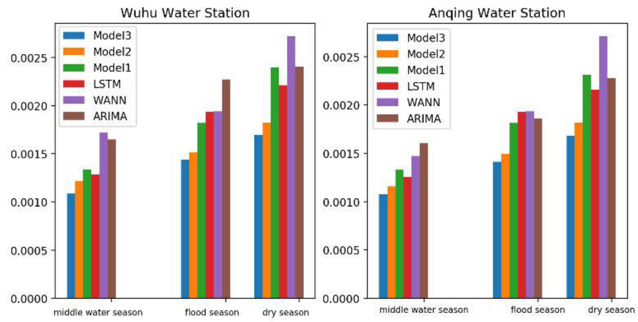


FIGURE 10. Histogram of RMSE of the different models.

3) In all seasons, the three all evaluation factors of "Model2" and "Model3" which based on three water stations are better than "Model1" which based single water station. Especially, the "Model3" based on CNN-GRU has the highest accuracy.

The advantage of the CNN-GRU-based prediction model includes that the GRU part can learn the changing trend of the water level, and the CNN part can learn the spatial correlation among water level data from the adjacent water stations. From the visualization analysis of the water level data, some extent delay of values can be found from an upper station to an adjacent lower station, and the closer water stations are, the higher their water level value are correlative. By the one-dimensional convolution, the CNN network part can learn kernels to capture relationship features among adjacent water stations, and utilize the water level value of the adjacent water stations with different weights. Therefore, the interference of abnormal value of a single water station can be weaken to some extent, and the generalization ability and robustness of the prediction model will be improved thereby.

4) The accuracy with the factor of NSE or RMSE of each model increases in order flood, dry, middle water season. It may reflect that the water level data in middle water season has smoother trend, and that in flood season is more fluctuating. However, when using MRE, accuracy increases in order dry, middle water, flood season. That is mainly because the MRE is a relative error which would obtain smaller evaluation value when the observe value is larger.

5) Regardless of which period, no matter which model, for Anqing water station, all three evaluation factors of prediction result are slightly better than that of the Wuhu water station. This is mainly because the Wuhu water station is closer to the estuary, and its water level is affected by tides more obviously. So, under the joint action of the water flow from upstream and the counter flow of tidal water, the change of water level is more complicated, which may lead to lower prediction accuracy.

6) Figure 11 shows the further detail of the RMSE of every day in the dry season. It can be observed that the error increases and the accuracy decreases gradually as the prediction time increases. The water level is affected by various uncertain factors, the change does not always follow a certain regularity, but has a certain degree of randomness. The longer

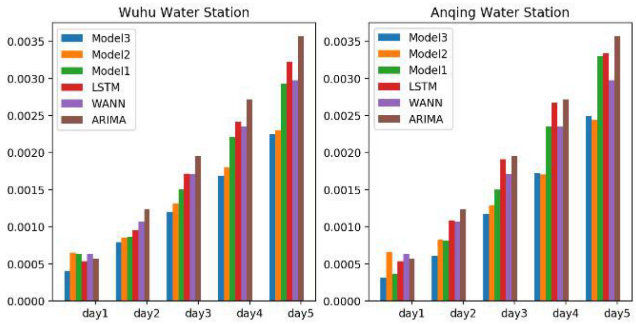


FIGURE 11. Comparison of daily RMSE of different models in dry season.

the prediction time, the harder it is for the model to capture the interference caused by randomness. And, from Figure 11, it can also be observed that the multi-day prediction accuracy of the all GRU-based models are higher than that of the ARIMA and WANN models. It reflects that the GRU network is less affected by randomness due to its consideration of long-term water lever states.

VIII. CONCLUSION AND FUTURE WORK

In this paper, deep neural networks are used to investigate water level prediction of inland rivers. The GRU-based prediction model is used to extract the time series wave shape features of water level changes, and a prediction model based CNN-GRU with multiple water stations is proposed for water level prediction. Practical experiments based on a 30-years water level dataset of the Yangtze River were performed to evaluate the proposed model. From the evaluation and comparison results, the following conclusions can be drawn:

1) The proposed CNN-GRU-based prediction model is superior to the classical ARIMA and WANN models. It has higher prediction accuracy, as it considers more data from multiple water stations with spatial correlation to predict the water level of a middle station, which is helpful for it to decrease the affection of the data randomness of a single water station to some extent.

2) For GRU-based models, the better the smoothness of time series data of water level, the higher prediction accuracy. In the Yangtze River, the accuracy of prediction increases in the order of flood, dry and middle water season. And, the farther the water station is away from the sea, the less affected by tides, and the higher prediction accuracy. In the next step of the study, the influence of tides on the water level prediction accuracy in the lower tidal water section of the Yangtze River will be investigated in detail, the observation data of water level and tide information will be combined to improve the prediction model.

3) Even the CNN-GRU-based model, its prediction errors increase with the forecasting days. Therefore, in practice, the long-term prediction more than 5 days is not very efficient.

The researched models of this paper have been used on a public service platform of the Yangtze River Nanjing Waterway Bureau, to provide intelligent water level prediction services in the form of REST interfaces. The intelligent prediction service can be integrated in many other application systems to aid vessels to navigate more safely or develop voyage plan more reasonably.

In the future, cluster methods can be used to analyze the water levels of different river segments for clustering and extracting important environmental factors. Furthermore, the clustered datasets can be adopted to train the CNN-GRU-based prediction model for the improvement of water level prediction.

ACKNOWLEDGMENT

The authors would like to thank the Changjiang Nanjing Waterway Bureau of the People's Republic of China for providing the dataset of water level.

REFERENCES

[1] M. Behzad, K. Asghari, M. Eazi, and M. Palhang, "Generalization performance of support vector machines and neural networks in runoff modeling," *Expert Syst. Appl.*, vol. 36, no. 4, pp. 7624–7629, May 2009.

[2] T. Guo, W. He, Z. Jiang, X. Chu, R. Malekian, and Z. Li, "An improved LSSVM model for intelligent prediction of the daily water level," *Energies*, vol. 12, no. 1, p. 112, 2019.

[3] R. Adnan, F. A. Ruslan, A. M. Samad, and Z. M. Zain, "Flood water level modelling and prediction using artificial neural network: Case study of Sungai Batu Pahat in Johor," in *Proc. IEEE Control Syst. Graduate Res. Colloq.*, Shah Alam, Malaysia, Jul. 2012, pp. 16–17.

[4] H. Galavi, M. Mirzaei, L. T. Shul, and N. Valizadeh, "Klang river-level forecasting using ARIMA and ANFIS models," *J. Amer. Water Works Assoc.*, vol. 105, no. 9, pp. E496–E506, Sep. 2013.

[5] A. Wibowo and S. H. Arbain, "Time series methods for water level forecasting of Dungun river in Terengganu Malaysia," *Int. J. Eng. Sci. Technol.*, vol. 4, no. 4, pp. 1803–1811, 2012.

[6] W.-C. Wang, K.-W. Chau, D.-M. Xu, and X.-Y. Chen, "Improving forecasting accuracy of annual runoff time series using ARIMA based on EEMD decomposition," *Water Resour. Manage.*, vol. 29, no. 8, pp. 2655–2675, Jun. 2015.

[7] G. Lee, S. Lee, and H. Park, "Improving applicability of neuro-genetic algorithm to predict short-term water level: A case study," *J. Hydroinform.*, vol. 16, no. 1, pp. 218–230, Jan. 2014.

[8] Z. Yao, J. Xu, J. Kong, and S. Liu, "Prediction of river water level by GA-Elman model," *J. Yangtze River Sci. Res. Inst.*, vol. 35, no. 9, pp. 34–37, 2018.

[9] X. Fan, Z. Li, C. Yang, and X. Cao, "Short-term water level prediction method based on multiple linear regression analysis," *Port Waterway Eng.*, no. 487, pp. 159–162, 2014.

[10] Z. Zhang, Z. Liang, B. Feng, J. Huang, and D. Wu, "Groundwater level forecast based on principal component analysis and multivariate time series model," *Adv. Water Sci.*, vol. 28, no. 3, pp. 415–420, 2017.

[11] J.-H. Yang, C.-H. Cheng, and C.-P. Chan, "A time-series water level forecasting model based on imputation and variable selection method," *Comput. Intell. Neurosci.*, vol. 2017, pp. 1–11, Nov. 2017.

[12] J. Adamowski and H. F. Chan, "A wavelet neural network conjunction model for groundwater level forecasting," *J. Hydrol.*, vol. 407, nos. 1–4, pp. 28–40, Sep. 2011.

[13] Y. Seo, S. Kim, O. Kisi, and V. P. Singh, "Daily water level forecasting using wavelet decomposition and artificial intelligence techniques," *J. Hydrol.*, vol. 520, pp. 224–243, Jan. 2015.

[14] D. Wang, A. G. Borthwick, H. He, Y. Wang, J. Zhu, Y. Lu, P. Xu, X. Zeng, J. Wu, L. Wang, X. Zou, J. Liu, Y. Zou, and R. He, "A hybrid wavelet de-noising and rank-set pair analysis approach for forecasting hydro-meteorological time series," *Environ. Res.*, vol. 160, pp. 269–281, Jan. 2018.

[15] N. T. N. Anh, N. Q. Dat, N. T. Van, N. N. Doanh, and N. Le An, "Wavelet-artificial neural network model for water level forecasting," in *Proc. Int. Conf. Res. Intell. Comput. Eng. (RICE)*, San Salvador, El Salvador, Aug. 2018, pp. 1–6.

[16] R. Fu, Z. Zhang, and L. Li, "Using LSTM and GRU neural network methods for traffic flow prediction," in *Proc. 31st Youth Acad. Annu. Conf. Chin. Assoc. Automat. (YAC)*, Wuhan, China, Nov. 2016, pp. 11–13.

[17] B. Liu, C. Fu, A. Bielefield, and Y. Liu, "Forecasting of Chinese primary energy consumption in 2021 with GRU artificial neural network," *Energies*, vol. 10, no. 10, p. 1453, 2017.
[18] R. Zhao, D. Wang, R. Yan, K. Mao, F. Shen, and J. Wang, "Machine health monitoring using local feature-based gated recurrent unit networks," *IEEE Trans. Ind. Electron.*, vol. 65, no. 2, pp. 1539–1548, Feb. 2018.
[19] D. Zhang and M. R. Kabuka, "Combining weather condition data to predict traffic flow: A GRU-based deep learning approach," *IET Intell. Transp. Syst.*, vol. 12, no. 7, pp. 578–585, Sep. 2018.
[20] Y. Deng, L. Wang, H. Jia, X. Tong, and F. Li, "A sequence-to-sequence deep learning architecture based on bidirectional GRU for type recognition and time location of combined power quality disturbance," *IEEE Trans Ind. Informat.*, vol. 15, no. 8, pp. 4481–4493, Aug. 2019.
[21] W. Li, T. Logenthiran, and W. L. Woo, "Multi-GRU prediction system for electricity generation's planning and operation," *IET Gener., Transmiss. Distrib.*, vol. 13, no. 9, pp. 1630–1637, May 2019.
[22] N. Q. K. Le, E. K. Y. Yapp, and H.-Y. Yeh, "ET-GRU: Using multi-layer gated recurrent units to identify electron transport proteins," *BMC Bioinf.*, vol. 20, no. 1, p. 377, Dec. 2019.
[23] L.-X. Luo, "Network text sentiment analysis method combining LDA text representation and GRU-CNN," *Pers. Ubiquitous Comput.*, vol. 23, nos. 3–4, pp. 405–412, Jul. 2019.
[24] Q. Tao, F. Liu, Y. Li, and D. Sidorov, "Air pollution forecasting using a deep learning model based on 1D convnets and bidirectional GRU," *IEEE Access*, vol. 7, pp. 76690–76698, 2019.
[25] X. Li, J. Li, Y. Qu, and D. He, "Gear pitting fault diagnosis using integrated CNN and GRU network with both vibration and acoustic emission signals," *Appl. Sci.*, vol. 9, no. 4, p. 768, 2019.
[26] G. Zhang and B. Liu, "Research on time series classification using CNN and bidirectional GRU," *J. Frontiers Comput. Sci. Technol.*, vol. 13, no. 6, pp. 916–927, 2019.
[27] H. Liu, G. Sun, and R. Cao, "The application of GM (1, 1) dynamic model in the forecast of groundwater level in Wujiang city," *J. Geol. Hazards Environ. Preservation*, vol. 19, no. 3, pp. 47–51, 2008.
[28] D. Luo, S. Liu, and Y. Dang, "The optimization of grey model GM (1, 1)," *Eng. Sci.*, vol. 5, no. 8, pp. 50–53, 2003.
[29] M. Frigge, D. C. Hoaglin, and B. Iglewicz, "Some implementations of the boxplot," *Amer. Statist.*, vol. 43, no. 1, pp. 50–54, Feb. 1989.
[30] J. Chen, P. Jönsson, M. Tamura, Z. Gu, B. Matsushita, and L. Eklundh, "A simple method for reconstructing a high-quality NDVI time-series dataset based on the Savitzky–Golay filter," *Remote Sens. Environ.*, vol. 91, nos. 3–4, pp. 332–344, 2004.
[31] H. H. Madden, "Comments on the Savitzky–Golay convolution method for least-squares-fit smoothing and differentiation of digital data," *Anal. Chem.*, vol. 50, no. 9, pp. 1383–1386, Aug. 1978.
[32] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," 2014, *arXiv:1412.3555*. [Online]. Available: http://arxiv.org/abs/1412.3555
[33] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980
[34] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015.
[35] S. Lu and Y. Liu, "Impact of extreme water levels on characteristics of vessel traffic flow in the middle reaches of Yangtze River," *J. Chongqing Jiaotong Univ.*, vol. 36, no. 03, pp. 103–107, 2017.
[36] R. H. McCuen, Z. Knight, and A. G. Cutter, "Evaluation of the Nash–Sutcliffe efficiency index," *J. Hydrol. Eng.*, vol. 11, no. 6, pp. 597–602, Nov. 2006.

MINGYANG PAN received the Ph.D. degree in traffic information engineering and control from Dalian Maritime University, Dalian, China, in 2004.

He is currently an Associate Professor with the Navigation College, Dalian Maritime University. He is also the Director of the Technical Institute of Navigation, Dalian Maritime University. His research interests are in the fields of electronic chart display and information system (ECDIS), digital waterway systems, and intelligent waterway transportation systems.

HAINAN ZHOU received the bachelor's degree from Ningbo University, in 2017. He is currently pursuing the master's degree with Dalian Maritime University.

His research interests include time series modeling, natural language processing, and computer vision.

JIAYI CAO received the bachelor's degree from the Shenyang Institute of Engineering University, in 2018. He is currently pursuing the master's degree with Dalian Maritime University.

His research interests include data augmentation and computer vision.

YISAI LIU received the bachelor's degree with Dalian Maritime University, in 2017, where he is currently pursuing the master's degree.

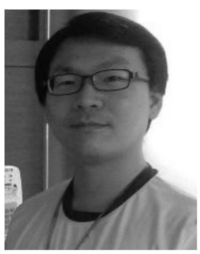
His research interests include data analysis and computer vision.

JIANGLING HAO received the master's degree in systems engineering from Northern Jiaotong University, Beijing, China, in 1999.

He is currently an Associate Professor with the Navigation College, Dalian Maritime University. He is also working with the Institute of Navigation Technology, Dalian Maritime University. He has long been engaged in the research and application of ECS, electronic chart display and information system (ECDIS), ship dynamic monitoring systems, ship navigation systems, navigational telemetry remote control systems, digital waterway systems, and intelligent waterway transportation systems.

SHAOXI LI received the master's degree in traffic information engineering and control from Dalian Maritime University, Dalian, China, in 2004.

He is currently an Associate Professor with the Navigation College, Dalian Maritime University. His research interests focus on electronic chart display and information system (ECDIS), intelligent waterway transportation systems, ocean mapping, cartography, and marine geography information systems (MGISs).



CHI-HUA CHEN (Senior Member, IEEE) is currently a Chair Professor with Dalian Maritime University, China. He is also a Distinguished Professor with Fuzhou University. He has published over 270 academic articles and patents. His recent research interests include the Internet of Things, machine learning, deep learning, big data, cellular networks, and intelligent transportation systems.