

A hybrid Wavelet-CNN-LSTM deep learning model for short-term urban water demand forecasting

Zhengheng Pu^{1,2}, Jieru Yan^{1,2}, Lei Chen^{1,2}, Zhirong Li^{1,2}, Wenchong Tian^{1,2}, Tao Tao^{1,2}, Kunlun Xin (✉)^{1,2}

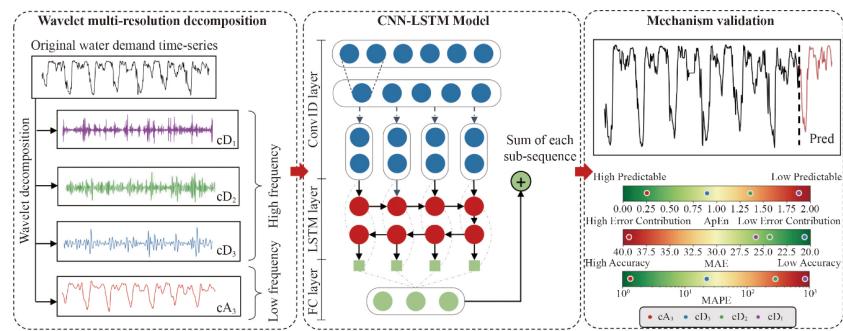
1 College of Environmental Science and Engineering, Tongji University, Shanghai 200092, China

2 Smart Water Joint Innovation RD Center, Tongji University, Shanghai 200092, China

HIGHLIGHTS

- A novel deep learning framework for short-term water demand forecasting.
- Model prediction accuracy outperforms other traditional deep learning models.
- Wavelet multi-resolution analysis automatically extracts key water demand features.
- An analysis is performed to explain the improved mechanism of the proposed method.

GRAPHIC ABSTRACT



ARTICLE INFO

Article history:

Received 3 March 2022

Revised 13 July 2022

Accepted 25 July 2022

Available online 10 September 2022

Keywords:

Short-term water demand forecasting
Long-short term memory neural network
Convolutional Neural Network
Wavelet multi-resolution analysis
Data-driven models

ABSTRACT

Short-term water demand forecasting provides guidance on real-time water allocation in the water supply network, which help water utilities reduce energy cost and avoid potential accidents. Although a variety of methods have been proposed to improve forecast accuracy, it is still difficult for statistical models to learn the periodic patterns due to the chaotic nature of the water demand data with high temporal resolution. To overcome this issue from the perspective of improving data predictability, we proposed a hybrid Wavelet Multi-Resolution Analysis (MRA) and implement it into an advanced deep learning model, CNN-LSTM. Four models - ANN, ConvID, LSTM, GRUN - are used to compare with Wavelet-CNN-LSTM, and the results show that Wavelet-CNN-LSTM outperforms the other models both in single-step and multi-steps prediction. Besides, further mechanistic analysis revealed that MRA produce significant effect on improving model accuracy.

© Higher Education Press 2023

1 Introduction

With the rapid urbanization and the requirements of national energy-saving policies, fine management with high efficiency in water supply system is imperative. As a basic strategic management task in urban water distribution system, short-term water demand forecasting plays an important role in pump scheduling optimization.

✉ Corresponding author

E-mail: xkl@tongji.edu.cn

Special Issue—Artificial Intelligence/Machine Learning on Environmental Science & Engineering (Responsible Editors: Yongsheng Chen, Xiaonan Wang, Joe F. Bozeman III & Shouliang Yi)

More specifically, accurate water demand forecast can provide guidance on real-time water allocation in the water supply network, which can save unnecessary water supply costs and avoid serious water supply accident (Billings and Jones, 2011). However, short-term water demand (e.g., hourly, or sub-hourly) usually exhibit complex nonlinear behaviors in a dynamic manner, which is a result of many factors, such as temperature, rainfall, and historical urban water demand (Xenochristou et al., 2017). Thus it is challenging to obtain accurate water demand forecasting.

Many studies have been conducted to obtain accurate water demand forecasting. According to different principles behind, these methods can be divided into

linear and non-linear methods (Zhang, 2001). Many linear statistical methods were applied in some early studies, such as multiple linear regression (MLR) (Jowitt and Xu, 1992; Bougadis et al., 2005; Zhou et al., 2000) and autoregressive integrated moving average model (ARIMA) (Maidment and Parzen, 1984; Chen and Boccelli, 2014). The advantages of the linear models are that the model structures are relatively simple and usually they have good interpretability. However, the assumption of linearity limits their applicability in water demand forecasting problems for their poor capability of describing the nonlinearity in such problems.

To address this issue, many non-linear methods have been developed to predict water demand. The representative models include artificial neural networks (ANN) (Adamowski, 2008; Ghiassi et al., 2008; Firat et al., 2010), support vector machines (SVM) (Peña-Guzmán et al., 2016; Candelieri et al., 2019; Herrera et al., 2010) and random forests (RF) (Herrera et al., 2010; Chen et al., 2017; Duerr et al., 2018). Some advanced techniques exploited, such as activation functions (Sharma et al., 2017), kernel functions (Amari and Wu, 1999), enable these methods to describe non-linear relationship. But there still exist some problems, such as complex feature engineering. In machine learning, feature engineering is of paramount importance. If the key features that determine the change in water demand were not extracted properly, the desired prediction results would be difficult to achieve.

In recent years, deep learning methods have proven to be very effective in various prediction areas, such as stock price forecasting (Cao and Wang, 2019), electricity forecasting (Torres et al., 2017; Bedi and Toshniwal, 2019; Hafeez et al., 2020), traffic forecasting (Huang et al., 2017; Yu et al., 2017a; 2017b) and so on. Among the various deep learning methods, Recurrent neural networks (RNNs) have proven their superiority over other neural networks. As a deep network structure, it enables the neural network to extract features from the input data adaptively to achieve an End-to-end prediction model (Dara and Tumma, 2018). Moreover, recurrent neural networks can model the dependencies between sequences and are therefore suitable for time series prediction (Sherstinsky, 2020). For instance, a hybrid model of CNN and Bi-LSTM was proposed to predict daily water demand. The key features in the historical water demand data and meteorological data are extracted through one-dimensional convolution and pooling layers, which are then fed into the Bi-LSTM model to construct temporal dependencies (Hu et al., 2019). In addition, to take full advantage of the rich water demand monitoring data, deep learning models are used for higher temporal resolution water demand forecasting. A multi-headed GRU network structure was constructed with multi-timescale input features to predict the water demand of a DMA in a residential area in Changzhou City in China with a time

step of 15 min (Guo et al., 2018). A fully connected neural network was also constructed to avoid the accumulation of prediction errors in continuous prediction. Meanwhile, a long short-term memory (LSTM)-based model was also applied to predict short-term urban water demand at different timescale for Hefei City in China (Mu et al., 2020). These studies have shown that deep learning models outperform traditional predictive models by a significant margin.

However, there is still room for optimization of the prediction model because of the stochastic fluctuation of water demand. It is well known that what is important for predictive models is to discover the cyclical patterns of data variation and thus estimate the future state. Therefore, how to decompose the corresponding predictable and random components from the data and enhance the overall predictability of the data is the key to improve the prediction accuracy of the model. Nevertheless, most of the current studies focus on feature selection and model selection instead of improving the predictability of the water demand data itself.

In this paper, we aim to make contributions to addressing the issues on short-term water demand forecasting. First, a new model, Wavelet-CNN-LSTM, is proposed, which combines time-frequency decomposition characteristics of Wavelet Multi-Resolution Analysis (MRA) and implement it into an advanced deep learning model, CNN-LSTM. Second, MRA was applied to extract predictable components in different frequency domains influenced by different scale features in water demand data, combined with CNN-LSTM deep learning models for separate training of different sub sequence, so as to minimize the impact of random components on the prediction results, thus improving the prediction accuracy of the model. To demonstrate the model performance, the proposed Wavelet-CNN-LSTM model for short-term water demand forecasting is compared with another four widely used deep learning models. Third, a measure termed the approximate entropy is introduced to quantify the predictability of water demand data for explaining the superiority of the Wavelet-CNN-LSTM model. The results show that the proposed Wavelet-CNN-LSTM outperforms the other four deep learning models in both single-step and multi-step prediction by improving the predictability of the water demand data.

2 Methodology

2.1 One-dimension convolution neural network (1D-CNN)

Convolution neural network (Yang and Li, 2017) is a feed-forward neural network evolved from multi-layer perceptron algorithm (MLP), which makes convolution neural network excel in image processing field due to its structural features such as local connection, weight

sharing, and down sampling.

The basic structure of a convolutional neural network consists of an input layer, a convolutional layer, a pooling layer, a fully connected layer and an output layer (Bouvie, 2006). To be able to extract high-dimensional feature from complex data, a multi-blocks network structure is generally used, which contains multiple convolutional and pooling layers in each block (Simonyan and Zisserman, 2014).

Traditionally, two-dimensional convolutional is mainly used in image processing. To take advantage of the powerful local feature extraction capability of CNNs also in 1D sequence data, the classical 2D CNNs are extended to 1D. The computational principle of 1D convolutional neural network is very similar to that of 2D convolutional neural network, the only difference is that 1D convolutional neural network performs convolutional operation with 1D input vector through a 1D convolutional kernel, as shown by the schematic diagram in Fig. 1.

The expressions of the one-dimensional convolution operation are shown in Eqs. (1) and (2).

Convolution layer:

$$y_p^{(l+1)} = f(y_p^{(l)} \otimes \text{rot180}(\omega_p^{(l)}) + b_p^{(l)}). \quad (1)$$

Pooling layer:

$$y_p^{(l+1)} = \text{pooling}(y_p^{(l)}), \quad (2)$$

where $y^{(l)}$ denotes the activation value of the l -th layer; $\omega_p^{(l)}$ and $b_p^{(l)}$ are the weight and bias matrices, respectively; \otimes is the symbol of convolutional operation; $\text{rot180}(\cdot)$ means to rotate the input by 180° .

2.2 Long-short term memory neural network (LSTM)

Long short-term memory neural network (LSTM) is an advanced recurrent neural network structure for several

learning problems related to sequence data. It has been proven that LSTM has better performance in capturing long-term temporal dependencies compared with traditional recurrent neural networks (Hochreiter and Schmidhuber, 1997). Thus, LSTM is widely used to solve such problems in various fields, which include language modeling and translation (Sundermeyer et al., 2015; Kurata et al., 2017), time series modeling (Song et al., 2020), speech acoustic modeling (Zen, 2015), protein secondary structure prediction (Sønderby and Winther, 2014; Cheng et al., 2020), video data analysis (Ullah et al., 2017), etc. The core idea of the LSTM network structure design mainly consists of a memory unit that can maintain its state over time, and a nonlinear gating unit that controls the flow of information. Specifically, the gating mechanism of LSTM mainly includes forget gate, update gate and output gate, as listed below:

Forget gate:

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f). \quad (3)$$

Input gate:

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i). \quad (4)$$

Output gate:

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o), \quad (5)$$

where f_t , i_t , o_t represent forgot gate, input gate and output gate respectively; x_t is hidden status unit at time $t-1$; $W_f, W_i, W_o, U_f, U_i, U_o$ are the related weight matrices; b_f, b_i, b_o are the related biases.

The gating mechanism refers to mapping the result of a linear combination of parameters learned by the model with the input through a sigmoid function into an output value between 0 and 1, which is used to determine the proportion of information passed. The LSTM gating unit structure is shown in Fig. 2.

The forget gate controls the proportion of internal state

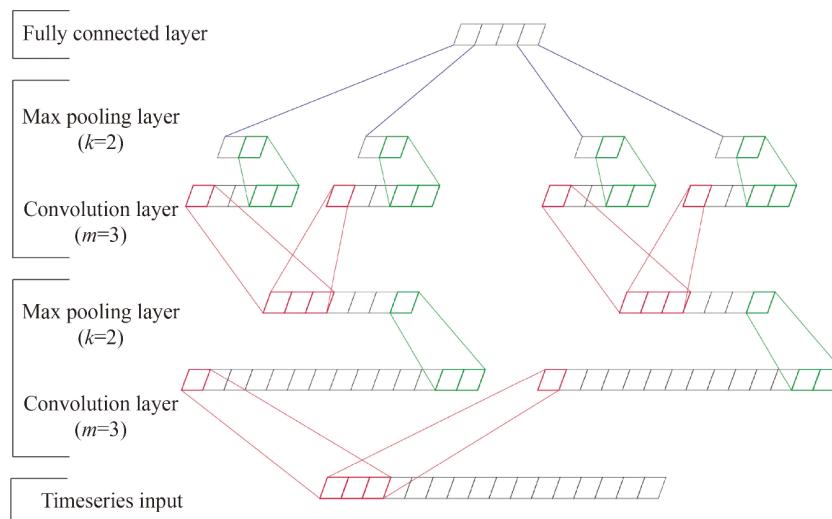


Fig. 1 Schematic diagram of 1D-CNN.

(C_t) that needs to be forgotten at the previous moment; the input gate controls the proportion of new information that needs to be updated at the current moment; and the output gate controls the proportion of internal state (C_t) that needs to be output to the hidden state at the current moment. The formula for each memory cell of the LSTM model can be written as:

$$\bar{C}_t = \sigma(W_c x_t + U_c h_{t-1} + b_c), \quad (6)$$

$$C_t = f_t \otimes C_{t-1} + i_t \otimes \bar{C}_t, \quad (7)$$

$$h_t = o_t \otimes \tanh(C_t), \quad (8)$$

where W_c and U_c are the weight matrices for x_t and h_{t-1} , respectively; b_c is the respective bias; C_t is the internal status unit at time t and \bar{C}_t is the candidate status information at time t .

2.3 Wavelet-CNN-LSTM model

In our problem, there is a significant autocorrelation in water demand data, so history water demand is often used as an input in water demand forecasting models. However, the prediction accuracy of the LSTM model is limited by the length of the input sequence, which is known as the gradient disappearance problem, resulting in some important historical information that may not be available to the model (Hochreiter, 1998).

In contrast, CNN-LSTM architecture involves using CNN layers for feature extraction on input data combined with LSTM to support sequence prediction. The key features of the novel model structure are that it can extract the key temporal features from the input sequence into a shorter sequence by exploiting the filters in the convolutional layers (Sainath et al., 2015), thus ensuring

that the subsequent LSTM layer can encode the temporal dependencies in the time series more effectively.

Moreover, one of the most difficult challenges of the short-term water demand forecasting problem is to control the impact of the random components in the time series on the forecasting accuracy. Overall, urban water demand data has a significant daily pattern of variation, but at specific times, it fluctuates randomly within a certain range of water demand due to the influence of various factors at different scale. For example, urban water demand is varying owing to different short (e.g., climate) or long-term factors (e.g., population) that reflect different frequencies of water demand variation pattern (Xenochristou et al., 2017; Goodchild, 2003). To solve this problem, both time domain and frequency domain information is needed.

In signal processing, the Fourier transform (FT) has been the standard method for spectral decomposition (Bracewell, 1989). However, the basis function of FT is defined over the entire time domain, which means the loss of frequency information over time. To overcome this weakness of FT, Short Time Fourier Transform (STFT) has been proposed (Nussbaumer, 1981). STFT uses a fixed time window to split the signal for the Fourier transform. Depending on the compactly supported basis function, STFT can record both time and frequency domain information, but it is difficult to choose a suitable window width because both time and frequency resolutions are fixed and cannot be dynamically adjustable with frequency, and Wavelet multi-resolution analysis (MRA) is proposed to address this problem.

MRA is a method applied to signal time-frequency analysis by dividing the time-frequency plane in an adaptive manner (Boggess and Narcowich, 2005). (i.e., a

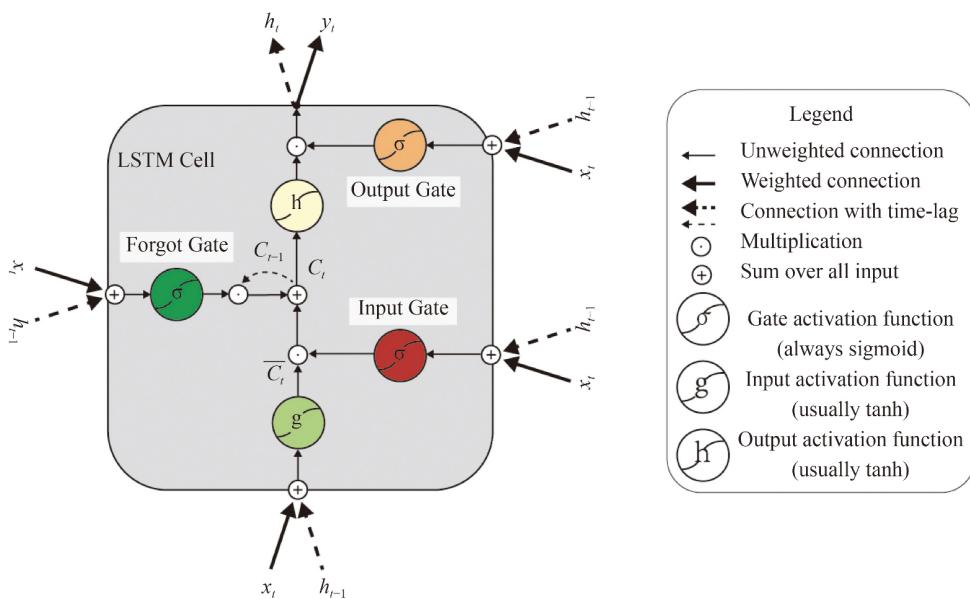


Fig. 2 Structure of a long-short term memory (LSTM) network cell.

short window is used at high frequencies and a long window is used at low frequencies), which allows for explicit capture of the local dynamics within a time series by the parameterized scale function $\phi_{j,k}(t)$ and corresponding mother wave function $\varphi_{j,k}(t)$. By dilation and translation, the explicit representation of the scale function and the mother wavelet can be derived as:

$$\phi_{j,k} = 2^{-\frac{j}{2}} \phi(2^{-j}t - k), \quad (9)$$

$$\varphi_{j,k} = 2^{-\frac{j}{2}} \varphi(2^{-j}t - k), \quad (10)$$

where the subscript j, k represents the amplitude in frequency and the offset in time respectively. Because of orthogonality between $\phi_{j,k}(t)$ and $\varphi_{j,k}(t)$, the wavelet multi-resolution decomposition for an arbitrary time series can be written as:

$$x(t) = \sum_{k \in Z} \alpha_{J,k} \phi_{J,k}(t) + \sum_{j \leq J} \sum_{k \in Z} d_{j,k} \varphi_{j,k}(t), \quad (11)$$

where the first term represents the approximation at the level J , and the second term represents the corresponding detail term on each scale j ; $\alpha_{J,k}$ and $d_{j,k}$ represents the wavelet coefficients of the approximation and the details respectively.

Therefore, we adopt a wavelet multi-resolution decomposition method into the original CNN-LSTM model structure to improve the prediction accuracy of the model. As shown in Fig. 3, in the Wavelet-CNN-LSTM model, the original time series is decomposed into the approximation term cA3 and the detail terms cD1, cD2 and cD3 by the three-level wavelet transform. Each component is reconstructed into a sequence of the same length as the original time series. Then, CNN-LSTM models are constructed for each decomposed subsequence, and the sum of the output from each model is used as the overall output. This model structure for water

demand forecasting is designed to allow for the separation of low-frequency periodic components and high-frequency random components in the original time series, such that the subsequent CNN-LSTM model can more efficiently capture the characteristics of patterns of variation in short-term water demand data.

2.4 Performance indices

Four different performance indices were adopted for a more comprehensive evaluation of model performance, which includes coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and mean absolute percent error (MAPE). These performance indices are defined as follows:

2.4.1 The coefficient of determination (R^2)

$$R^2 = \left(\frac{\sum_{i=1}^n (O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^n (O_i - \bar{O})^2 \sum_{i=1}^n (P_i - \bar{P})^2}} \right)^2, \quad (12)$$

where O_i and P_i represents observation and prediction respectively, \bar{O} and \bar{P} represents the mean of water demand observation and prediction, n is the number of data points. R^2 reveals the degree of correlation between prediction and observation, where 1 represents perfect correlation and 0 represents no correlation at all.

2.4.2 Root-mean-square error (RMSE)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (O_i - P_i)^2}. \quad (13)$$

The smaller the value, the closer the prediction is to the observed value.

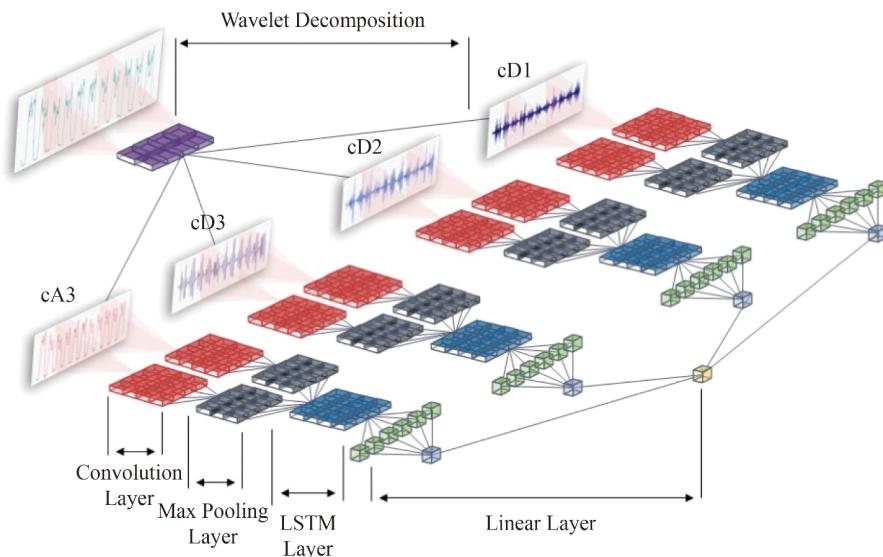


Fig. 3 A flowchart of Wavelet-CNN-LSTM.

2.4.3 Mean absolute error (MAE)

$$MAE = \frac{1}{n} \sum_{i=1}^n |O_i - P_i|. \quad (14)$$

The smaller the value of MAE, the closer the prediction is to the observed value. Meanwhile the value of the mean absolute error has a physical significance and can be combined with actual application scenarios to evaluate the model prediction accuracy.

Mean absolute percent error (MAPE)

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|O_i - P_i|}{O_i}. \quad (15)$$

MAPE represents the relative percentage of the mean absolute error. Importantly, since MAPE is a dimensionless quantity, it can be used to compare the prediction performance between different models.

2.5 Predictability indices

Although the prediction accuracy of water demand forecasting has been increased due to the application of more advanced algorithm and the availability of vast data, there still exists a gap from the theoretical accuracy upper bound (TAUB). In general, the observed data C_o can be considered as having two components, i.e., the interpretable component C_i and the stochastic component C_s , then the TAUB can be defined as follows:

$$TAUB = Metrics(C_o, C_s) \quad (16)$$

where $Metrics(\cdot)$ represents the corresponding error metrics used in model evaluation. This means that the upper bound on the theoretical prediction accuracy can only be approximated if the random components contribute sufficiently little to the model input. But in most studies, the predictability of water demand data is seldom considered as a reason for the bottleneck of prediction accuracy. In this study, we use the approximate entropy (ApEn) to evaluate whether wavelet decomposition can improve the predictability of the data.

Approximate entropy is a statistical measure that quantifies the regularity of time series data (Pincus, 1995). The calculation is as follows:

1) For a timeseries $X(i) : i = 1, 2, \dots, N$, forms $N - m + 1$ sub sequences $X_m(j) : j = 1, 2, \dots, (N - m + 1)$, with a sliding window of length m .

2) Calculate the distance between each sequence and the other k sub-sequences, the distance between two such vectors is defined as the maximum difference of its corresponding element:

$$d_{ij} = \max |x_{k+j}(t), x_{k+i}(t)|, k = 1, 2, \dots, (m - 1). \quad (17)$$

3) Define $C_j^m(r)$ for each $j = 1, 2, \dots, (N - m + 1)$:

$$C_j^m(r) = (\text{number of } X_m(j) \text{ such that } d_{ij} \leq r) / (N - m + 1) \quad (18)$$

where r is a positive real number that represents the

measure of similarity.

4) Define $\varphi^m(r)$ as the average logarithmic conditional over j :

$$\varphi^m(r) = \frac{1}{N - m + 1} \sum_{j=1}^{N-m+1} \log(C_j^m(r)). \quad (19)$$

5) Calculate the approximate entropy is defined as follows Eq. (20):

$$ApEn(m, r, N) = \varphi^m(r) - \varphi^{m+1}(r). \quad (20)$$

For given m , r , and time series $X(i) : i = 1, 2, \dots, N$, the stronger the regularity, the lower the approximate entropy, and vice versa.

3 Case study

3.1 Study area and data description

The study area is in the eastern part of Qingpu District in Shanghai. One water supply plant with a daily capacity of 200000 m³ serves this area. In addition, there are four booster pump stations used for temporary pump scheduling, two of which have a storage capacity of 7500 m³. 24 flow meters are installed in the study area, mainly record the flow and pressure data of the main pipes with a diameter of 500–1000 m; Besides, there are 27 more pressure sensors installed for pressure monitoring of the whole area. A water demand data ranges from 1st October 2019 to 30th October 2021 with a 15-min interval was collected through an SCADA system, which has a total of 70909 observations.

3.2 Model input selection

The short-term water demand forecasting model is a black-box model, of which the principle is to build the mapping between inputs and outputs by linear or nonlinear methods. Although deep learning is capable of automatically encoding features from the input data, selecting appropriate input of model and ensuring that a mapping relationship exists between the inputs and outputs are necessary. Many studies have focused on analyzing the variation patterns in water demand data. For example, Adamowski analyzed summer water consumption data for Ottawa West Centre (OWC) from 1993–2002 and found that there were distinct morning and evening peaks on the daily scale, around 9:00–10:00 a.m. and 20:00–21:00 p.m., respectively (Adamowski, 2008). We have also found the similar pattern in our data set, as shown in Fig. 4. For a deterministic function $z = f(x, y)$, its conditional probability $P(z_i|x_i, y_i)$ equals to 1. As for water demand data, one of the investigations that can be obtained from Fig. 4 is that water demand data has a clear daily pattern and fluctuates within a specific range of values at different times of the day. That is to

say, if the historical water demand is given, we can infer the distribution of future water demand based on the pattern of historical water demand. This inference can be written formally as Eqs. (21) and (22):

$$P(D_t \in I_t | D_{t-1} \dots D_{t-n}) > P(D_t \notin I_t | D_{t-1} \dots D_{t-n}), \quad (21)$$

$$r_p = P(D_t \in I_t | D_{t-1} \dots D_{t-n}) / P(D_t \notin I_t | D_{t-1} \dots D_{t-n}), \quad (22)$$

where D_t represents the water demand at time t ; I_t represents the corresponding water demand interval at time t ; r_p represents odds ratio: the larger the odds ratio, the stronger is the mapping relationship between model inputs and outputs. Therefore, water demand data from the previous day (i.e., $D_{t-1}, D_{t-2}, \dots, D_{t-96}$ with 15 min sampling interval) are considered as model input to ensure a strong mapping relationship and maintain a simple input.

3.3 Model development

The water demand dataset was divided into training set, validation set and test set, where the training set was for fitting the mapping relationships in the training data, the validation set was for verifying the convergence of the model to prevent overfitting and for hyperparameter tuning, and the test set was for evaluating the real prediction performance of the model. Regarding the partitioning of the dataset, a simple approach is to partition all the datasets according to a fixed ratio. However, for time series data, the pattern of variation varies with seasons. In other words, the partitioning method for the dataset mentioned above does not take seasonal factors into consideration, which may have an influence on prediction accuracy. A more reasonable way is to choose a fixed-length training set, but there exists a problem that if the length is too short, the model will be underfit with insufficient samples, while if it is too long, there is no significant difference compared with the fixed-

ratio partitioning method mentioned before. To identify a suitable length of the training set, a parametric grid search method was employed, and the length of 30 d was determined. Based on this, we have divided the entire data set into subsets with the length of 36 d: 30 d for training, 3 d for validation, and 3 d for testing. The corresponding training set was selected automatically by choosing a specific prediction time. Besides, the training set and validation set were shuffled randomly to eliminate the potential impact from temporal order.

In the training process, a mini-batch gradient descent algorithm was used to ensure the convergence stability of the optimization algorithm and avoid excessive computational expenses. Moreover, we have employed the early-stopping strategy, which allows the model to stop training when the minimum loss is achieved on the validation set, thus ensuring the best results of the model training. Then the performance of the model will be evaluated on the test set.

In this study, a total of five models (ANN, Conv1D, LSTM, GRUN and Wavelet-CNN-LSTM) were constructed to allow for a comparison study. Among them, ANN models were widely used for short-term water demand forecasting and can serve as a benchmark to evaluate the prediction performance, while Conv1D, LSTM, and GRUN are deep learning models for water demand forecasting proposed in recent years (Guo et al., 2018; Hu et al., 2019; Mu et al., 2020).

For construction of the neural networks, main hyperparameters include the number of hidden layers, the number of neurons in each layer, the learning rate, the batch size, and the choice of activation function. Based on the previous studies, some classical model structure and hyperparameter config were applied in our model construction. The specific model structures and parameter configs are shown in Table 1.

Besides, the grid search algorithm and cross-validation

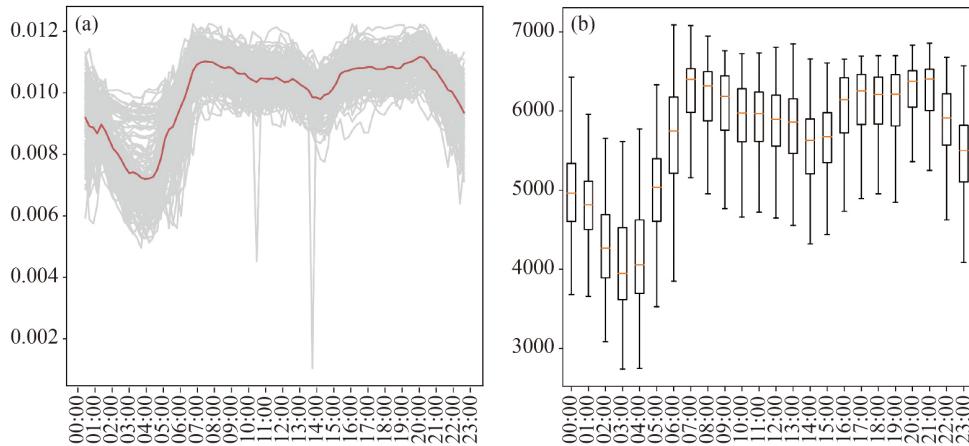


Fig. 4 Water demand distribution on a daily scale (a) Average water demand variation pattern (b) variation interval of water demand in different time periods.

was used to find the optimum hyperparameter values to ensure the model performance on dataset. A series of ranges were specified for the parameters to be optimized, and the parameters were adjusted successively based on the step size. The model was then trained using the adjusted parameters to tune the parameters with the highest accuracy on the validation set. The hyperparameter search ranges and optimization configuration of ANN, Conv1D, LSTM and Wavelet-CNN-LSTM are detailed in [Table 2](#).

Table 1 Model structure and model parameters

Parameter	ANN		Conv1D		LSTM	
	Dense Layer		Convolutional Layer	Dense Layer	LSTM Layer	Dense Layer
Number of layers	3		3	2	1	1
Number of Nodes	32,8,1		16,32,64	512,32	50	50
Learning rate	0.0005			0.0005		0.0005
Optimizer	Adam			Adam		Adam
Batch size	100			100		100
Kernel size				3		

Parameter	GRUN		Wavelet-CNN-LSTM		
	GRU Layer	Dense Layer	Convolutional Layer	LSTM Layer	Dense Layer
Number of layers	3	7	2	1	1
Number of Nodes	48,32,32	64,32,16,8,4,2,1	16,32	50	50
Learning rate		0.0005		0.0005	
Optimizer		Adam		Adam	
Batch size		100		100	
Kernel size				3	

Table 2 The hyper-parameter search range and optimal configuration of models

Model	Hyper parameters	Search range	Optimal configuration
ANN	Learning rate	ranging from 0.01 to 0.0005 with a decrease of 0.0005	0.0005
	Batch size	[50,100,150,200]	100
	Number of layers	[2,3,4]	3
	Number of Nodes(1st)	[48,32,24]	32
	Number of Nodes(2nd)	[16,8,4]	8
Conv1D ¹	Number of layers	[2,3,4]	3
	Number of Nodes(1st)	[4,8,16]	16
	Kernel size	[3,5,7]	3
	Number of Nodes(dense)	[64,32,16]	32
LSTM ²	Number of Nodes	[50,49,48,47,46,32,24]	50
Wavelet-CNN-LSTM	Number of layers	[2,3,4]	2
	Number of Nodes(1st)	[4,8,16]	16
	Kernel size	[3,5,7]	3
	Number of Nodes (LSTM layer)	[50,49,48,47,46,32,24]	50

Note: 1. The number of channels in a convolutional layer is generally the previous layer multiplied by two, so only the number of layers and the number of channels in the first layer need to be determined. 2. A two-layer grid search is used to adjust the number of nodes in the LSTM hidden layer, the first layer [48,32,24] and the second layer [50,49,48,47,46]

4 Results and discussion

4.1 Model performance analysis

To verify the model performance, single-step predictions for the next 15 min and multi-step predictions for the next 1 h tested, and the average results evaluated from 20 randomly selected subsets are summarized for all the five models in [Table 3](#). From the results, it can be observed that all the models could predict the overall trend of the

observations with an acceptable MAPE error (e.g., an MAPE error of less than 3 % for the 15 min predictions, and an MAPE error of less than 5 % for the 1h predictions have been achieved by all these models), but Wavelet-CNN-LSTM outperforms the other four models significantly in terms of all the four prediction indices (e.g., average MAPE 1.25 % in single-step prediction and 2.79 % in multi-step prediction).

In addition, to further analyze the generalization capability of the model, the error distributions of the four indicators in different prediction tasks are shown in Figs. 5 and 6. It can be observed from the figures that Wavelet-CNN-LSTM has better generalization capability compared to the other four models, which can be explained by the tight boxes in the box plots.

Moreover, Figs. 7 and 8 present the predictions verse

observations for the five models applied on the one of test datasets for a more graphic comparison of these difference. Regardless of the 15-min prediction or 1-h prediction, the relative error of the results from the Wavelet-CNN-LSTM model is significantly smaller than those from the other four models when the water demand fluctuates significantly. Besides, the difference between the 1-h prediction and the 15-min prediction is that the Wavelet-CNN-LSTM model have a larger prediction error at some points compared to other models for the 1-h prediction, this is because the low prediction accuracy of the high-frequency components. However, as the prediction accuracy of the low-frequency components increases, the overall model prediction accuracy is still outperformed other models (e.g., the mape error of each model in Fig. 8: ANN-4.35 %, CNN-4.74 %, LSTM-4.83 %,

Table 3 Average performance indicators of prediction models evaluated from 20 testing datasets

Model	Prediction step	R^2	MAPE (%)	MAE (m^3/h)	RMSE (m^3/h)
ANN	15 min	0.79	2.59	133.71	182.24
	1 h	0.63	4.08	210.74	276.01
Conv1D	15 min	0.71	3.05	154.01	197.68
	1 h	0.53	4.67	246.85	310.23
LSTM	15 min	0.85	2.32	118.31	164.89
	1 h	0.62	4.49	231.79	304.13
GRUN	15 min	0.86	2.39	122.72	170.71
	1 h	0.61	4.31	223.54	294.3
Wavelet-CNN-LSTM	15 min	0.93	1.25	62.86	79.23
	1 h	0.71	2.97	150.28	201.13

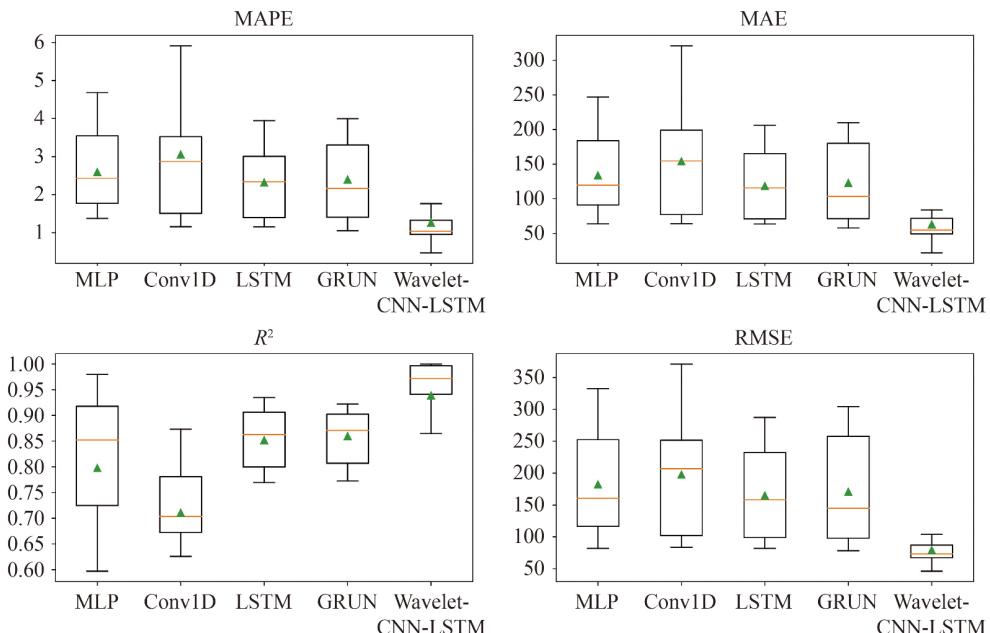


Fig. 5 Box plots of the predication performance indicators for the 15-min prediction (evaluated from 20 randomly selected test sets).

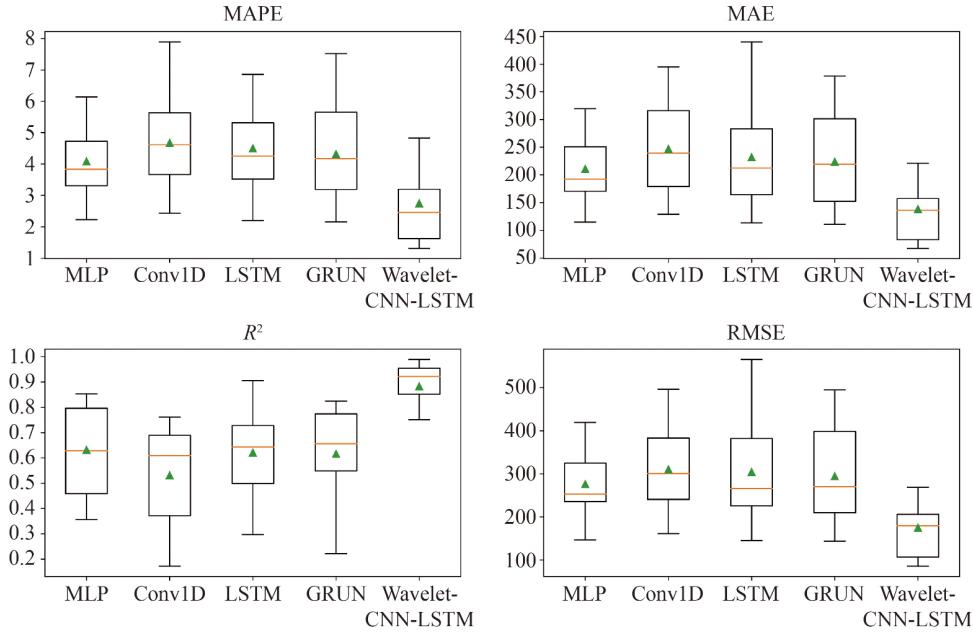


Fig. 6 Box plots of the predication performance indicators for the 1-h prediction (evaluated from 20 randomly selected test sets).

GRUN-4.80 %, Wavelet-CNN-LSTM-4.32 %), and a similar conclusion can be verified in Fig. 6.

4.2 Model mechanism analysis

Although the above-mentioned extensive results have demonstrated the performance and stability of the Wavelet-CNN-LSTM model compared with the other four deep learning models applied in short-term water demand forecasting, it is necessary to explain the reason

behind (i.e., what gains brought by MRA). In fact, the proposed Wavelet-CNN-LSTM model is based on the hypothesis that the noise of random components in the time series can be avoided by modeling the subseries separately after MRA. To prove this hypothesis, we analyzed the predictive performance for each sub model and quantitatively evaluated the data predictability gain by the implementation of MRA using the predictability indices, i.e., approximate entropy. As shown in Fig. 9, the model can capture the trend of the low-frequency term

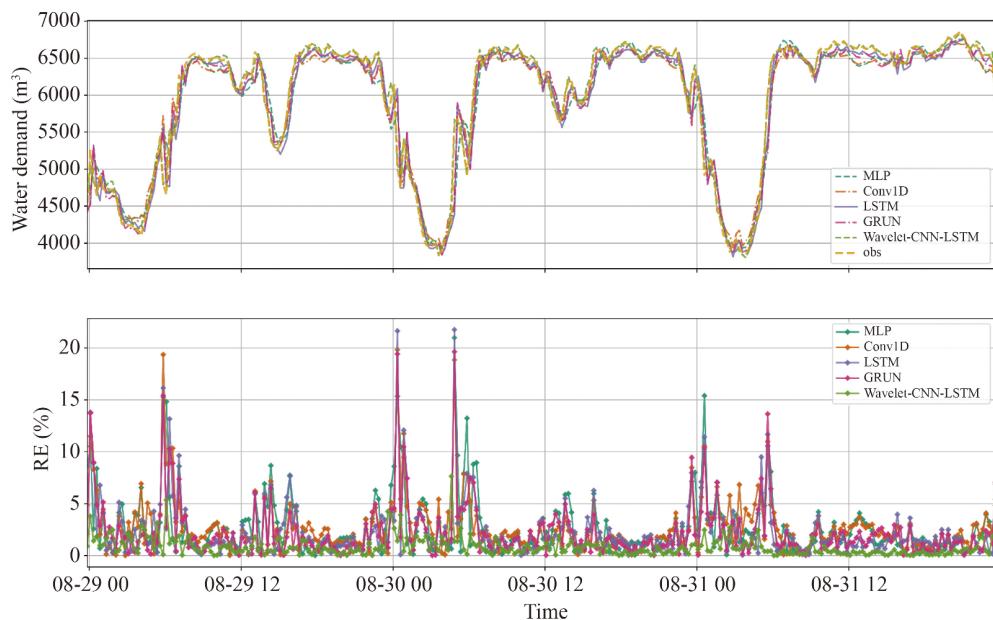


Fig. 7 15-min predictions versus observations for the five models.

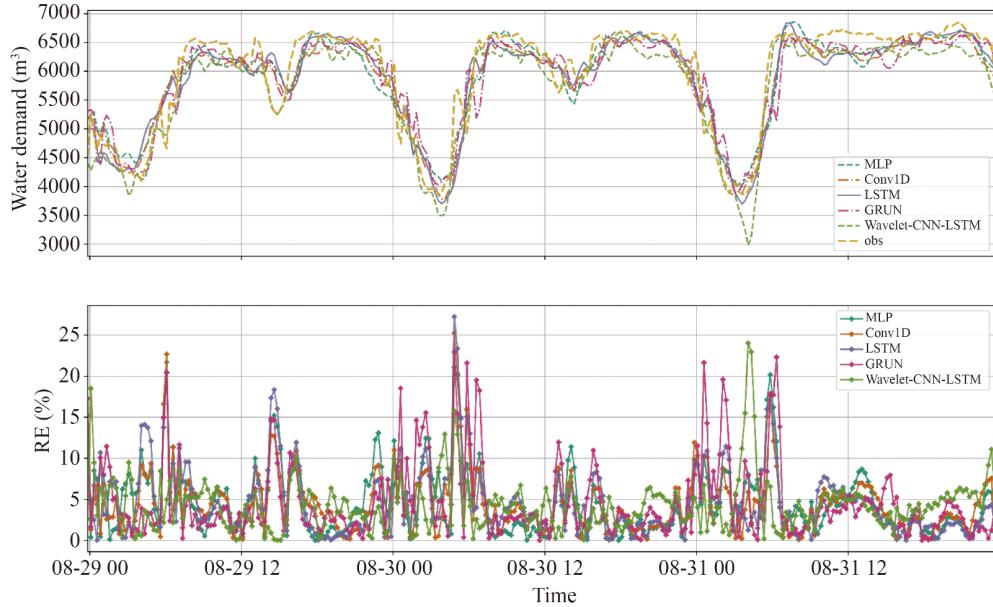


Fig. 8 1-h predictions versus observations for the five models.

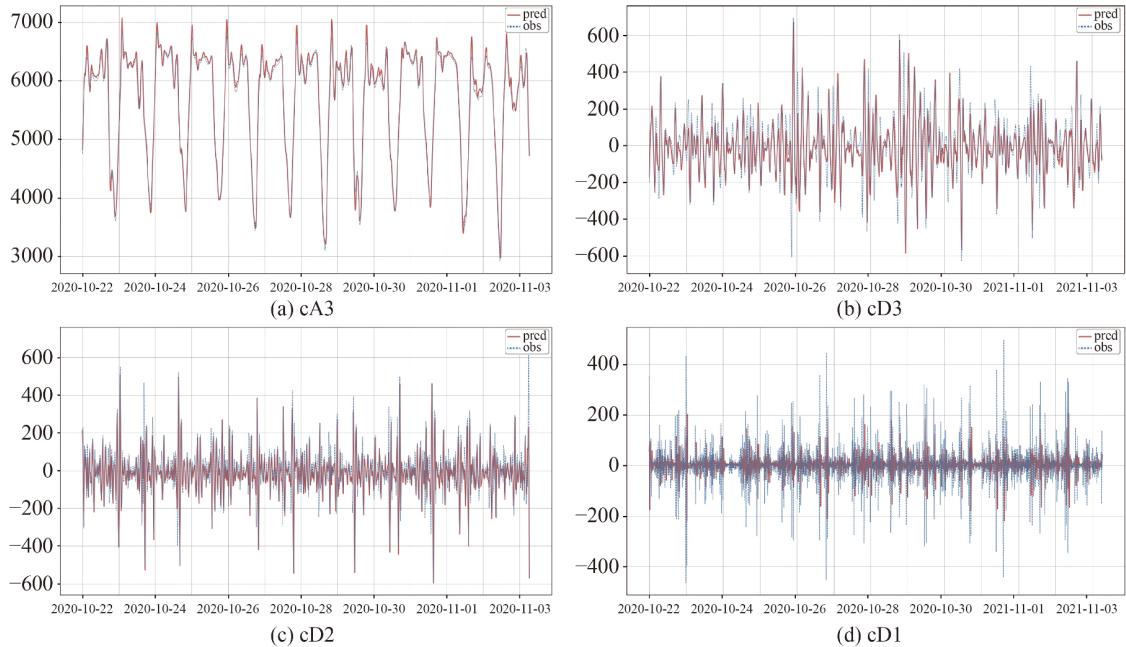


Fig. 9 Sub series model performance at different scales (a) cA3 (b) cD3 (c) cD2 (d) cD1.

cA3 perfectly, but it is powerless for high-frequency term cD3, cD2 and cD1. This fully illustrates that it is very difficult for deep learning models to predict the high-frequency random components in the original time series, thus decreasing prediction accuracy of the models. To obtain a more convincing conclusion, we have randomly selected 20 different test sets and plot the distribution of the predictability indicator (approximate entropy) and the performance indicators (MAPE and MAE) of the subseries at different decomposition scales in Fig. 10.

From Fig. 10, the following conclusions can be drawn: 1) there exists a significant positive correlation between predictability and model performance, which provides the evidence to support the previous conclusions that the predictability of data is an important factor in determining the accuracy of model predictions. 2) Although the prediction accuracy for the low frequency component is very high (with an average MAPE of 1.2 %), it contributes significantly to the overall error (40 m³/h), while the prediction accuracy for the high frequency

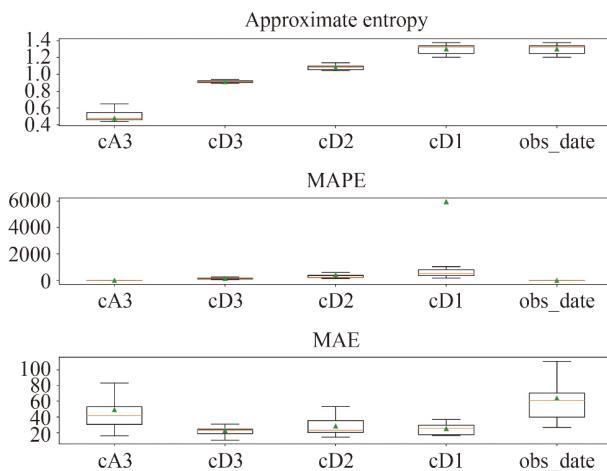


Fig. 10 Predictability indicators and performance indicators distribution at different scales evaluated from the 20 randomly selected test sets

component is very low ($MAPE >> 100\%$), they have less impact on the overall error ($20\text{--}30\text{ m}^3/\text{h}$). This fully demonstrates that the prediction accuracy of low-frequency components is far more important than those of the high-frequency components. The reason why wavelet decomposition can improve the prediction accuracy is that by separating low-frequency and high-frequency terms, it reduces the influence from high-frequency random components, to better capture the long-time dependence in low-frequency components. Although the model has very poor prediction accuracy for the high-frequency components, they do not have a significant effect on the overall prediction results.

5 Conclusions

This paper proposed a novel Wavelet-CNN-LSTM model for short-term urban water demand forecasting. To systematically verify the proposed method, four different deep learning models (ANN, Conv1D, LSTM and GRUN) applied in short-term water demand forecasting were also constructed to be compared with the Wavelet-CNN-LSTM model. The main conclusions obtained as follows:

1) The proposed Wavelet-CNN-LSTM model exhibits better performance than traditional deep learning models in prediction accuracy and stability both in single-step prediction and multi-steps prediction.

2) MRA can produce a remarkable effect on improving the model accuracy by separating the different frequency components in water demand data. Specifically, the lower-frequency component with higher predictable contributes more significantly to the final estimates, while the higher-frequency component with lower predictable contributes less significantly to the final estimates. By

building models for each sub sequence in parallel during prediction, the perturbation from high-frequency random components can be avoided, thus allowing for a better modeling of the longer temporal dependencies in water demand data, and as a result, a higher predication accuracy.

Moreover, more features that are explanatory to the prediction results should be considered to incorporate as input variables. This is because the model is naturally constructed with the assumption that there is a static mapping relationship between the inputs and outputs (Oliveira et al., 2017), but in fact it evolves through time, which is known as concept drift in machine learning. A simple approach to this problem, as mentioned above, is to periodically retrain, but a better choice is to add more explanatory features to ensure that the mapping relationships learned by the model do not fail quickly. Specifically for short-term water demand forecasting, some short-term hydraulics features should be considered (e.g., pressure at the most disadvantageous point). For example, the low-pressure state prior to the morning peak of water consumption can be seen as a signal of future water consumption variability, and those pressure data may have a positive effect on the water demand prediction.

Acknowledgements This work was financially supported by the National Natural Science Foundation of China (No. 51978494), and the Science and Technology Innovation Program Project of Shanghai City Investment Co., Ltd. (No. CTKY-ZDXM-2020-012).

Data Accessibility Statement The data supporting the findings of this study are available within the article and its supplementary materials. The code not available due to intellectual property rights of cooperative institute restrictions.

Electronic Supplementary Material Supplementary material is available in the online version of this article at <https://doi.org/10.1007/s11783-023-1622-3> and is accessible for authorized users.

References

- Adamowski J F (2008). Peak daily water demand forecast modeling using artificial neural networks. *Journal of Water Resources Planning and Management*, 134(2): 119–128
- Amari S I, Wu S (1999). Improving support vector machine classifiers by modifying kernel functions. *Neural networks: the official journal of the International Neural Network Society or Neural Netw*, 12(6): 783–789
- Bedi J, Toshniwal D (2019). Deep learning framework to forecast electricity demand. *Applied Energy*, 238: 1312–1326
- Billings R B, Jones C V (2011). *Forecasting Urban Water Demand*. Washington, DC: America Water Works Association
- Boggess A, Narcowich F J (2015). *A first course in wavelets with Fourier analysis*. 2nd ed. John Wiley & Sons, 183–217
- Bougadis J, Adamowski K, Diduch R (2005). Short-term municipal water demand forecasting. *Hydrological Processes*, 19(1): 137–148

- Bouvie J (2006). Notes on convolutional neural networks. Bracewell R N (1989). The Fourier transform. *Scientific American*, 260(6): 86–95
- Candelieri A, Giordani I, Archetti F, Barkalov K, Meyerov I, Polovinkin A, Sysoyev A, Zolotykh N (2019). Tuning hyperparameters of a SVM-based water demand forecasting system through parallel global optimization. *Computers & Operations Research*, 106: 202–209
- Cao J, Wang J (2019). Stock price forecasting model based on modified convolution neural network and financial time series analysis. *International Journal of Communication Systems*, 32(12): e3987
- Chen G, Long T, Xiong J, Bai Y (2017). Multiple random forests modelling for urban water consumption forecasting. *Water Resources Management*, 31(15): 4715–4729
- Chen J, Boccelli D (2014). Demand forecasting for water distribution systems. *Procedia Engineering*, 70: 339–342
- Cheng J, Liu Y, Ma Y (2020). Protein secondary structure prediction based on integration of CNN and LSTM model. *Journal of Visual Communication and Image Representation*, 71: 102844
- Dara S, Tumma P (2018). Feature extraction by using deep learning: a survey. In: Second International Conference on Electronics 2018, Communication and Aerospace Technology (ICECA), IEEE, Coimbatore, RVS Technical Campus, Coimbatore, India, 1795–1801
- Duerr I, Merrill H R, Wang C, Bai R, Boyer M, Dukes M D, Bliznyuk N (2018). Forecasting urban household water demand with statistical and machine learning methods using large space-time data: a comparative study. *Environmental Modelling & Software*, 102: 29–38
- Firat M, Turan M E, Yurdusev M A (2010). Comparative analysis of neural network techniques for predicting water consumption time series. *Journal of Hydrology (Amsterdam)*, 384(1-2): 46–51
- Ghiassi M, Zimbra D K, Saidane H (2008). Urban water demand forecasting with a dynamic artificial neural network model. *Journal of Water Resources Planning and Management*, 134(2): 138–146
- Goodchild C (2003). Modelling the impact of climate change on domestic water demand. *Water and Environment Journal: the Journal/the Chartered Institution of Water and Environmental Management*, 17(1): 8–12
- Guo G, Liu S, Wu Y, Li J, Zhou R, Zhu X (2018). Short-term water demand forecast based on deep learning method. *Journal of Water Resources Planning and Management*, 144(12): 04018076
- Hafeez G, Alimgeer K S, Khan I (2020). Electric load forecasting based on deep learning and optimized by heuristic algorithm in smart grid. *Applied Energy*, 269: 114915
- Herrera M, Torgo L, Izquierdo J, Pérez-García R (2010). Predictive models for forecasting hourly urban water demand. *Journal of Hydrology (Amsterdam)*, 387(1-2): 141–150
- Hochreiter S (1998). The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-based Systems*, 6(2): 107–116
- Hochreiter S, Schmidhuber J (1997). Long short-term memory. *Neural Computation*, 9(8): 1735–1780
- Hu P, Tong J, Wang J, Yang Y, de Oliveira Turci L (2019). A hybrid model on CNN and bi-LSTM for urban water demand prediction. In: 2019 IEEE Congress on evolutionary computation (CEC), Wellington, New Zealand, 1088–1094
- Huang C W, Chiang C T, Li Q (2017). A study of deep learning networks on mobile traffic forecasting. In: IEEE 28th annual international 2017 symposium on personal, indoor, and mobile radio communications (PIMRC), Montreal, Quebec, Canada, 1–6
- Jowitt P W, Xu C (1992). Demand forecasting for water distribution systems. *Civil Engineering Systems*, 9(2): 105–121
- Kurata G, Ramabhadran B, Saon G, Sethy A (2017). Language modeling with highway lstm. In: IEEE Automatic Speech Recognition and Understanding Workshop (ASRU) 2017, Okinawa Japan, 244–251
- Maidment D R, Parzen E (1984). Time patterns of water use in six Texas cities. *Journal of Water Resources Planning and Management*, 110(1): 90–106
- Mu L, Zheng F, Tao R, Zhang Q, Kapelan Z (2020). Hourly and daily urban water demand predictions using a long short-term memory-based model. *Journal of Water Resources Planning and Management*, 146(9): 05020017
- Nussbaumer H J 1981 The Fast fourier transform. In: *Fast Fourier Transform and Convolution Algorithms*, 80–111
- Oliveira G H, Cavalcante R C, Cabral G G, Minku L L, Oliveira A L (2017). Series forecasting in the presence of concept drift: A pso-based approach. In: IEEE 29th International Conference on Tools with Acritical Intelligence (ICTAI), Boston, USA, 239–246
- Pincus S (1995). Approximate entropy (apen) as a complexity measure—Chaos: an interdisciplinary Journal of Nonlinear Science, 5(1): 110–117
- Peña-Guzmán C, Melgarejo J, Prats D (2016). Forecasting water demand in residential, commercial, and industrial zones in Bogotá, Colombia, using least-squares support vector machines. *Mathematical Problems in Engineering*, 2016
- Sainath T N, Vinyals O, Senior A, Sak H (2015). Convolutional, long short-term memory fully connected deep neural networks. In: 2015 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, Brisbane, Queensland, Australia, 4580–4584
- Sharma S, Sharma S, Athaiya A (2017). Activation functions in neural networks. *Towards Data Science*, 6(12): 310–316
- Sherstinsky A (2020). Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Physica D. Nonlinear Phenomena*, 404: 132306
- Simonyan K, Zisserman A (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:14091556*
- Sønderby S K, Winther O (2014). Protein secondary structure prediction with long short-term memory networks. *arXiv preprint arXiv:14127828*
- Song X, Liu Y, Xue L, Wang J, Zhang J, Wang J, Jiang L, Cheng Z (2020). Time-series well performance prediction based on long short-term memory (LSTM) neural network model. *Journal of Petroleum Science Engineering*, 186: 106682
- Sundermeyer M, Ney H, Schlüter R (2015). From feedforward to recurrent LSTM neural networks for language modeling. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 23(3): 517–529

- Torres J F, Fernández A M, Troncoso A, Martínez-Álvarez F (2017). Deep learning-based approach for time series forecasting with application to electricity load. In: International Work-Conference on the Interplay between Natural and Artificial Computation. Berlin: Springer. 203–212
- Ullah A, Ahmad J, Muhammad K, Sajjad M, Baik S W (2018). Action recognition in video sequences using deep bi-directional LSTM with CNN features. *IEEE Access: Practical Innovations, Open Solutions*, 6: 1155–1166
- Xenocristou M, Kapelan Z, Hutton C, Hofman J (2017). Identifying relationships between weather variables and domestic water consumption using smart metering. *Proceedings of the CCWI*
- Yang J, Li J (2017). Application of deep convolution neural network. In: 14th International Computer Conference on Wavelet Active Media and Technology Information Processing 2017 (ICCWAMTIP), University of Electronic Science And Technology of China, China, 229–232
- Yu B, Yin H, Zhu Z (2017a). Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting. *arXiv preprint arXiv:170904875*
- Yu R, Li Y, Shahabi C, Demiryurek U, Liu Y (2017b). Deep learning: A generic approach for extreme condition traffic forecasting. In: *Proceedings of the 2017 SIAM international Conference on Data Mining*, Houston, Texas, USA, 777–785
- Zen H (2015). Acoustic modeling in statistical parametric speech synthesis-from HMM to LSTM-RNN: Conference: RTTH Summer School on Speech Technology, A Deep Learning Perspective, Barcelona, Spain
- Zhang G P (2001). An investigation of neural networks for linear time-series forecasting. *Computers & Operations Research*, 28(12): 1183–1202
- Zhou S L, McMahon T A, Walton A, Lewis J (2000). Forecasting daily urban water demand: a case study of Melbourne. *Journal of Hydrology (Amsterdam)*, 236(3–4): 153–164