

A hybrid hourly natural gas demand forecasting method based on the integration of wavelet transform and enhanced Deep-RNN model

Huai Su ^a, Enrico Zio ^{b, c}, Jinjun Zhang ^{a*}, Mingjing Xu ^b, Xueyi Li ^a, Zongjie Zhang ^{a, d}

^a National Engineering Laboratory for Pipeline Safety/ MOE Key Laboratory of Petroleum Engineering /Beijing Key Laboratory of Urban Oil and Gas Distribution Technology, China University of Petroleum-Beijing, 102249, Beijing, China

^b Dipartimento di Energia, Politecnico di Milano, Via La Masa 34, 20156, Milano, Italy

^c Chair System Science and the Energy Challenge, Fondation Electricité de France (EDF), CentraleSupélec, Université Paris Saclay, Bâtiment Bouygues - Laboratoire LGI – 2ème étage 3 rue Joliot-Curie 91192, Gif-sur-Yvette cedex, France

^d Petrochina West East Gas Pipeline, Dongfushan Road 458, Pudong District 200122, Shanghai, China

Abstract

The rapid development of big data and smart technology in the natural gas industry requires timely and accurate forecasting of natural gas consumption on different time horizons. In this work, we propose a robust hybrid hours-ahead gas consumption method by integrating Wavelet Transform, RNN-structured deep learning and Genetic Algorithm. The Wavelet Transform is used

* Corresponding author. Address: College of Mechanical and Transportation Engineering, China University of Petroleum, Fuxue Road 18, Changping District 102249, Beijing, China.
Tel.: +86-10-8973 4627; fax: +86-10-8973 4627. E-mail address: zhangjj@cup.edu.cn

to reduce the complexity of the forecasting tasks by decomposing the original series of gas loads into several sub-components. The RNN-structured deep learning method is built up via combining a multi-layer Bi-LSTM model and a LSTM model. The multi-layer Bi-LSTM model can comprehensively capture the features in the sub-components and the LSTM model is used to forecast the future gas consumption based on these abstracted features. To enhance the performance of the RNN-structured deep learning model, Genetic Algorithm is employed to optimize the structure of each layer in the model. Besides, the dropout technology is applied in this work to overcome the potential problem of overfitting. In this case study, the effectiveness of the developed method is verified from multiple perspective, including graphical examination, mathematical errors analysis and model comparison, on different data sets.

Key words: Natural gas demand forecasting; Deep Learning; Recurrent Neural Network; Genetic Algorithm; Long Short Time Memory model

1. Introduction

A rapid increase of demand of natural gas, as an important source of clean energy, is occurring in many countries. The important role of natural gas in the world energy portfolio and the increasing awareness of environment issues have accelerated the development of natural gas industry. The usage of natural gas has penetrated in varies field, e.g., power generation, urban heating supplying, public transportation, manufacturing and so on. On the other hand, the uncertainties in natural gas demand increase the difficulty of management of the gas production and distribution system and the risk of interruption of gas supply, which poses threats on the economy and society [1]. Robust and accurate forecasting of demand of natural gas is one of the

critical problems for maintaining a reliable supply of gas for different applications. In this work, our research focuses on hourly gas forecasting at the customer level. In other words, the forecasting method aims to be used on the customer level, like power plants, factories and distribution companies.

Efforts have been made to forecast well the gas demand. The literature surveys [2], [3] indicate that the exploration of natural gas forecasting can be divided according to different rules, such as forecasting horizon, forecasting tools, data type and applied area. For different types of forecasting work, i.e., different horizons and applied area, the methods used are different. According to Reference [2], the forecasting horizon can be hourly, daily, monthly, annually and combined. The applied area can refer to world level, national level, regional level, gas distribution system level and individual customer level. Methods to forecast natural gas demand should be carefully selected to fit the specific conditions of the forecasting problem. According to the literature research, the methods for gas demand forecasting can be mainly grouped as time series model, regression model, artificial neural network and hybrid method [4]. Generally, the choice of the forecasting method depends on the forecasting scenario and the type of input data.

Time series (TS) models are used to forecast the gas demand based on the collected data, without prior knowledge [4]. The reports in literature indicate that TS models can be applied for a wide range of forecasting horizons (from annual to hourly). ARIMA model was applied to forecast annual or monthly gas demand of Turkey, with the consideration of GDP and price of gas [5]. A forecasting model based on SARIMAX was developed for short-term prediction of the daily gas demand, with the consideration of temperature, pressure, humidity and cloudiness [6]. Structural time series model was used to forecast the annual gas demand considering multiple factors, such

as future trend of natural gas consumption, determinants income and natural gas price [7]. The results of literature indicate that SARIMA and SARIMAX have better performances in capturing seasonal factors in the time series of the demand than ARIMAX, which is able to provide qualified annual forecasting for demand of gas [4].

Besides time series models, regression models are also widely used for natural gas demand forecasting. Generally, linear regression models are preferred for long-term horizons, country level forecasting, based on some main independent factors [8], [9]. For example, linear regression was applied to forecast annual natural gas demand in South Korea with four variables: population, GDP, export and import amounts [10]. Linear regression model considering temperature, GDP per capita and natural gas price was applied for long-term forecasting of natural gas demand at the country level [11]. Besides linear regression models, the OLS (ordinary least squares) regression model [6] and some nonlinear-regression based statistical methods [12] have shown good abilities for natural gas demand forecasting at different levels.

In recent years, a number of forecasting models based on artificial neural networks have been developed and these models significantly improve the accuracy and efficiency of natural gas demand forecasting. Feedforward neural network, fuzzy neural network, recurrent neural network or some hybrid neural networks have been applied at different horizons and levels [13]–[15]. The comparison of the forecasting results show that the neural network-based models have strong abilities for natural gas demand forecasting [16]. For example, reference [17] indicated that the developed neural network model outperforms the condition demand analysis method and the engineering model. The comparison carried out in [10] showed that the developed multilayer perceptron model has better prediction performance than a linear regression model and an

exponential model.

Among the neural network models, recurrent neural network (RNN) models, which process data by internal memory loops and maintain a chain-structure, are more suitable to learn the features of time series data [18], [19]. However, the deep chain-like structure increase the difficulty of training RNN models by backpropagation. This is not the case for long-short time memory (LSTM) model [20], whose great power of LSTM mode for analyzing data sequence has been proved by successful applications in many areas, including speech recognition [21], human trajectory prediction [22], traffic prediction [23], etc. However, the potential advantages of LSTM model for forecasting are far from being totally exploited because most of LSTM-based forecasting models are shallow-structured and the knowledge embedded in the data can not be fully learned [24]. Also, because time series data are fed chronologically to a LSTM model, the information is passed in forward direction along the chain-like structure and the traditional LSTM model can learn only the forward relationship in the data. This results in the fact that LSTM models may filter out valuable information of backward dependencies of data. This is quite relevant for the case that of interest because, generally, demands of natural gas have relatively strong periodicity and regularity, which means that backward temporal dependencies constitute an important part of the natural gas demand pattern.

Various works have explored the abilities of hybrid forecasting models. In general, hybrid forecasting models have better performance in flexibility, robustness, computing efficiency than most individual forecasting methods. For example, in reference [25], different methods including wavelet transform, genetic algorithm, adaptive neuro-fuzzy inference system and feedforward neural network were integrated for day-ahead demand forecasting in Greece.

According to the above literature survey, we can tentatively conclude that the critical part of demand forecasting relates to finding the inherent features and relationships hidden in the natural gas demand data. In this paper, we originally develop a robust natural gas demand forecasting model by integrating Wavelet Transform, stacked Bi-Directional LSTM model, Genetic Algorithm (GA) and LSTM model. The wavelet transform is used to decompose the original demand data, to reduce the difficulties to learn the relationships in data. The stacked Bi-Directional LSTM model can comprehensively learn the inherent knowledge from both forward and backward directions in each decomposed component. GA method is used to optimize the structure of the stacked Bi-Directional LSTM model to improve its performance of feature learning. Finally, the LSTM model with a dropout part is applied for hourly demand forecasting based on chronologically-arranged data. This hourly forecasting method can contribute to fulfill research and application for predictive optimization of natural gas pipeline networks and real-time demand side management in gas supply systems.

The main contribution of this work is summarized as follows:

(1) This paper proposes a hybrid model, which is able to effectively learn the knowledge in natural gas demand data and make accurate forecasting, with high efficiency. Natural gas consumption data has the characteristics of series data, but these have not been paid enough attention for gas demand forecasting. Under this consideration, the deep RNN model for time series data analysis is originally used for natural gas demand data forecasting in this work. To the best of the authors' knowledge, this is the first time that deep recurrent neural networks are used for natural gas demand forecasting.

(2) This paper originally proposes the idea of bi-direction feature learning, to fundamentally

improve the performance of natural gas demand forecasting by effectively mining the features from the sequential structured consumption data. The results prove that considering the backward relationship can significantly improve the accuracy of forecasting and shorten the model training process.

2. Development of natural gas demand forecasting method

For a clear illustration, the developed forecasting method is introduced in three parts: data decomposing part, data feature learning part and demand forecasting part.

2.1 Decomposition of the original time series of natural gas demand

Data pre-processing is usually performed to improve the performances of prediction. Most of the time series data of natural gas demand contain trends and volatilities, which means forecasting difficulty. To handle this difficulty, wavelet transform is applied to decompose the original time series of gas demands into several high-frequency and one low-frequency subseries in the wavelet domain, in order to reduce the difficulty of feature learning and improve the forecasting performance [26], [27].

A wavelet transform can provide useful information in both time and frequency, especially when the time series is non-stationary, like natural gas demand. Generally, wavelet transforms can be classified in Discrete Wavelet Transform (DWT) and Continuous Wavelet Transform (CWT). The CWT is performed based on continuously scaling and translating the mother wavelet, which leads to a great amount of redundant information. DWT samples the coefficients to reduce the information redundancy[25]. In the DWT method, the coefficient $W(a, b)$ is:

$$W(a, b) = 2^{-\left[\frac{a}{2}\right]} \sum_{t=0}^{T-1} f(t) \phi\left(\frac{t-b2^a}{2^a}\right) \quad (1)$$

where f denotes the original series; ϕ denotes the mother wavelet; T denotes the length of the original series; t represents the index of the discrete time. A fast DWT [28] is used here which contains four filters: decomposition high-pass filter, decomposition low-pass filter, reconstruction high-pass filter and reconstruction low-pass filter. The Fig. 1 presents the process in which the original series can be successively decomposed into lower resolution components. In this paper, we perform a 3-level wavelet decomposition of the original natural gas demand series using the order 5 Daubechies wavelet. The original series can be obtained by reversely summing the high-frequency components (d1-d3) and the low-frequency component a3.

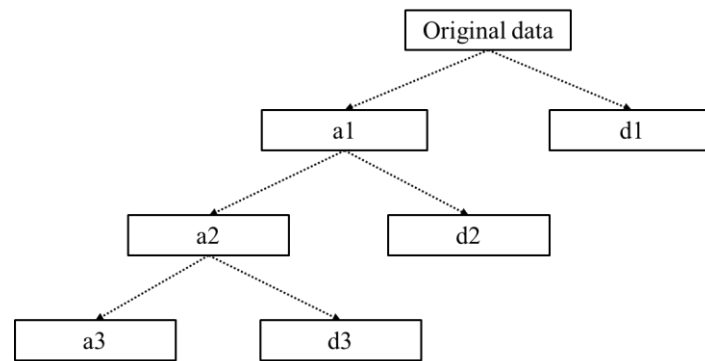


Fig. 1 Multi-level wavelet decomposition

2.2 The deep forecasting model based on bidirectional LSTM and LSTM model

2.2.1 Bidirectional LSTM and LSTM model

In recent years, deep learning has showed great powers in many applications. As a representative deep learning method, the great power of LSTMs for sequence data processing has been proved by its successful application in many real-world problems [29]. Many results have shown that LSTM works well on sequence data with long time dependencies [24], [30]. The structure of a LSTM memory cell is shown in Fig. 2. The self-loop in the cell makes it able to store temporal information encoded into the state of this cell.

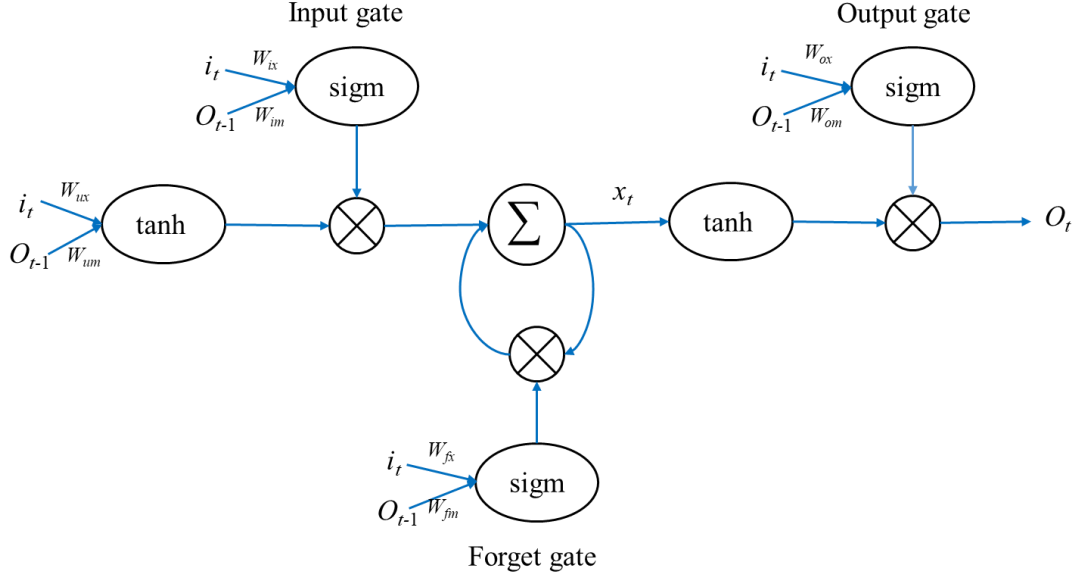


Fig. 2 The structure of a LSTM memory cell

To overcome the training problem of exploding/vanishing in traditional RNN models, three types of operations are supplemented in the LSTM cell, including reading, writing and erasing [31]. These operations are carried out by the output gate, input gate and forget gate, respectively. For example, the input gate is able to decide whether the updated data should modify the memory state by applying an activation function, which works as a switch depending on the previous output and the current input. The memory state will not be affected by the updated data if the related input gate value is close to zero. The mathematical representations of the operation in the LSTM cell are introduced by the following equations:

$$i^g = \text{sigm}(i_t W_{ix} + O_{t-1} W_{im} + b_i) \quad (2)$$

$$f^g = \text{sigm}(i_t W_{fx} + O_{t-1} W_{fm} + b_f) \quad (3)$$

$$O^g = \text{sigm}(i_t W_{ox} + O_{t-1} W_{om} + b_o) \quad (4)$$

$$u = \text{tanh}(i_t W_{ux} + O_{t-1} W_{um} + b_u) \quad (5)$$

$$x_t = f^g \circ x_{t-1} + i^g \circ u \quad (6)$$

$$O_t = O^g \circ \text{tanh}(x_t) \quad (7)$$

where o^s represents the output gate, i^s represents the input gate, f^s represents the forget gate, x_t denotes the state at time step t , u represents the update signal and O_t is the output of the cell.

An efficient forecasting method of natural gas demands need to comprehensively capture their features, especially the regularity and periodicity. However, the LSTM model can only learn forward dependency of arranged sequence data, which means that valuable information on backward relationships is ignored. In this research, the bidirectional LSTM (Bi-LSTM) model is applied to perform feature-learning by considering the dependencies of demand data from both forward and backward directions [32].

The structure of the bidirectional LSTM model is presented in Fig. 3. The mechanism of the Bi-LSTM mode can be interpreted as the combination of a forward LSTM and a backward LSTM, which are used to process the time series data in positive time sequence (from t_0 to t_n) and reversed sequence (from t_n to t_0), respectively. In the two LSTMs of different directions, the outputs obtained based on the Equations 8. Then, these output are combined into the outputs of the Bi-LSTM model:

$$Y_t = G(\bar{O}_t, \bar{O}_t) \quad (8)$$

where function G is used to generate new outputs based on the outputs of the forward LSTM and the backward LSTM. The type of function G can be multiplication function, average function, summation function and so on, and should be selected based on the problem and the data.

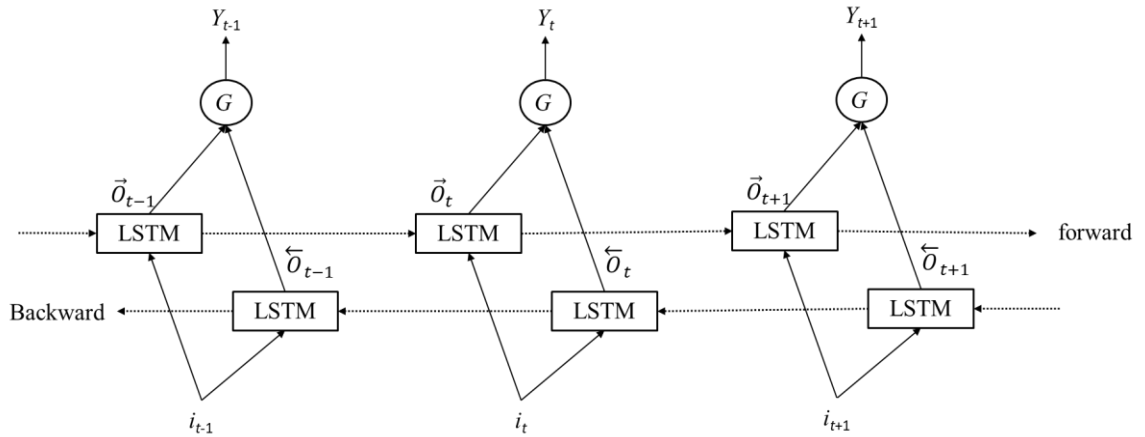


Fig. 3 Illustration of the unfolded structure of a Bi-LSTM model

2.2.2 Stacked Bi-LSTM and LSTM model

RNN models have been used in many real world forecasting problems. Most of the proposed models are developed based on shallow structures with one hidden layer [33]. Recent studies indicate that deep-structured RNNs with several hidden layers can be very effective in sequence data learning [34]. Deep RNN architectures (Fig. 4) are built up by stacking several RNN neural networks together, in which the output of the former RNN is fed to the subsequent layer as input. Such type of deep structure is here used to improve the ability of feature learning and forecasting of natural gas demands.

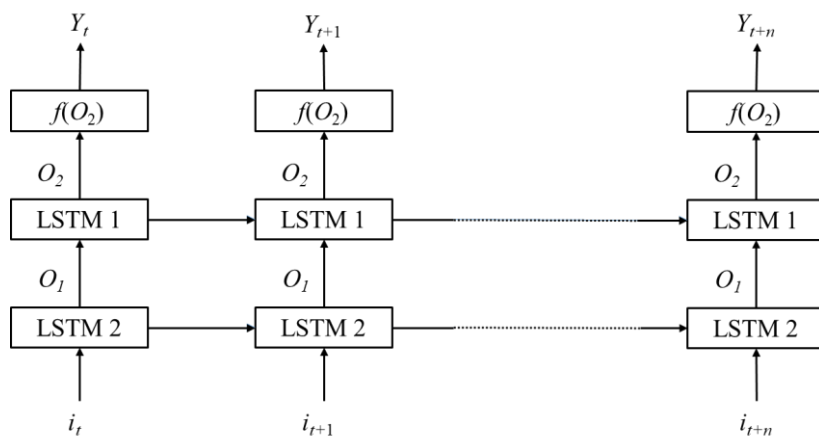


Fig. 4 Illustration of the deep structure RNN model

To comprehensively learn the complex features of natural gas demand data, several Bi-LSTM

layers are used. In this feature learning process, the characteristic of Bi-LSTM can help to effectively capture the information of sequence data and learn the complex relationships between the decomposed gas demand data and the selected factors (for example, calendar and weather conditions). On the top of the deep architecture, a LSTM layer is used to predict the future value along the forward direction based on the learned features from the stacked Bi-LSTM layers. Finally, the forecasting method is built up by combining the wavelet transform and the developed deep RNN model.

2.2.3 Enhancement of the performance of the deep forecasting model

The developed deep-structure network model is capable to learn the complex relationships of the data and the selected factors, but many of these so-called relationships may be just noises. In this condition, overfitting would be a serious problem, threatening the forecasting accuracy. To address this problem, the dropout technique is introduced in this model. The dropout technique can efficiently prevent overfitting by randomly dropping a unit [35]. With reference to the generic layer, classical dropout mechanism can be explained by the following equations:

$$r^{(l)} \sim \text{Bernoulli}(p) \quad (9)$$

$$y^{(l)} = r^{(l)} * y^{(l)} \quad (10)$$

$$z_i^{(l+1)} = w_i^{(l+1)} y^{(l)} + b_i^{(l+1)} \quad (11)$$

$$y_i^{(l+1)} = f(z_i^{(l+1)}) \quad (12)$$

where * represents an element-wise product. For the specific layer l , $r^{(l)}$ denotes a vector of independent Bernoulli random variables, whose probability of being equal to 1 is p . This vector is randomly sampled and, then, multiplied element-wise by the outputs of the layer ($y^{(l)}$), to obtain the “thinned” outputs $\hat{y}^{(l)}$. Then the outputs $\hat{y}^{(l)}$ are fed into the next layer. The dropout operation

is repeated at each layer.

It is necessary also to optimize the architecture of the proposed deep-learning model, i.e., the number of neurons in each layer, to enhance its forecasting performance[36]. In this work, we used Genetic Algorithm, which has been effectively employed to solve many kinds of optimization problems [25], [37]–[39]. In this work, the GA minimizes the Root Mean Square Error (RMSE) between the forecasted natural gas demands \mathbf{F} and the real natural gas demand data \mathbf{T} . The optimization problem is formulated by Equations 13-14 below. The framework of the deep forecasting method and the enhancements are shown in Fig. 5.

$$\text{Minimize : } RMSE(\mathbf{F}, \mathbf{T}) \quad (13)$$

$$\mathbf{F} = P(\mathbf{D}; M_1, M_2, M_3, M_4, N_1, N_2, N_3, N_4) \quad (14)$$

in which P represents the deep forecasting model needing to be optimized. \mathbf{D} represents the vector of the input data of natural gas, M_i denotes the vector of the decision variables which determine the structure of the deep RNN model for the component i , N_i denotes the number of neurons of the LSTM model for the component i , T represents the real natural gas demand data. The ranges for the search of the decision variables are pre-defined.

The prediction model P is composed by four major sub-models, including Discrete Wavelet Transform model DWT (shown in Equation 1), stacked Bidirectional LSTM model Γ^* (shown in Equation 2), LSTM model Γ (shown in Equation 3) and Inverse Wavelet Transform model InverseWT (shown in Equation 4).

DWT model takes the input \mathbf{D} and provides in output the decomposed signal components $\mathbf{d1}$, $\mathbf{d2}$, $\mathbf{d3}$ and $\mathbf{a3}$, as follows:

$$[\mathbf{d1}; \mathbf{d2}; \mathbf{d3}; \mathbf{a3}]_{4 \times N} = DWT(\mathbf{D}) \quad (15)$$

where $\mathbf{d1}$, $\mathbf{d2}$, $\mathbf{d3}$, $\mathbf{a3}$ and $\mathbf{D} \in \mathbf{R}^{1 \times N}$.

The stacked Bi-LSTM model $\Gamma^{*(j)}$ takes in input the decomposed signal component $\mathbf{i}^{(j)}$, outputs the feature $\mathbf{Y}^{*(j)} \in \mathbf{R}^{1 \times N}$, with the number of neurons in the Bi-LSTM layers given by \mathbf{M}_j .

$$\mathbf{Y}^{*(j)} = \Gamma^{*(j)}\left(\mathbf{i}^{(j)}; \mathbf{M}_j\right) \quad (16)$$

where j denotes the index of the decomposed signal component obtained by DWT in Eq. 15, $j=1,2,3,4$, corresponding to the components $\mathbf{d1}$, $\mathbf{d2}$, $\mathbf{d3}$, $\mathbf{a3}$, $\mathbf{i}^{(j)} \in \mathbf{R}^{1 \times N}$ denotes the input, i.e., the j -th decomposed signal component, $i \in \{\mathbf{d1}, \mathbf{d2}, \mathbf{d3}, \mathbf{a3}\}$, \mathbf{M}_j denotes the number of neurons in the Bi-LSTM layers.

The LSTM model $\Gamma^{(j)}$ takes in input the output $\mathbf{Y}^{*(j)}$ of the stacked Bi-LSTM model $\Gamma^{*(j)}$ and provide in output the prediction of the decomposed signal component $\mathbf{Y}^{(j)}$, with the number of neurons in the LSTM layer N_j given by:

$$\mathbf{Y}^{(j)} = \Gamma^{(j)}\left(\mathbf{Y}^{*(j)}; N_j\right) \quad (17)$$

where j denotes the index of the decomposed signal component obtained by DWT, $j=1,2,3,4$, corresponding to the component $\mathbf{d1}$, $\mathbf{d2}$, $\mathbf{d3}$, $\mathbf{a3}$.

The *InverseWT* model takes in input all the predictions of the decomposed signal component and provides in output the prediction of the original data of natural gas demand \mathbf{D} :

$$\mathbf{F} = \text{InverseWT}\left(\mathbf{Y}^{(1)}, \mathbf{Y}^{(2)}, \mathbf{Y}^{(3)}, \mathbf{Y}^{(4)}\right) \quad (18)$$

Finally, let $DWT(D)^{(j)}$ ($j=1,2,3,4$) denotes the decomposed signal component $\mathbf{d1}$, $\mathbf{d2}$, $\mathbf{d3}$, $\mathbf{a3}$, respectively. We get the detailed mathematical presentation of Eq. 14 as follows:

$$\mathbf{F} = \text{InverseWT}\left(\Gamma^{(1)}\left(\Gamma^{*(1)}\left(DWT(\mathbf{D})^{(1)}; M_1\right); N_1\right), \dots, \Gamma^{(4)}\left(\Gamma^{*(4)}\left(DWT(\mathbf{D})^{(4)}; M_4\right); N_4\right)\right) \quad (19)$$

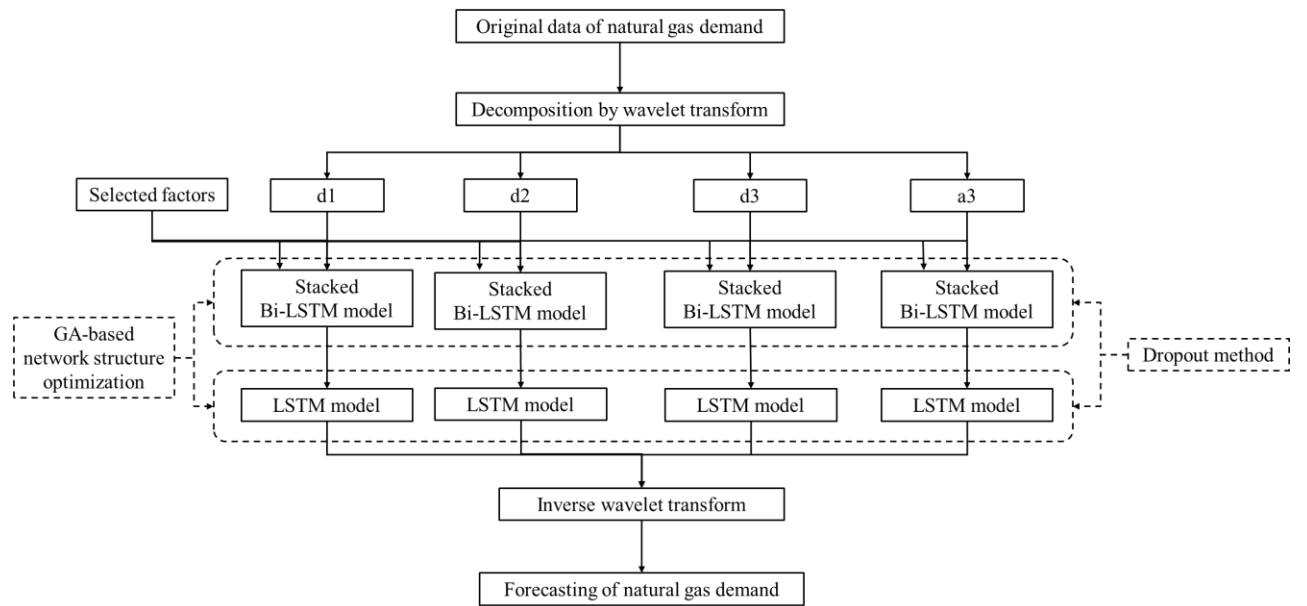


Fig. 5 Schematic representation of the integrated method with performance enhancement

To illustrate the operation process of the forecasting method, the flowchart is shown in Fig. 6:

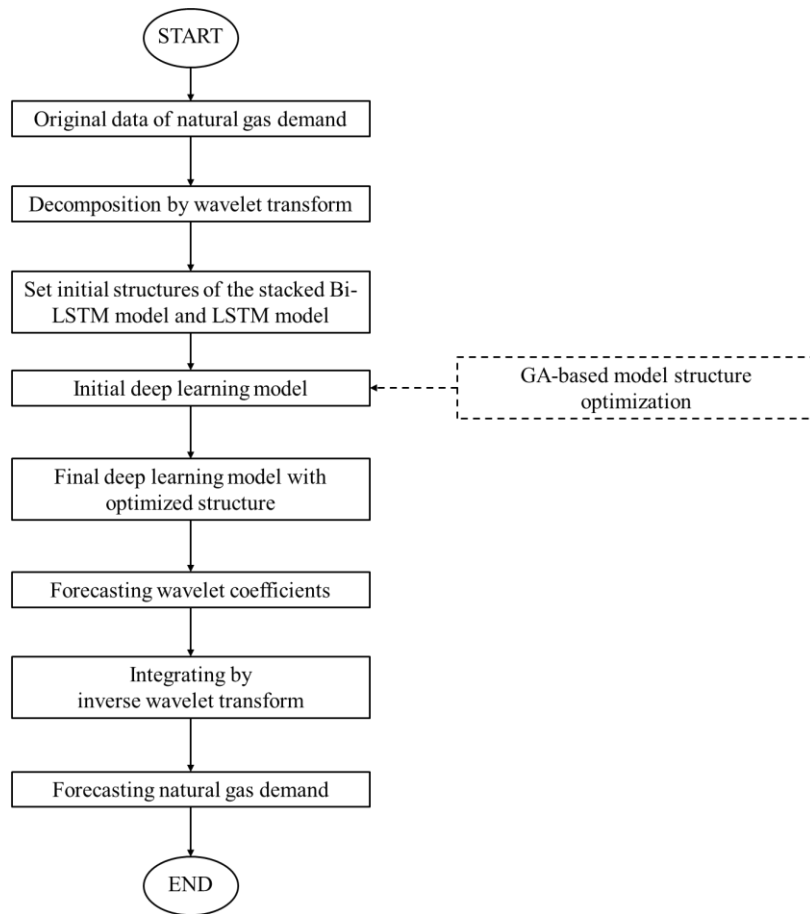


Fig. 6 The flowchart of the forecasting method

3. Applications

3.1 Preparation

The developed forecasting model is applied to two types of data. One type of data is artificially generated by the Mackey-Glass time-series model (Equation 15), which is often used to verify the effectiveness of prediction models because of its chaotic and periodical characteristics [40]:

$$\frac{dx(t)}{dt} = \frac{ax(t-\tau_M)}{1+x^c(t-\tau_M)} - bx(t) \quad (20)$$

where the parameters are set to be $a=0.2$, $b=0.1$, $c=10$, by trial and error, to generate a pattern of data similar to the fluctuation of gas consumption. The parameter τ_M determines the chaotic property of the defined time series. In this work, the value of τ_M is set to be 20. To simulate the time series, the 4th Runge-Kutta method is used here. Then the generated data is sampled at the interval of one hour. Further, to better test the predictive ability, a random term (1% of the value of the generated data) is introduced.

As a second application, data of natural gas consumption is taken from OpenEI, a platform set up by the United States Department of Energy, providing structured energy information of different sectors.

To quantify the performance of the forecasting method, four criteria are used here, which are Mean Absolute Error (MAE), Relative Error (RE), Mean Relative Error (MRE) and Root Mean Squared Error (RMSE):

$$MAE = \frac{1}{N} \sum_{i=1}^N |F_i - T_i| \quad (21)$$

$$RE = \sum_{i=1}^N \left| \frac{F_i - T_i}{T_i} \right| \quad (22)$$

$$MRE = \frac{1}{N} \sum_{i=1}^N \left| \frac{F_i - T_i}{T_i} \right| \quad (23)$$

$$RMSE = \frac{1}{N} \sqrt{\sum_{i=1}^N (F_i - T_i)^2} \quad (24)$$

where F_i denotes the forecasting results and T_i denotes the target value.

Before training the model, we need to standardize the wavelet components:

$$s = \frac{c - \mu}{\sigma} \quad (25)$$

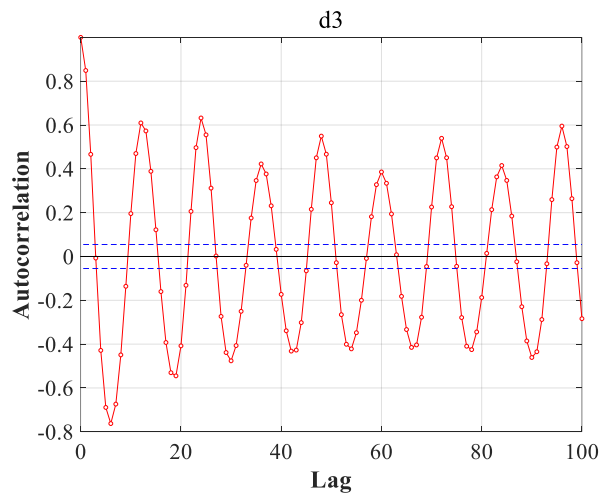
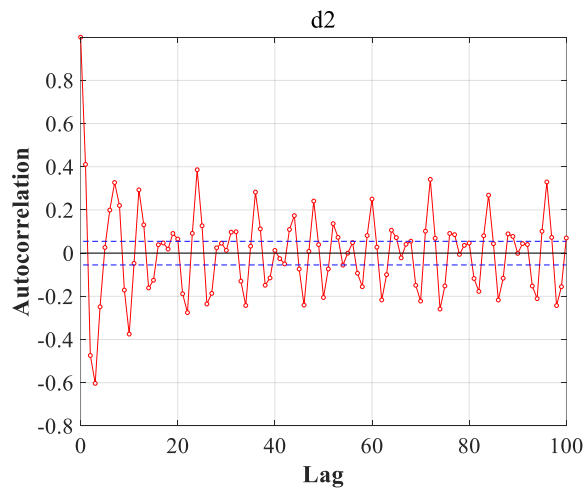
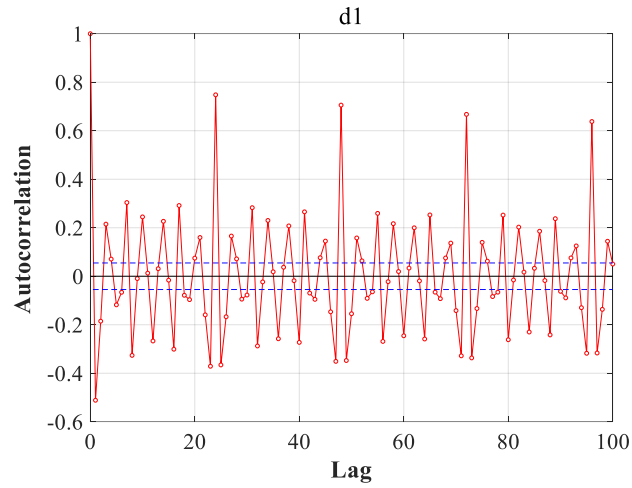
where s represents the standardized value of component c ; μ represents the mean value of c ; σ denotes the standard deviation.

The selection of the input size is important for the success of the prediction. The Pearson correlation coefficient is often used to explore the periodicity of the natural gas consumption [25]; this coefficient has a good ability to research the linear correlation of data. However, the regularities behind the data of natural gas consumption are much more complicated than linear correlation. Following the literature [41], the autocorrelation function (ACF) is used here to describe the correlation for time series data:

$$ACF(Y, k) = \frac{\frac{1}{T} \sum_{t=1}^{T-k} (y_t - \bar{y})(y_{t+k} - \bar{y})}{\sigma_y} \quad (26)$$

where \bar{y} is the mean value of the time series Y ; k denotes the lag of forecasting; σ_y is the variance of the time series. The calculation of ACF is performed for every wavelet component and the input sizes of the components are determined by the lengths of their first period (Figure 7). For example, the values of ACF of wavelet components of one group of data collected on OpenEI indicate that the input sizes for the wavelet components should be longer than around 25 hours (Figure 7). In the cases based on the real world data, the available information of date and time

(hour) is also chosen as another input of the forecasting model and some other important information, such as weather, climate and gas price, should also be carefully selected as the inputs of gas consumption forecasting.



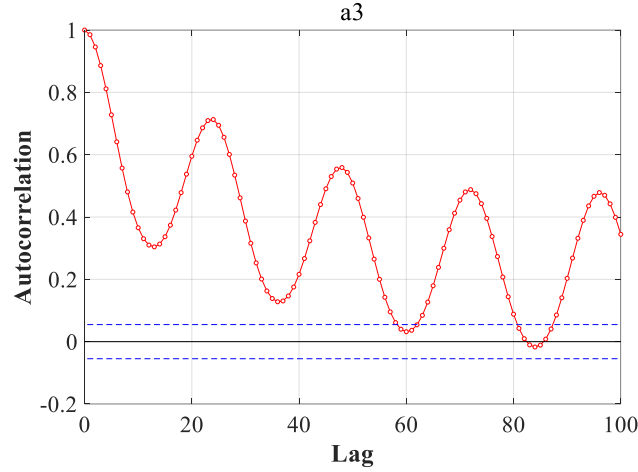


Fig. 7 ACF analysis for different wavelet components

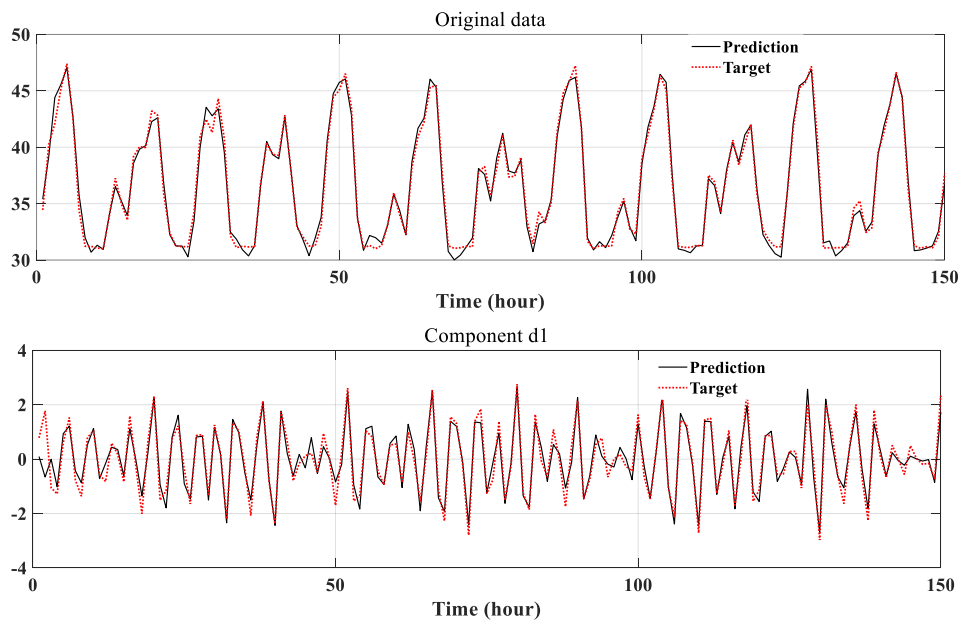
3.2 Performance evaluation

For each deep RNN model, the number of Bi-LSTM layer is set to 2, by trial and error. The numbers of neurons in the Bi-LSTM layers and the LSTM layer are optimized by GA in a search within the range of [100, 110, 120, 130, 140, 150], to find good structures for each RNN layer. The main parameters of the GA are set by trial and error as: population size=30, crossover probability=0.4 and mutation probability=0.6. For the application of the GA, one needs to trade off the performance improvement and the computational burden for the optimality search. For this reason, the number of generations in the GA is set to 100, leading to a significant improvement of the forecasting accuracy with acceptable computation time. Considering that, the maximal number of generations in the GA are set by 100, which can effectively improve the forecasting accuracy under acceptable computation time consuming. During the training process, the maximum number of epochs is 600. The dropout rate is set by 0.4, to avoid the overfitting problem.

The developed forecasting model is used on the simulated data by the Mackey Glass model and to predict the gas consumption based on two sets of real world data (named as Set I and Set II). For these latter, only the consumption data of winter are used. Considering this research focuses

on hourly gas load forecasting, the prediction time interval of these presented applications is chosen as 10 hours. For the three different forecasting tasks, the structures of the deep RNN models, each of which includes four stacked Bi-LSTM layers and four LSTM layers, are optimized by the GA. The optimization variables are the numbers of neurons in every layer of the deep RNN model. For example, in the model for forecasting the simulated data, the optimized numbers of the neurons of the two-layer Bi-LSTM parts and the LSTM parts for the wavelet components of d1, d2, d3 and a3 are [130, 120, 130], [130, 110, 100], [120, 130, 120], [130, 110, 100], respectively. During the training processes of the deep RNN models, the developed models show good convergence rates. The normalized RMSEs of the training processes are lower than 0.1 after around 100 epochs. Some oscillation are observed because of the dropout technique.

Both the prediction results of the original series and the wavelet components are shown in the Figs 8-10, to give a comprehensive picture of the performances of the forecasting model.



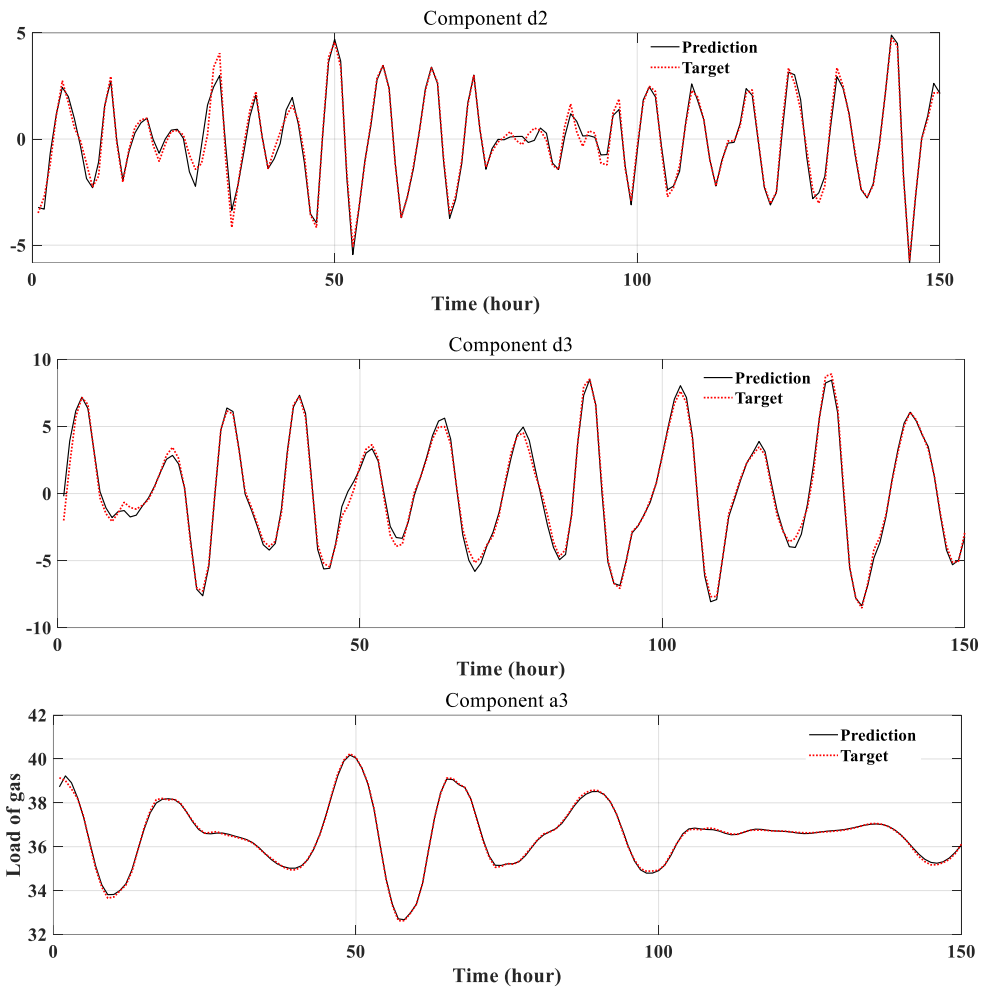
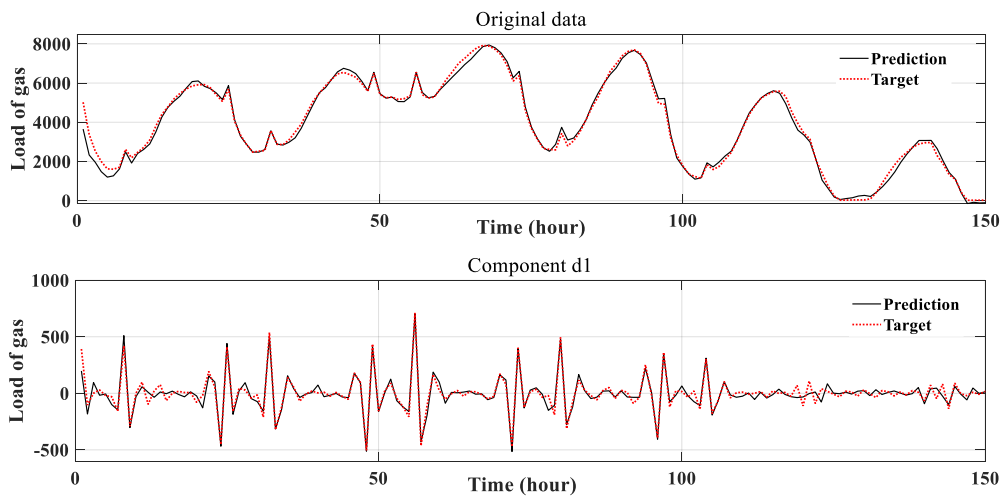


Fig. 8 “Predicted demands” compared with “actual demands” on data generated by Mackey- Glass model



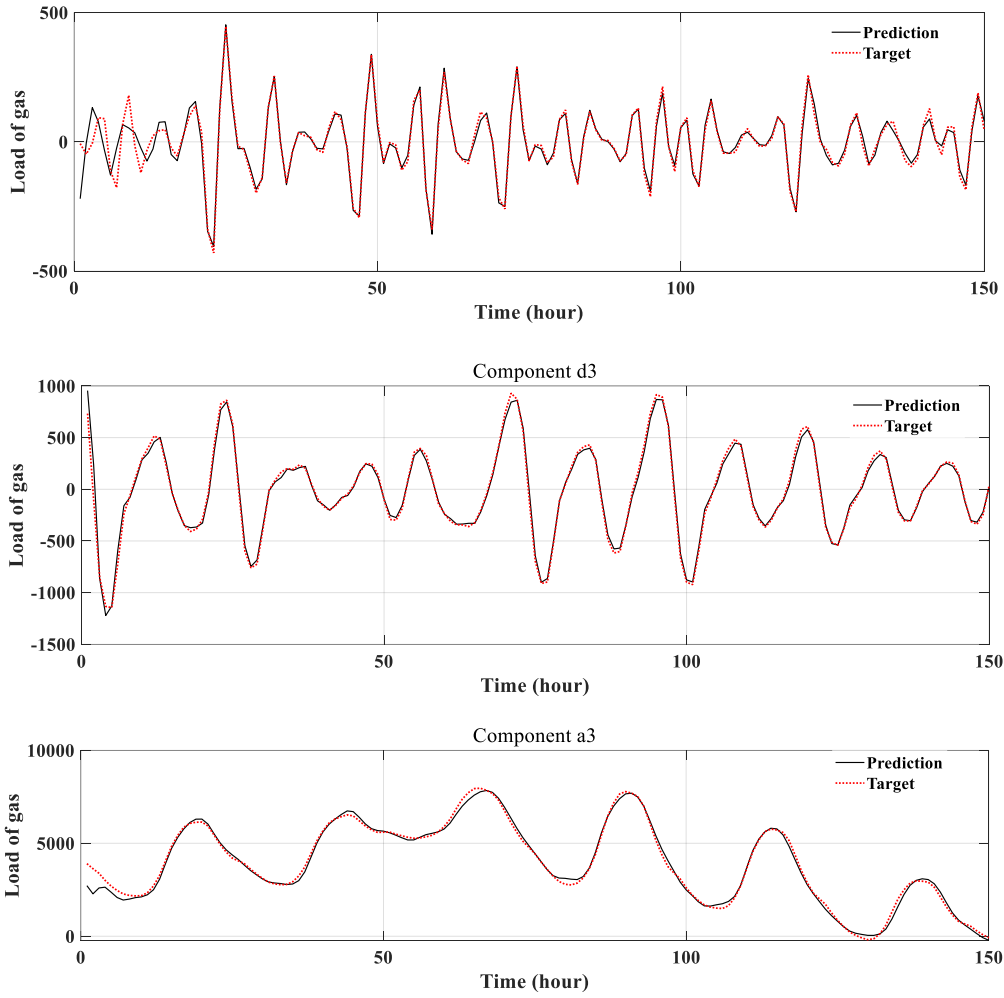
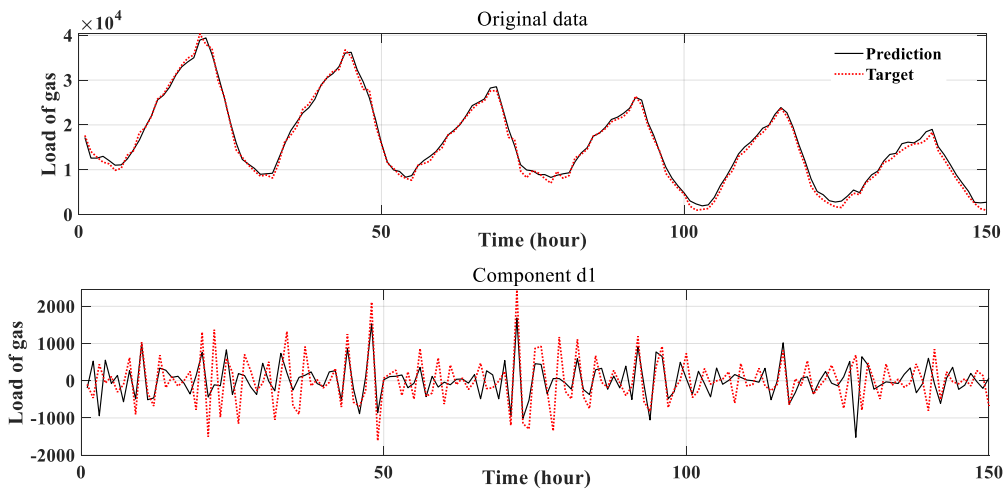


Fig. 9 Predicted demands compared with actual demands on data Set I



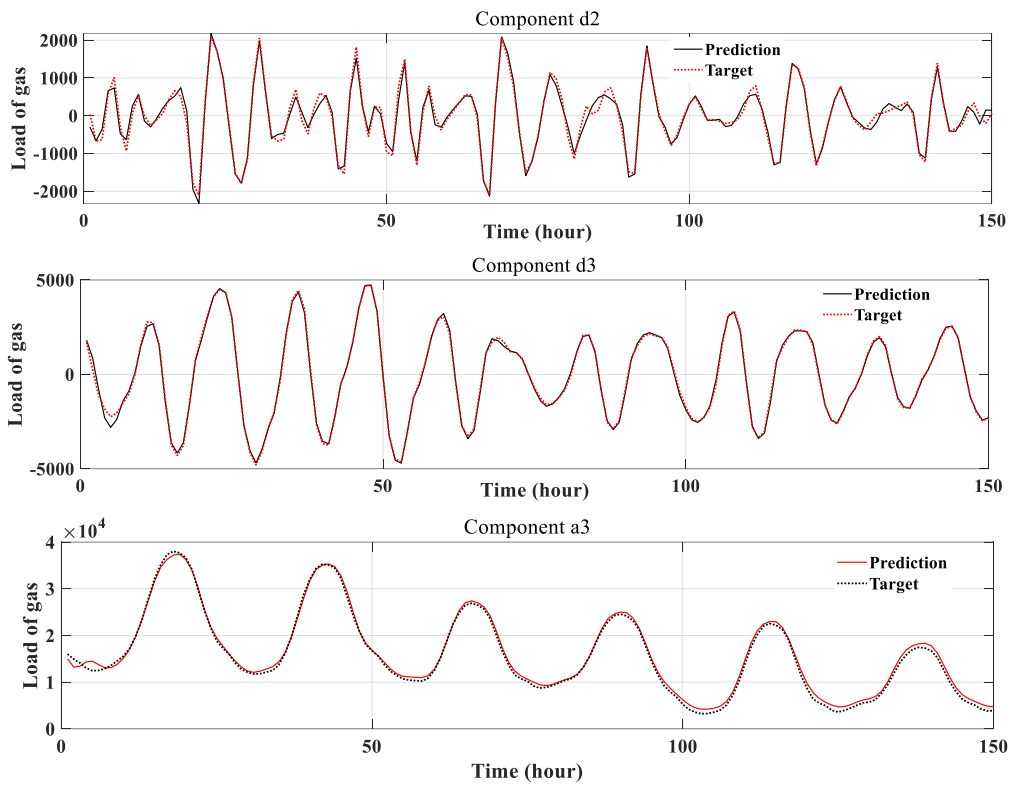


Fig. 10 Predicted demands compared with actual demands on data Set II

These Figures show that the developed forecasting model is able to make accurate predictions and this indicates that the deep RNN model has a strong ability to capture the features behind the gas consumption data. However, the model shows relatively poor performance on the wavelet components of d1, whose periodical behaviors are not obvious compared with the other components. From the opposite perspective, this observation indicates that the data decomposition process, via the wavelet method, can reduce the complexity of the forecasting task and improve the accuracy of the overall results.

The error measures are listed in Table 1. The errors are given based on different lengths of forecasting horizon, which are 1 hours, 5 hours and 10 hours, to present a comprehensive information of model performance and its sensitivity under different requirements of forecasting time horizon.

Table 1 Prediction performances for different forecasting horizons (winter data for data Set I and Set II)

Data basis	Task	MAE	MRE	RMSE
The Mackey Glass series	1 hour forecasting	0.0596	0.0017	0.1182
	5 hours forecasting	0.1960	0.0054	0.5537
	10 hours forecasting	0.5724	0.0157	1.0719
Set I	1 hour forecasting	19.0497	0.0058	25.9278
	5 hours forecasting	53.9272	0.0180	69.6089
	10 hours forecasting	125.7692	0.0584	167.1491
Set II	1 hour forecasting	81.4731	0.0061	109.5693
	5 hours forecasting	124.0326	0.0119	154.3480
	10 hours forecasting	594.2678	0.0678	744.4329

The error analysis results in Table 1 indicate that the developed model is capable to perform accurate forecasting on different data sets. The forecasting based on Mackey Glass series has an accuracy level of about 99%, even when the forecasting horizon is increased to 10 steps ahead. For the real world data in Set I and Set II, the MREs of the forecasting errors are maintained (at acceptable values of 5.84% and 6.78%), even when the model is used to perform 10-steps ahead predictions.

We notice that the forecasting performance of the developed model degenerates as the forecasting horizon extends. Generally, the reason of this degeneration is that the strength of the relationship between the future data of consumption and the current data decreases as the forecasting horizon increases, and this increases the difficulty for the Deep RNN model to learn such relationship. Hence, if we need to make a forecasting for a relatively long time in a real application, it is not a good choice to extend the forecasting horizon of the model without

limitation. To overcome this problem, one needs to firstly determine the limitation of the forecasting model based on the data and the accuracy requirement, and, then, apply controlled recursive processes that make use of the forecasting results to enhance the ability of the deep learning model [42].

To further verify the effectiveness of the forecasting model, a comparison of the forecasting performance is performed among the developed model, a three-layer-LSTM model and a Non-linear Autoregressive (NAR) model. The three-layer-LSTM model is introduced here to verify the effectiveness of the Wavelet Transformation and the Bi-LSTM model. The Non-linear Autoregressive model is a classical method for time series prediction and is used here to compare the overall performance of the developed model. The three-layer-LSTM shares the same structure with the optimized Deep RNN model. These three models are used to forecast the data of 10 hours ahead (10 steps) in the future, based on the three data sets. Then, the forecasting performances are presented by relative errors and compared with each other in the form of Cumulative Distribution Function (CDF). The analysis results are shown in Figs. 11-13.

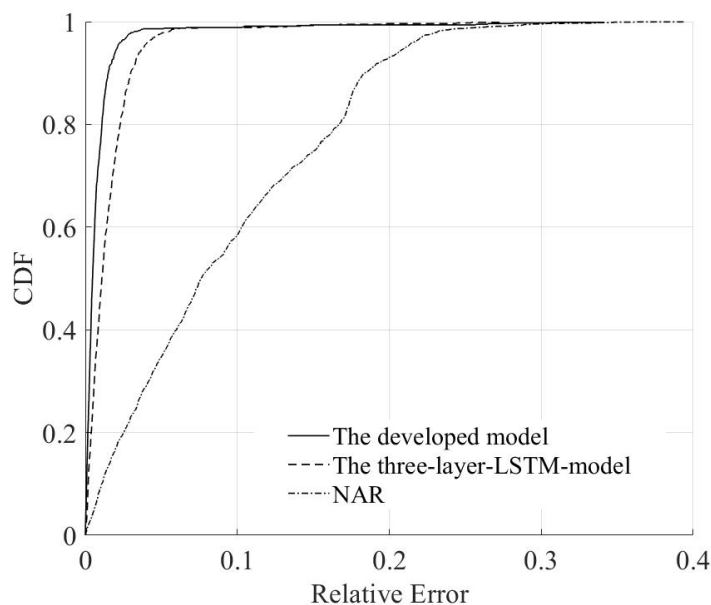


Fig. 11 Performance comparison based on the CDFs of the relative error values (the Mackey Glass

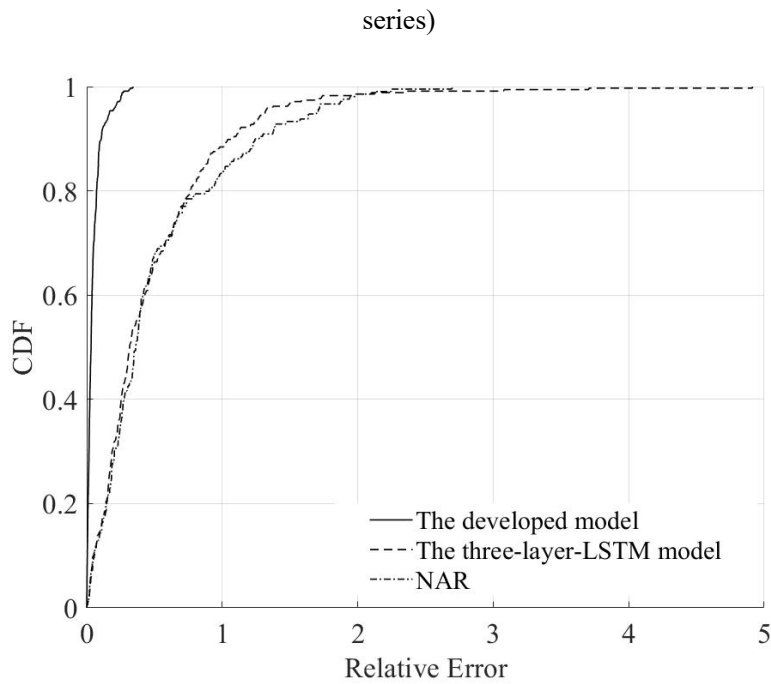


Fig. 12 Performance comparison based on the CDFs of the relative error values (the Set I)

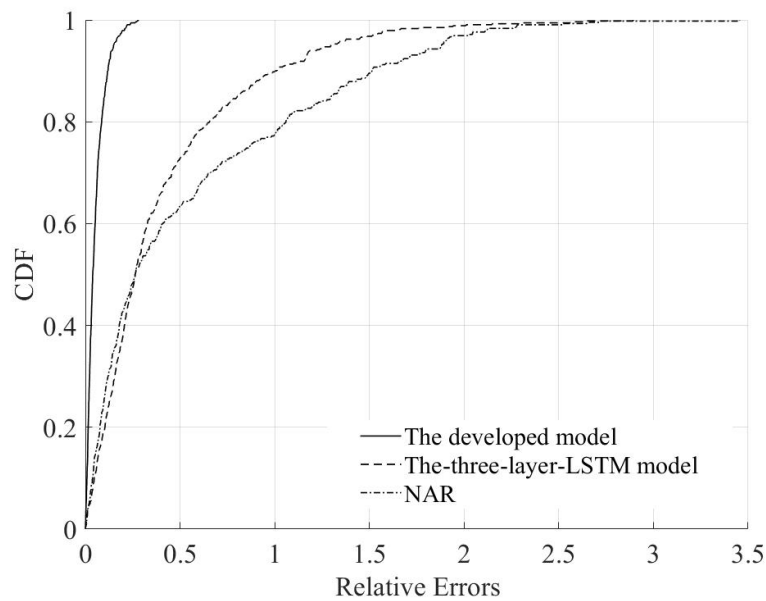


Fig. 13 Performance comparison based on the CDFs of the relative error values (the Set II)

From Figs. 11-13, we can conclude that the developed forecasting model outperforms the other models on different types of data sets. According to Figs. 12-13, we can also observe that the developed model has superior capacity compared to the three-layer-LSTM, in spite of the same deep structures of them. This is because the relationships behind the data of the real-world gas

consumption is complicated by the regularity, the customer habit and the market properties, and the deep Bi-LSTM is more powerful to capture this kind of relationship. Besides that, the Figures show that both the three-layer-LSTM and the developed model have better forecasting accuracies than the NAR model, which confirm the power of RNN models with deep structures, for natural gas demand forecasting.

The developed model has been observed to have a relatively good performance on the winter data. It remains to verify its performance for the other climate conditions. For this, we consider on the summer part of the data in Set II and on a 10-hours-forecasting. Figs. 14-16 present a relatively accurate forecasting for the gas demand also in the summer period: by comparing the forecasting performances of different forecasting horizons, we can observe very small differences between the real demands and the predictions.

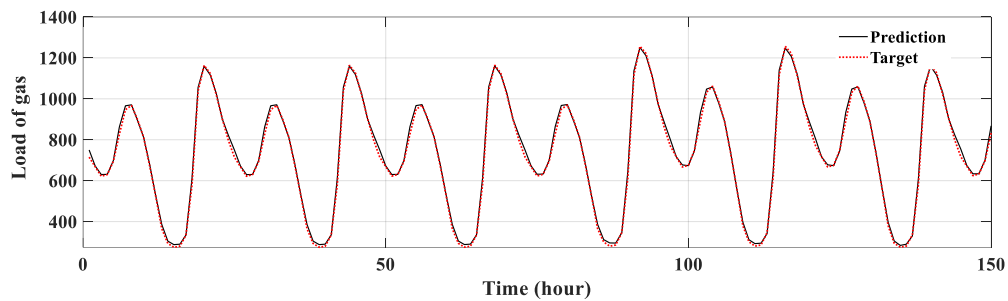


Fig. 14 Predicted demands compared with actual demands on the data summer (1 hour ahead)

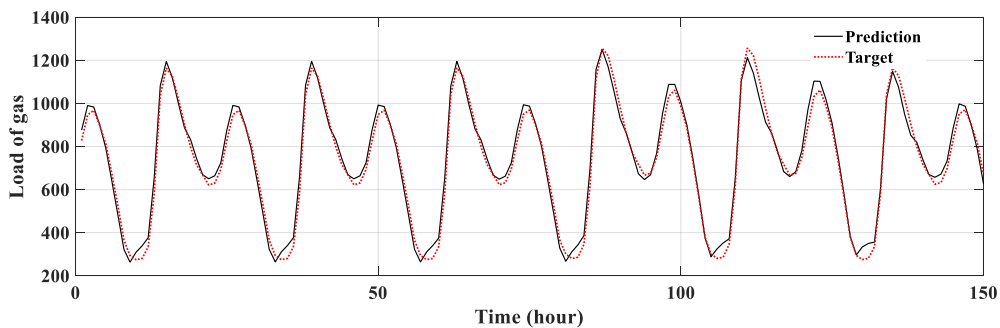


Fig. 15 Predicted demands compared with actual demands on the data summer (5 hours ahead)

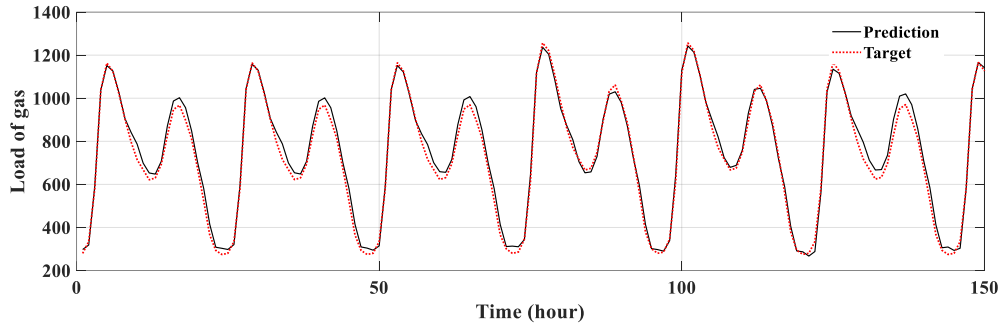


Fig. 16 Predicted demands compared with actual demands on the data summer (10 hours ahead)

The error measures are presented in Table 2.

Table 2 Prediction performances for different forecasting horizons (summer data for data Set II)

Data basis	Task	MAE	MRE	RMSE
Set II	1 hour forecasting	10.2701	0.0094	11.8352
	5 hours forecasting	11.9608	0.0110	14.1134
	10 hours forecasting	55.1135	0.0501	69.2558

4 Conclusions

The aim of this work is to introduce a new and highly reforming hourly forecasting method of natural gas demand forecasting. The method is developed based on the integrated Wavelet Transform, Bi-LSTM model, LSTM model and Genetic Algorithm. The Wavelet Transform is used to decompose the original series into sub-components, to reduce the difficulty of forecasting. Several Bi-LSTM models are stacked together to comprehensively learn the complicated relationships behind each sub-components and the LSTM model is adopted to forecast the future values of these components. To enhance the forecasting performances, the structures of the Bi-LSTM models and the LSTM models are optimized by the GA method. To avoid potential overfitting, dropout is performed during the training process.

The developed model is applied to three sets of data to verify its effectiveness: one simulated

set and two real gas consumption data. To test the robustness of the model, the forecasting is performed on different horizons, i.e., 1 hour, 5 hours and 10 hours. The experimental results show that the model is capable to achieve a high forecasting accuracy even when the horizon is increased to 10 steps ahead of the current data. The degeneration of performances at large forecasting horizons can be controlled by recursive processes method: this will be developed in the future research.

The forecasting performance of the developed model is compared with a three-layer-LSTM model and a Non-linear Autoregressive (NAR) model. The results of the comparison show that the developed model outperforms the three-layer-LSTM model and the NAR model, which indicates that the Bi-LSTM model has a solution capacity to learn the complicated features of the actual data.

In the future work, we will further improve the forecasting accuracy via exploring different advanced methods and performing more detailed analysis on different influencing variables. Furthermore, the ability of deep RNN models for natural gas demand forecasting on longer time horizons, e.g., days or months, will be explored.

Reference

- [1] M. Flouri, C. Karakosta, C. Kladouchou, and J. Psarras, "How does a natural gas supply interruption affect the EU gas security? A Monte Carlo simulation," *Renew. Sustain. Energy Rev.*, vol. 44, pp. 785–796, 2015.
- [2] B. Soldo, "Forecasting natural gas consumption," *Appl. Energy*, vol. 92, pp. 26–37, 2012.
- [3] J. G. Tamba *et al.*, "Forecasting natural gas: A literature survey," *Int. J. Energy Econ. Policy*, vol. 8, no. 3, pp. 216–249, 2018.
- [4] J. Szoplik, "Forecasting of natural gas consumption with artificial neural networks," *Energy*, vol. 85, pp. 208–220, 2015.
- [5] E. Erdogdu, "Natural gas demand in Turkey," *Appl. Energy*, vol. 87, no. 1, pp. 211–219, Jan. 2010.
- [6] F. Taşpinar, N. Çelebi, and N. Tutkun, "Forecasting of daily natural gas consumption on regional basis in Turkey using various computational methods," *Energy Build.*, vol. 56, pp. 23–31, 2013.
- [7] P. Potočnik, M. Thaler, E. Govekar, I. Grabec, and A. Poredoš, "Forecasting risks of natural gas consumption in Slovenia," *Energy Policy*, vol. 35, no. 8, pp. 4271–4282, Aug. 2007.
- [8] L. Zhu, M. S. Li, Q. H. Wu, and L. Jiang, "Short-term natural gas demand prediction based on support vector regression with false neighbours filtered," *Energy*, vol. 80, pp. 428–436, 2015.
- [9] J. H. Herbert, S. Sitzler, and Y. Eades-Pryor, "A statistical evaluation of aggregate monthly industrial demand for natural gas in the U.S.A.," *Energy*, vol. 12, no. 12, pp. 1233–1238, Dec. 1987.
- [10] Z. W. Geem and W. E. Roper, "Energy demand estimation of South Korea using artificial neural network," *Energy Policy*, vol. 37, no. 10, pp. 4049–4054, Oct. 2009.
- [11] M. A. Khan, "Modelling and forecasting the demand for natural gas in Pakistan," *Renew. Sustain. Energy Rev.*, vol. 49, pp. 1145–1159, Sep. 2015.
- [12] Q. Wang, S. Li, and R. Li, "Forecasting Energy Demand in China and India: Using Single-linear, Hybrid-linear, and Non-linear Time Series Forecast Techniques," *Energy*, Jul. 2018.
- [13] Y. Chen, W. S. Chua, and T. Koch, "Forecasting day-ahead high-resolution natural-gas demand and supply in Germany," *Appl. Energy*, vol. 228, pp. 1091–1110, Oct. 2018.
- [14] E. Čeperić, S. Žiković, and V. Čeperić, "Short-term forecasting of natural gas prices using machine learning and feature selection algorithms," *Energy*, vol. 140, pp. 893–900, Dec. 2017.
- [15] F. Yu and X. Xu, "A short-term load forecasting model of natural gas based on optimized genetic algorithm and improved BP neural network," *Appl. Energy*, vol. 134, pp. 102–113, 2014.
- [16] B. Soldo, "Forecasting natural gas consumption," *Appl. Energy*, vol. 92, pp. 26–37, Apr. 2012.
- [17] M. Aydinalp-Koksal and V. I. Ugursal, "Comparison of neural network, conditional demand analysis, and engineering approaches for modeling end-use energy consumption in the residential sector," *Appl. Energy*, vol. 85, no. 4, pp. 271–296, Apr. 2008.
- [18] M. Rigamonti, P. Baraldi, E. Zio, I. Roychoudhury, K. Goebel, and S. Poll, "Ensemble of optimized echo state networks for remaining useful life prediction," *Neurocomputing*, vol. 281, pp. 121–138, Mar. 2018.
- [19] J. Zheng, C. Xu, Z. Zhang, and X. Li, "Electric Load Forecasting in Smart Grid Using Long-Short-Term-Memory based Recurrent Neural Network Electric Load Forecasting in

- Smart Grid Using Long-Short-Term-Memory based Recurrent Neural Network,” *2017 51st Annu. Conf. Inf. Sci. Syst. (CISS)*, no. January, pp. 1–6, 2017.
- [20] A. Kamilaris and F. X. Prenafeta-Boldú, “Deep learning in agriculture: A survey,” *Comput. Electron. Agric.*, vol. 147, pp. 70–90, Apr. 2018.
- [21] D. Hirafuji Neiva and C. Zanchettin, “Gesture recognition: A review focusing on sign language in a mobile context,” *Expert Syst. Appl.*, vol. 103, pp. 159–183, Aug. 2018.
- [22] A. P. Marugán, F. P. G. Márquez, J. M. P. Perez, and D. Ruiz-Hernández, “A survey of artificial neural network in wind energy systems,” *Appl. Energy*, vol. 228, pp. 1822–1836, Oct. 2018.
- [23] A. M. Nagy and V. Simon, “Survey on traffic prediction in smart cities,” *Pervasive Mob. Comput.*, Jul. 2018.
- [24] T. Fischer and C. Krauss, “ARTICLE IN PRESS Deep learning with long short-term memory networks for financial market predictions,” *Eur. J. Oper. Res. Eur. J. Oper. Res. J.*, vol. 17, no. 0, pp. 48–0, 2017.
- [25] I. P. Panapakidis and A. S. Dagoumas, “Day-ahead natural gas demand forecasting based on the combination of wavelet transform and ANFIS/genetic algorithm/neural network model,” *Energy*, vol. 118, pp. 231–245, 2017.
- [26] A. Tascikaraoglu, B. M. Sanandaji, K. Poolla, and P. Varaiya, “Exploiting sparsity of interconnections in spatio-temporal wind speed forecasting using Wavelet Transform,” *Appl. Energy*, vol. 165, pp. 735–747, Mar. 2016.
- [27] P. M. R. Bento, J. A. N. Pombo, M. R. A. Calado, and S. J. P. S. Mariano, “A bat optimized neural network and wavelet transform approach for short-term price forecasting,” *Appl. Energy*, vol. 210, pp. 88–97, Jan. 2018.
- [28] M. Stocchi and M. Marchesi, “Fast wavelet transform assisted predictors of streaming time series,” *Digit. Signal Process.*, vol. 77, pp. 5–12, Jun. 2018.
- [29] H. Zheng, J. Yuan, and L. Chen, “Short-Term Load Forecasting Using EMD-LSTM Neural Networks with a Xgboost Algorithm for Feature Importance Evaluation,” *Energies*, vol. 10, no. 8, p. 1168, 2017.
- [30] H. I. Asala, J. Chebeir, W. Zhu, I. Gupta, A. D. Taleghani, and J. Romagnoli, “A Machine Learning Approach to Optimize Shale Gas Supply Chain Networks,” *SPE Annu. Tech. Conf. Exhib.*, no. October, pp. 0–28, 2017.
- [31] X. Qing and Y. Niu, “Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM,” *Energy*, vol. 148, pp. 461–468, Apr. 2018.
- [32] Z. Cui, R. Ke, and Y. Wang, “Deep Bidirectional and Unidirectional LSTM Recurrent Neural Network for Network-wide Traffic Speed Prediction,” pp. 22–25, 2018.
- [33] A. Mellit, S. A. Kalogirou, L. Hontoria, and S. Shaari, “Artificial intelligence techniques for sizing photovoltaic systems: A review,” *Renew. Sustain. Energy Rev.*, vol. 13, no. 2, pp. 406–419, Feb. 2009.
- [34] A. Rahman, V. Srikumar, and A. D. Smith, “Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks,” *Appl. Energy*, vol. 212, 2018.
- [35] Y. Gal and Z. Ghahramani, “A Theoretically Grounded Application of Dropout in Recurrent Neural Networks.”
- [36] J. Yang and J. Ma, “A structure optimization framework for feed-forward neural networks using sparse representation,” *Knowledge-Based Syst.*, vol. 109, pp. 61–70, Oct. 2016.

- [37] F. Yu and X. Xu, "A short-term load forecasting model of natural gas based on optimized genetic algorithm and improved BP neural network," *Appl. Energy*, vol. 134, pp. 102–113, Dec. 2014.
- [38] S. Askari, N. Montazerin, and M. H. F. Zarandi, "Forecasting semi-dynamic response of natural gas networks to nodal gas consumptions using genetic fuzzy systems," *Energy*, vol. 83, pp. 252–266, 2015.
- [39] S. Ene, İ. Küçüköğlü, A. Aksoy, and N. Öztürk, "A genetic algorithm for minimizing energy consumption in warehouses," *Energy*, vol. 114, pp. 973–980, 2016.
- [40] V. Sharma, D. Yang, W. Walsh, and T. Reindl, "Short term solar irradiance forecasting using a mixed wavelet neural network," *Renew. Energy*, vol. 90, pp. 481–492, 2016.
- [41] M. Bernas and B. Płaczek, "Period-aware local modelling and data selection for time series prediction," *Expert Syst. Appl.*, vol. 59, pp. 60–77, Oct. 2016.
- [42] H. Su, E. Zio, J. Zhang, Z. Yang, X. Li, and Z. Zhang, "A systematic hybrid method for real-time prediction of system conditions in natural gas pipeline networks," *J. Nat. Gas Sci. Eng.*, vol. 57, pp. 31–44, Sep. 2018.