

Article

Daily Natural Gas Load Forecasting Based on a Hybrid Deep Learning Model

Nan Wei ^{1,2,3}, Changjun Li ^{1,2,*}, Jiehao Duan ^{1,2}, Jinyuan Liu ^{1,2} and Fanhua Zeng ^{3,*}

¹ College of Petroleum Engineering, Southwest Petroleum University, Chengdu 610500, Sichuan, China; joey_weinan@126.com (N.W.); djh0114@163.com (J.D.); A7liujinyuan@163.com (J.L.)

² CNPC Key Laboratory of Oil & Gas Storage and Transportation, Southwest Petroleum University, Chengdu 610500, Sichuan, China

³ Faculty of Engineering and Applied Science, University of Regina, Regina, SK S4S 0A2, Canada

* Correspondence: lichangjunemail@sina.com (C.L.); fanhua.zeng@uregina.ca (F.Z.)

Received: 12 December 2018; Accepted: 7 January 2019; Published: 11 January 2019



Abstract: Forecasting daily natural gas load accurately is difficult because it is affected by various factors. A large number of redundant factors existing in the original dataset will increase computational complexity and decrease the accuracy of forecasting models. This study aims to provide accurate forecasting of natural gas load using a deep learning (DL)-based hybrid model, which combines principal component correlation analysis (PCCA) and (LSTM) network. PCCA is an improved principal component analysis (PCA) and is first proposed here in this paper. Considering the correlation between components in the eigenspace, PCCA can not only extract the components that affect natural gas load but also remove the redundant components. LSTM is a famous DL network, and it was used to predict daily natural gas load in our work. The proposed model was validated by using recent natural gas load data from Xi'an (China) and Athens (Greece). Additionally, 14 weather factors were introduced into the input dataset of the forecasting model. The results showed that PCCA–LSTM demonstrated better performance compared with LSTM, PCA–LSTM, back propagation neural network (BPNN), and support vector regression (SVR). The lowest mean absolute percentage errors of PCCA–LSTM were 3.22% and 7.29% for Xi'an and Athens, respectively. On these bases, the proposed model can be regarded as an accurate and robust model for daily natural gas load forecasting.

Keywords: artificial intelligence (AI); long short-term memory (LSTM); principal component analysis (PCA); natural gas load forecasting; deep learning (DL); recurrent neural networks (RNN)

1. Introduction

Natural gas load forecasting (NGLF) is an essential procedure for policy makers and related organizations within a natural gas system. Without accurate load forecasts, additional expenses due to uneconomic dispatch, over/under purchasing, and reliability uncertainty can cost a utility millions of dollars [1]. Therefore, many companies, organizations, and governments worldwide have already progressed the NGLF with versatile aspects [2]. The Energy Information Administration (EIA) forecasted the natural gas demand of the United States, finding that it could be 26.55 trillion cubic feet as of 2035 and that the annual energy demand would increase by 0.7% [3]. The National Natural Gas System Operator S.A. of Greece predicted the daily natural gas off-takes for the efficient operation of the natural gas supply system [4]. Thus, natural gas load is an important carrier in the energy mix and its accurate forecasting is of great significance to natural gas systems.

The accuracy of NGLF is mainly determined by forecasting models. Many methods have been developed for NGLF over the past several decades [5,6], including conventional statistical models [7–9]

and machine-learning (ML)-based models [10]. Since the nonlinear relationships among the parameters are complicated, conventional statistical models cannot obtain the desired prediction accuracy [11]. Lee and Tong [12] found that the forecasting accuracy of the autoregressive integrated moving average (ARIMA) model was poor for nonlinear data. In particular, when forecasting with the time series of China's energy consumption, the mean absolute percentage error (MAPE) of ARIMA is up to 25.89%. Conversely, ML-based models, such as the artificial neural network (ANN) [13–15], the back propagation neural network (BPNN) [16], and support vector regression (SVR) [17,18], have proved to be able to give better projected results compared with statistical models. Divina et al. [14] employed an ANN-based ensemble learning model to predict the short-term electricity consumption of Spain. Kaynar et al. [13] used ANN and ARIMA to predict the weekly natural gas consumption in Turkey. In the comparison of these two models, the MAPE of ANN was 5.47%, which decreased by 14.66% compared with ARIMA. Wei et al. [6] used an SVR-based hybrid model and over two years of daily consumption data to predict the natural gas consumption of Athens with the MAPE of 9.06%. Zhang et al. [19] proposed an SVR-based model to predict short-term electric load forecasting. In their work, the MAPE of the proposed model was reduced by 37.86% compared with the ARIMA-based model. Nowadays, deep learning (DL) models have achieved remarkable success in various artificial intelligence research areas [20]. As a specific subfield of ML, DL-based models are used in plenty of industries, from automated driving to medical devices [21], in which they have produced results comparable to and sometimes superior to human experts [22] and the conventional ML-based models [23,24]. Long short-term memory (LSTM) is one of the most famous DL-based models for solving forecasting problems. Gensler et al. [25] used an LSTM-based hybrid model to forecast the energy output of 21 solar power plants. In their work, the results presented that the LSTM-based model received a lower mean absolute error (MAE) and root-mean-square error (RMSE) compared with ANN as well as other reference models, such as physical models. Zaytar and Amrani [26] used multistacked LSTMs to forecast 24 and 72 h of weather data (temperature, humidity, and wind speed). Their study indicated that LSTM-based networks were becoming competitive compared with traditional methods and can be considered as a better alternative to forecast general weather conditions. He [24] used an LSTM-based model to forecast short-term electric load. By using three-year hourly load and temperature data from a northern city in China, the MAPEs of the proposed model were reduced by 52.2% and 18.3% compared with the linear regression approach and SVR, respectively.

The second aspect affecting the accuracy of NGLF is the redundant factors existed in original dataset. The natural gas load is affected by various factors, such as weather, economics, and politics [27]. The historical data of these factors consist of the input dataset of the forecasting models. After comprehensive data collection, the input datasets inevitably contain a large amount redundant information, which increases the computational complexity and decreases the accuracy of forecasting models [28]. Recently, many researchers have found that principal component analysis (PCA) can much improve forecasting accuracy by removing redundant components of the original dataset. Bianchi et al. [29] combined the echo state network (ESN) and PCA for predicting the short-term electric load of the power grid in Rome. In their work, the forecast error of PCA–ESN decreased by 7.34% compared with single ESN. Malvoni et al. [28] integrated PCA and the least-squares support vector machines (LSSVM) for forecasting the photovoltaic power in the day-ahead time. By introducing PCA, the training data size reduced to 30% of the original dimension, the normalized mean absolute error (NMAE) decreased by 4%, and the computational time reduced 70% compared with an implementation without PCA.

Based on the above brief literature survey, in the aspect of forecasting models, LSTM has already been employed to solve different kinds of forecasting tasks, and it has been proven to achieve better performance than ML-based models. Unlike ML-based models, LSTM performs the same task for every element in a sequence, with the output depending on the previous computations [30]. It seems that LSTM has a “memory” which captures information about what has been calculated before. This feature makes LSTM very suitable for predicting time series data. However, this new emerging model has

been rarely applied in daily NGLF. In the aspect of removing redundant components of raw datasets, PCA is a useful data preprocess technique, which has been widely used to preprocess the input datasets of various forecasting models. However, when observing the results of past works, which have focused on forecasting problems with conventional PCA, it can be conceivably stated that the conventional approach did not consider the correlation between the components and forecasting label. Although several methods have been proposed for selecting the components of PCA, including Kaiser's rule [31], cross validation [32], cumulative percent variance [33], and profile likelihood [34], using these methods to select the principal components from the original dataset will lose the information in minor components. These excluded components may have a higher correlation with the natural gas load compared with the principal components. Therefore, to better remove the redundant components of the original dataset to enhance the performance of the forecasting model, an improved PCA to extract components for the NGLF is necessary to be carried out.

The objective of this work is to develop a hybrid model for providing accurate forecasting of the natural gas load. The proposed model is a hybrid DL-based model combining principal component correlation analysis (PCCA) and the LSTM network, which is abbreviated as PCCA-LSTM. The preference of this paper for the NGLF over the others is justified by the following reasons:

- (a) This paper is mainly concerned with the recent data of Xi'an (China) and Athens (Greece). Contrary to the majority of the literature that focuses on the NGLF of cities in a certain country or area, this paper aims to develop a robust and feasible model to provide an accurate NGLF of cities worldwide. Using case studies, over two years of data on the natural gas load and 14 weather variations were used to validate the performance of the proposed model.
- (b) PCCA is first proposed in this paper. With PCCA, we developed a novel selection mechanism which considers correlation between each component in the eigenspace to select proper candidates for the input dataset of the forecasting model. This method can not only reduce the dimension of the input dataset but also automatically remove the redundant components.
- (c) A hybrid DL-based model is proposed for NGLF. This hybrid model combines PCCA and the LSTM network and is referred to as PCCA-LSTM. Practically, PCCA was used to remove redundant components and LSTM was applied to predict the natural gas load of cities in China and Greece. In addition, the comparative performance analysis of PCCA-LSTM with BPNN-, LSTM-, and SVR-based models is shown in this paper.

Our paper serves as an initial study focusing on the daily NGLF of cities worldwide. The objective of this work is to provide an accurate and robust forecasting tool to meet the actual demands of market participants, including natural gas companies, suppliers, and customers.

2. Natural Gas Load Forecasting Framework

2.1. Principal Component Correlation Analysis

Principal component analysis is a mathematical procedure that transforms some correlated variables into a number of uncorrelated variables called principal components. This classical technique has been widely applied to reduce the dimension of the original dataset, eliminate redundancy, and extract essential information from the factors [28]. Generally, selecting the principal components from the original dataset depending on the eigenvalue results in the loss of information in minor components, which may reduce the prediction accuracy of the model in certain situations. For instance, Figure 1 shows a comparison of the heatmap of an absolute correlation matrix before and after PCA. The dataset in Figure 1a uses the factors mentioned in [16]. The numbers 1–6 represent the previous day's consumption, the highest temperature, the lowest temperature, average temperatures, the weather conditions of the predicted day, and the data type, respectively. Figure 1b shows the heatmap of the data after PCA. It is clear that, after PCA, the correlation between the factors in the red box has decreased noticeably. This fact proves that PCA can extract the components from the original

dataset and reduce the correlation of the factors. However, the last factor in Figure 1b has a higher correlation with the natural gas load compared with the fourth factor, and the correlation coefficient of the penultimate factor is only 0.01 lower than the second factor. From this, we can conclude that the minor components obtained by PCA may have some useful information for NGLF. Using PCA to select the components based on the eigenvalue is inadequate for NGLF.

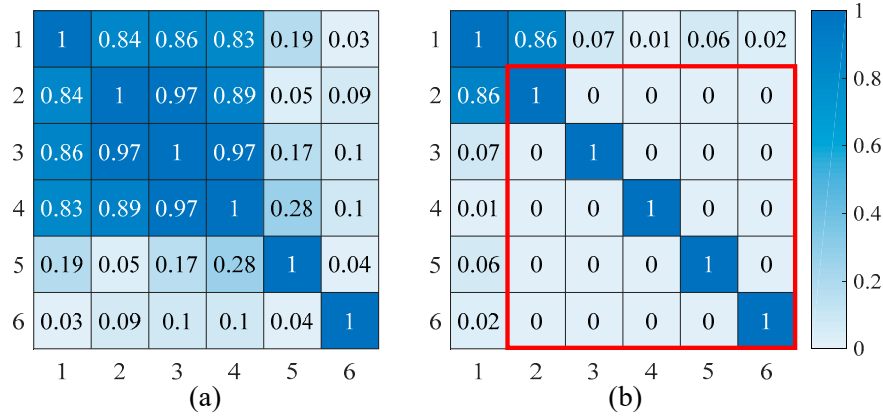


Figure 1. The heatmap of the correlation matrix compares the correlation between the factors before and after principal component analysis (PCA): (a) the correlation coefficient of the factors; (b) the correlation coefficient of the components extracted by PCA.

Principal component correlation analysis is a mathematical procedure that can be regarded as an effective data preprocess technique of NGLF. By combining the covariance method and correlation coefficient analysis, PCCA can automatically extract the components that influence the natural gas load and reduce the dimension of the raw dataset. Basically, PCCA can be divided into four steps.

Firstly, decentration and normalization: Given the original dataset X of dimensions $m \times n$, which consists of n correlated factors, and m is the data length of factors. Then, the matrix X will be decentralized and normalized as follows [35]:

$$\mu_j = \frac{1}{m} \sum_{i=1}^m x_j^i \quad (1)$$

$$x_j^i = x_j^i - \mu_j \quad (2)$$

$$\sigma_j^2 = \frac{1}{m} \sum_{i=1}^m (x_j^i)^2 \quad (3)$$

$$x_j^i = \frac{x_j^i}{\sigma_j} \quad (4)$$

where μ_j is the average of x_j^i , and σ_j is the variance of x_j^i .

Secondly, eigenvalue decomposition (ED): Denote C as the covariance matrix of X expressed as $C = \frac{1}{m} X^T X$. By introducing the ED technique, the covariance matrix can be split into eigenvalue matrix Λ and eigenvector matrix U .

Thirdly, correlation coefficient analysis: Make $L = XU$, which is consist of n -space vectors, represented as $L = \{l_1, l_2, \dots, l_n\}$. Then, analyze the correlation between label y and each column of L [36]:

$$\rho_j = \frac{|cov(l_j, y)|}{\sqrt{D(l_j)D(y)}} \quad (5)$$

where $cov(l_j, y)$ represents the covariance between component l_j and y ; y is the vector of label, which is the historical natural gas load in this paper; $D(l_j)$ is the variance of vector l_j ; and ρ_j represents the correlation coefficient, which ranges from 0 to 1, where 1 means total linear correlation and 0 means no linear correlation.

Finally, selection: Compute the contribution of each component in the matrix L :

$$b_j = \frac{\rho_j}{\sum_{j=1}^n \rho_j} \quad (6)$$

where b_j is the element of contribution vector B , which should be arranged in descending order. Let $r < n$, and compute the sum contributions of the first r value of vector B as $\sum_{j=1}^r b_j$. The sum of the contributions is at least bigger than a given threshold, for example, 90%. In L matrix, r columns are extracted, the subscripts of which are the same as those of first r items in the descending-ordered B vector. It can be observed that, with the different view of traditional PCA, PCCA selects the components based on the contribution of correlation coefficient rather than eigenvalue. This selection mechanism makes the PCCA more applicable for NGLF than conventional PCA.

2.2. Long Short-Term Memory

LSTM, as a useful architecture of DL-based models, was specifically designed to solve the forecasting problem [24]. By introducing a gate controller, LSTM is able to predict time series for long periods. Like other recurrent neural networks (RNNs), the LSTM network comprises many LSTM cells. The structure of an LSTM cell is shown in Figure 2.

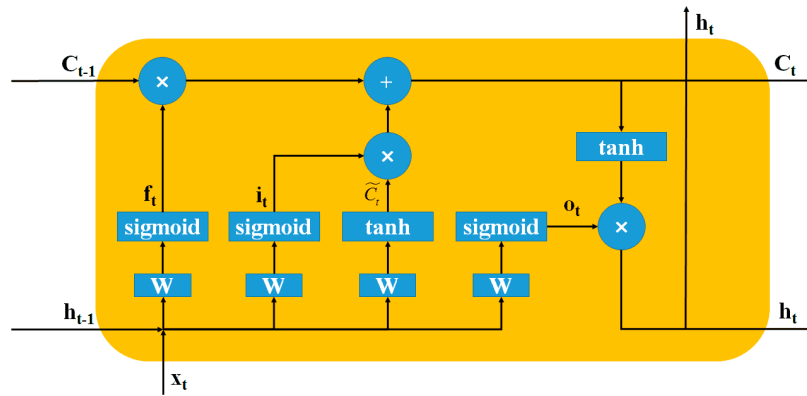


Figure 2. The structure of a single long short-term memory (LSTM) cell [37].

The LSTM cell can remember values over arbitrary time intervals and produce memories in LSTM. The memories are handled by three different gates: (a) input gate, (b) forget gate, and (c) output gate. The following equations express the computational process of a single LSTM [38]:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (7)$$

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (8)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (9)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (10)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (11)$$

$$h_t = o_t * \tanh(C_t) \quad (12)$$

where x_t is the input at time t ; h_t is the hidden state at time t ; i_t , f_t , and o_t represent the input gate, forget gate, and output gate, respectively; C_t is the cell state; \tilde{C}_t is the candidate cell state; and W and b are the parameters of the LSTM cell. Actually, by feeding the input time serial vector $X = \{x_1, x_2, \dots, x_n\}$ into the LSTM networks, the vector of hidden corresponding states $\{h_1, h_2, \dots, h_n\}$ as the outputs of the layer can be obtained. Furthermore, each element in the input vector X not only includes the historical natural gas load data but also contains the factors of natural gas load, such as temperature, wind speeds, humidity, etc.

In this study, we built a deep LSTM network for NGLF. The structure of this network is presented in Figure 3. The network consisted of training and forecasting parts, and the output layer in each part were RNN cells. Besides that, to make full use of the characteristics of the input data and achieve better performance, we introduced the LSTM cells in a hidden layer.

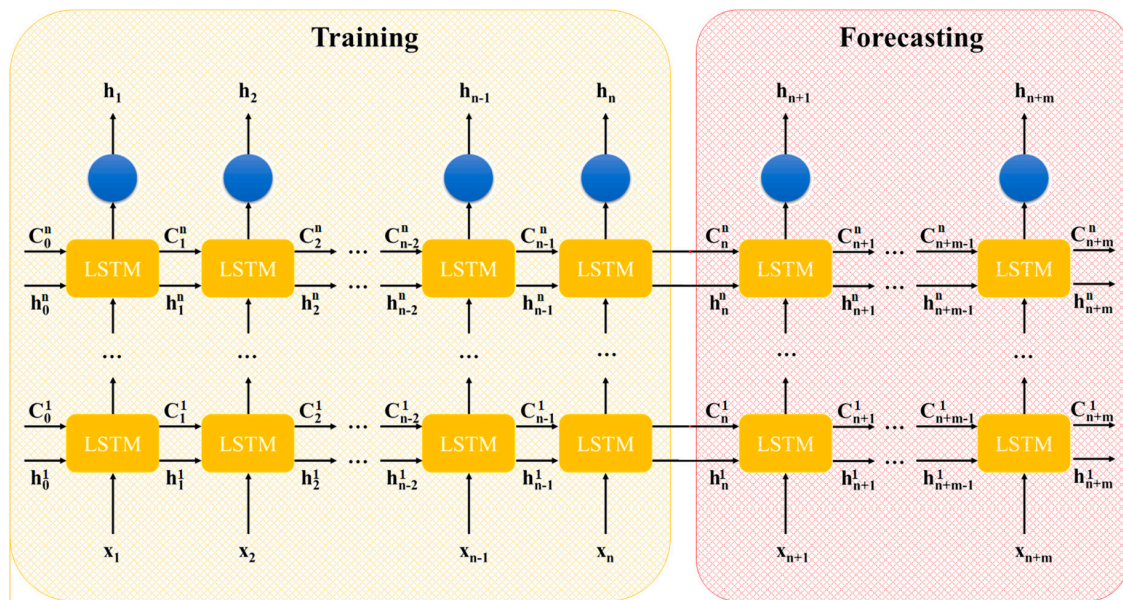


Figure 3. The LSTM network.

2.3. Construction of the Hybrid Model: PCCA–LSTM

In this paper, we propose a hybrid DL-based model, called PCCA–LSTM, for NGLF using the data of weather factors and the historical natural gas load as the foundation. First of all, comprehensive data collection was produced at this stage. Then, we employed PCCA to extract the components from the factors and to reduce the correlation between the factors. Finally, the input dataset of the forecasting model was formed by the outputs of PCCA and the historical natural gas load data, and the LSTM network was used to predict the daily natural gas load. The main procedure of the hybrid model is illustrated in Figure 4.

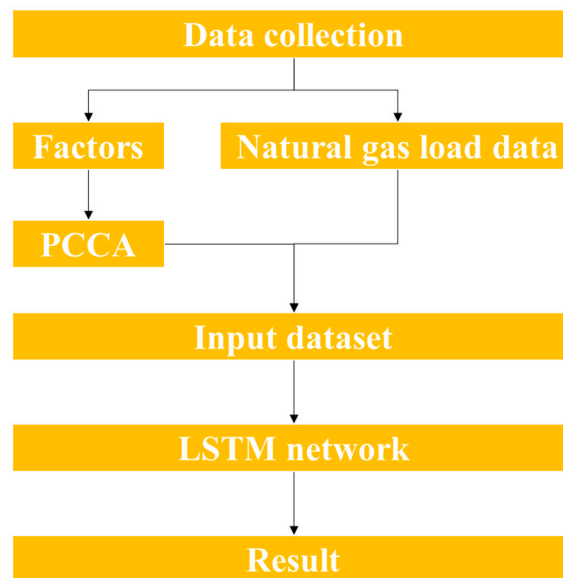


Figure 4. The flowchart of the proposed model, principal component correlation analysis (PCCA)–LSTM.

3. Case Studies

For this study, we designed two cases to test the performance of PCCA–LSTM. Among them, the first case was the Xi’an case, and the second one was the Athens case. In each case, the proposed model was compared with BPNN-, LSTM-, and SVR-based models optimized by the genetic algorithm (GA-SVR). The implementation of the LSTM network used in our work was based on the TensorFlow (TF) developed by Google Brain. The entire model was implemented by Pycharm 2017.2.4 (Python) (JetBrains Americas, Inc., Marlton, NJ, USA). Besides that, SVR and BPNN models were implemented by the relative toolbox in MATLAB R2017a (The MathWorks, Inc., Natick, NJ, USA). The optimal parameters of the models were determined via trial and error sets of simulation. The contribution thresholds of PCA and PCCA were both set as 99%.

3.1. Evaluation Framework

For all performed models, we quantified the prediction performance with the MAE, MAPE, RMSE, and mean absolute range normalized error (MARNE) [39]. These indicators can be defined as follows:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i^* - y_i| \quad (13)$$

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i^* - y_i}{y_i} \right| \times 100\% \quad (14)$$

$$\text{RMSE} = \frac{1}{N} \sqrt{\sum_{i=1}^N (y_i^* - y_i)^2} \quad (15)$$

$$\text{MARNE} = \frac{1}{N} \sum_{i=1}^N \frac{|y_i^* - y_i|}{\max(y_i^*)} \times 100\% \quad (16)$$

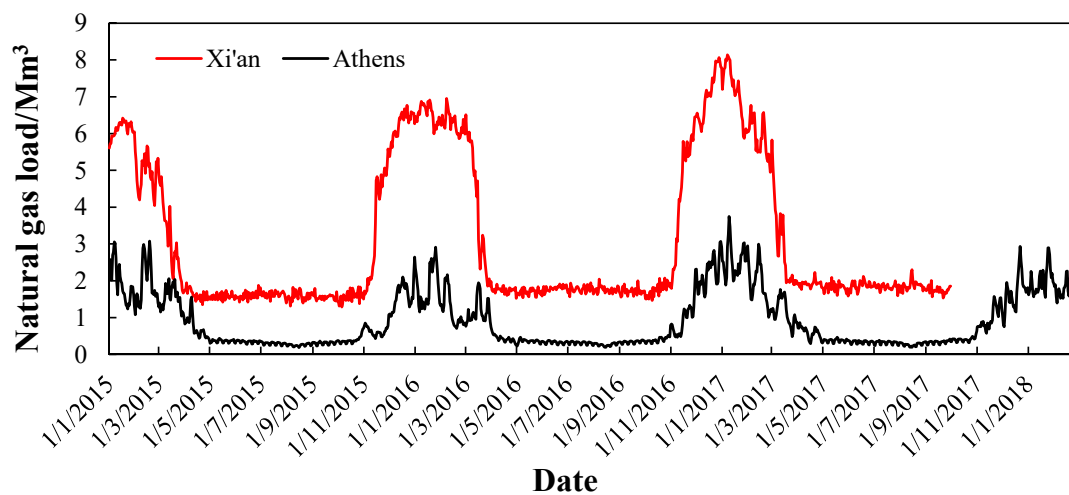
where N is the length of the vector of data, y_i is the actual data, and y_i^* is the forecasted value. Furthermore, Table 1 shows the benchmark of accuracy evaluation proposed by Lewis [40].

Table 1. Benchmark for modeling accuracy evaluation.

MAPE	$\leq 10\%$	10%–20%	20%–50%	$\geq 50\%$
Evaluation	Highly accurate	Good	Reasonable	Inaccurate

3.2. Data Description

In order to test the performance and robustness of the proposed model in different cities, we collected the recent natural gas load data of Xi'an (China) and Athens (Greece). Xi'an is the most populous city in Northwest China and has a population of 12 million as of 2018 [41]. Athens is the capital and largest city of Greece. As of 2018, the population of Athens is 0.7 million [42]. The natural gas loads of the cities are presented in Figure 5. The natural gas load data of Xi'an was obtained through the Xi'an Gas Company, while the natural gas load data of Athens was acquired from the official website of the National Natural Gas Transmission System (NNGTS) of Greece [3]. Figure 5 demonstrates that both cities consume more natural gas in winter months than in summer months. Furthermore, the natural gas load of Xi'an is much higher than that of Athens, owing to the huge gap in population between the two cities.

**Figure 5.** The natural gas loads of Xi'an and Athens.

According to the different natural gas loads of Xi'an and Athens, we designed two case studies for this study. The data used for the training and testing process in each case is presented in Table 2.

Table 2. The arrangement of data in the training and testing process.

City	Training		Testing	
	Data	Length	Data	Length
Xi'an	1 January 2015 to 31 August 2017	974	1–30 September 2017	30
Athens	1 January 2015 to 31 December 2017	1096	1 January 2018 to 28 February 2018	59

Actually, natural gas load is affected by various factors. In daily NGLF, weather factors are widely used because they are easily accessible and highly relevant [6,43]. In this paper, 14 weather factors were selected, including the lowest temperature (LT), average temperature (AT), highest temperature (HT), lowest dew point (LD), average dew point (AD), highest dew point (HD), lowest humidity (LH), average humidity (AH), highest humidity (HH), lowest visibility (LV), average visibility (AV), average air pressure (AA), average wind speed (AW), and precipitation (PP). The historical data of these factors were obtained by The Weather Company of IBM (www.wunderground.com).

3.3. Xi'an Case

First, we calculated the absolute correlation coefficient between the factors with Equation (5). Figure 6 shows that the biggest factors that affect the natural gas load are temperature factors (LT, AT, and HT), followed by dew point factors (LD, AD, and HD) and visibility factors (LV and AV). Among these factors, the correlation coefficient between AT and natural gas ranked first, with 0.86. By observing the correlation between these highly correlated factors, their correlation coefficients were higher than 0.69. According to the high correlation coefficient between these highly correlated factors, these factors had a similar effect on the natural gas load. Additionally, the low correlated factors to Xi'an's natural gas load were PP, AA, AW, and humidity factors.

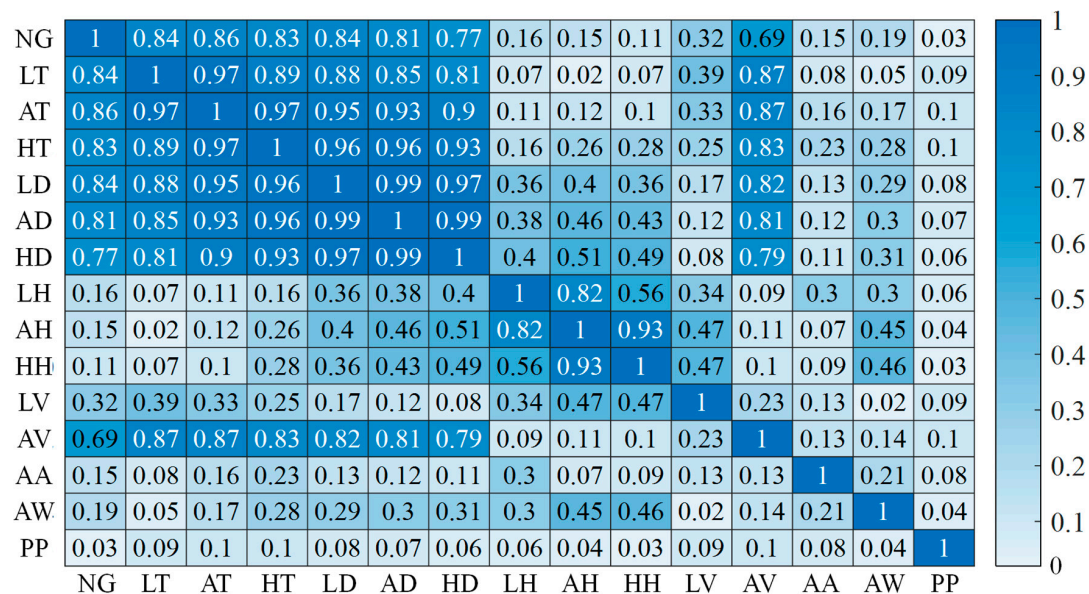


Figure 6. The absolute correlation coefficient between the factors of Xi'an case.

By using conventional PCA, nine principal components were extracted according to the eigenvalue of themselves (Figure 7). The correlation coefficient between the components was 0, which meant that each component was unrelated to others. The correlation coefficient between the first component and natural gas (NG) was 0.69, followed by component 2 (0.5), component 9 (0.27), and component 6 (0.16). The other principal components' correlation coefficient with NG was lower than 0.1.

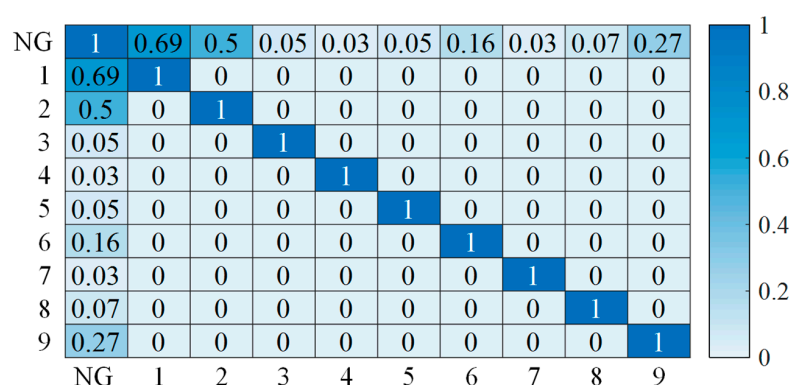


Figure 7. The principal components obtained by PCA (Xi'an case).

When applying PCCA, 13 unrelated components were extracted according to the correlation coefficient. Among these, five components (Nos. 1–5) that had over 0.1 correlation coefficient with NG were obtained. Figure 8 shows that component 5, the correlation coefficient of which was 0.11,

was omitted by PCA, as it was regarded as a minor component. However, its correlation coefficient with NG was much higher than many other components. The omitting of such components may affect the forecasting performance of models.

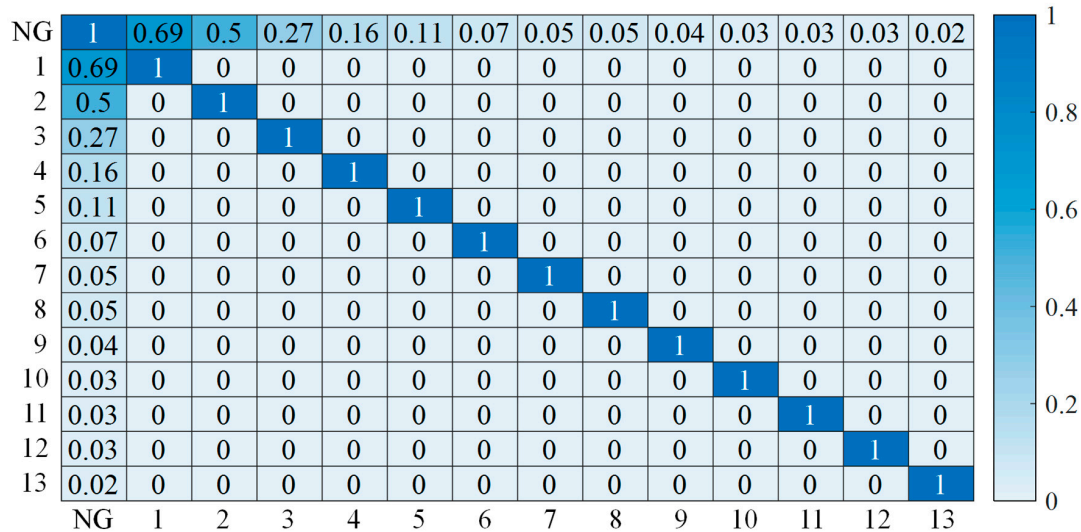


Figure 8. The components obtained by PCCA (Xi'an case).

After data preprocessing, we forecasted the NGLF of Xi'an using BPNN, GA-SVR, and LSTM models. Table 3 presents the forecasting results in this case. Regarding the forecasting time, GA-SVR took 241.23 s, which was much longer than BPNN and LSTM. Using PCA and PCCA to extract components and to reduce the dimension of the input dataset can effectively reduce the computation time. In this way, the times of PCA-LSTM and PCCA-LSTM were 110.07 and 162.31 s, respectively. From the perspective of forecasting errors, the highest forecasting error was for BPNN, the MAE, RMSE, MARNE, and MAPE of which were 0.17, 0.04, 9.32%, and 10.04%, respectively. According to Lewis's benchmark, BPNN achieved good performance in the NGLF of Xi'an. Except for BPNN, other models have made highly accurate forecasting in this case. When forecasting with the raw dataset, LSTM showed better forecasting accuracy compared with GA-SVR, and the MAPE of LSTM was 6.56%. This fact shows that the LSTM model outperformed BPNN and SVR in NGLF. When PCA was introduced, the forecasting errors of PCA-LSTM were higher than that of LSTM, and its MAE, RMSE, MARNE, and MAPE were 0.11, 0.03, 6.03%, and 6.99%, respectively. The best forecasting performance was obtained by PCCA-LSTM. The MAPE of the proposed model was 3.22%, which was a MAPE reduction of 51.91% compared to LSTM. Given this, it can be concluded that using PCA to refine the factors cannot improve forecasting accuracy, which conflicts with the results of [28,29]. Conversely, PCCA can effectively enhance the forecasting accuracy of LSTM. This proved that using PCA to extract the components from the raw dataset will omit some minor components that affect the natural gas load and ultimately worsen forecasting accuracy.

Table 3. The forecasting results in Xi'an case.

Items	BPNN	GA-SVR	LSTM	PCA-LSTM	PCCA-LSTM
Time (s)	155.77	241.23	186.34	110.07	162.31
MAE	0.17	0.15	0.10	0.11	0.06
RMSE	0.04	0.03	0.03	0.03	0.01
MARNE	9.32%	7.98%	5.87%	6.03%	3.02%
MAPE	10.04%	8.23%	6.56%	6.99%	3.22%
Evaluation	Good	Highly accurate	Highly accurate	Highly accurate	Highly accurate

3.4. Athens Case

In the Athens case, the most effective factors for natural gas load were temperature factors (LT, AT, and HT), which had a higher than 0.8 correlation coefficient to natural gas load (Figure 9). Dew point (LD, AD, and HD) factors were highly related to temperature factors and their correlation coefficients to natural gas load were in the range of 0.70–0.78. Additionally, the correlation coefficients of humidity factors (LH, AH, and HH) and AA to NG were in the range of 0.36–0.42. Unlike the Xi'an case, AV had low correlation with NG and its correlation coefficient was only 0.09. This result shows that the factors that affect natural gas load are quite different for different cities, which is consistent with the general conclusion made by Wei et al. [6].

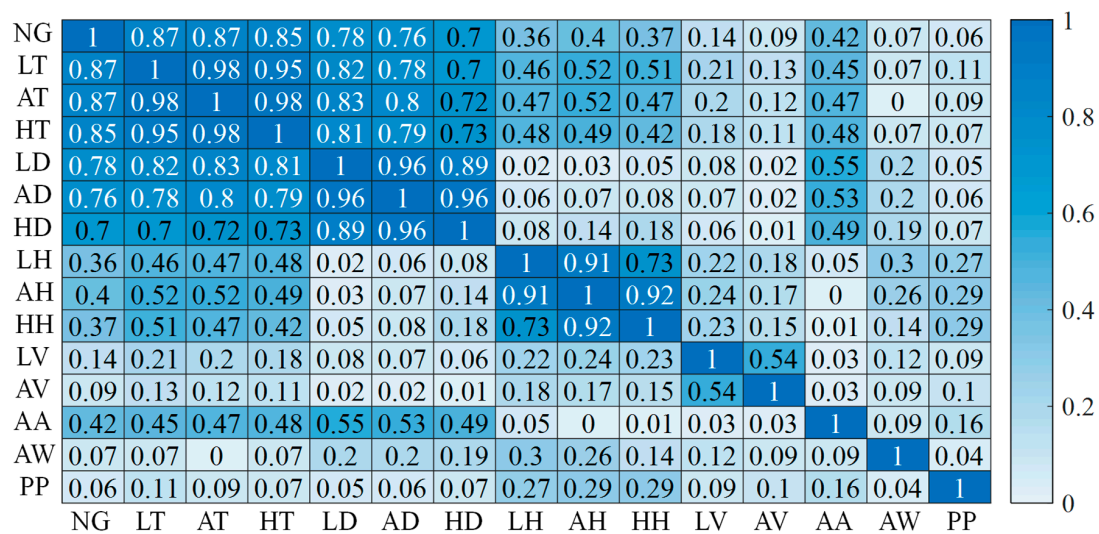


Figure 9. The absolute correlation coefficient between the factors of the Athens case.

When PCA was introduced, in this case, six principal components were extracted from the original dataset. Figure 10 shows that, among these factors, component 1 had the highest correlation with natural gas load, reaching a 0.88 correlation coefficient. The first component obtained by PCA often had the highest correlation coefficient with the forecasting object, while the correlation coefficients between other factors and NG were in the range of 0.01–0.07. The lowest correlation coefficient to NG was for component 6, which was only 0.01.

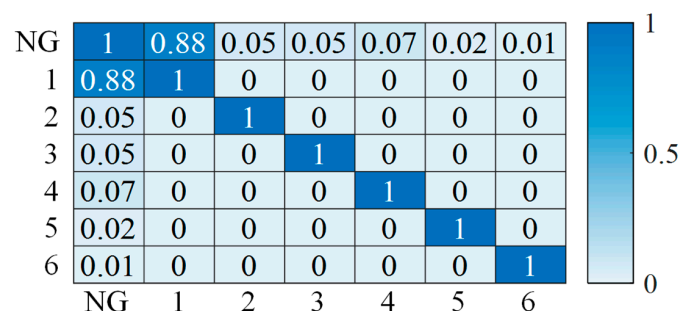


Figure 10. The principal components obtained by PCA (Athens case).

Figure 11 presents 10 components obtained by PCCA in the Athens case. The most effective component ranked in No. 1 had a 0.88 correlation coefficient to NG. The correlation coefficient of other factors to NG was in the range of 0.02–0.07. More remarkably, the component that had a 0.01 correlation coefficient to NG was not extracted by PCCA. This suggests that some principal components with a low correlation coefficient to the forecasting object might be extracted by PCA, but those can be excluded by PCCA.

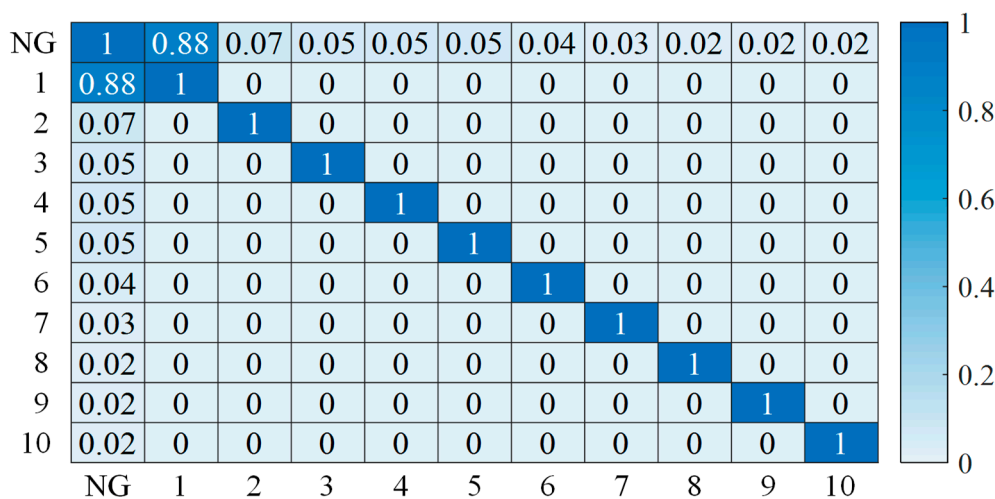


Figure 11. The components obtained by PCCA (Athens case).

By observing the prediction results in Table 4, when forecasting with the original dataset, LSTM achieved highly accurate forecasting, while the performances of BPNN and GA-SVR were good. The MAPEs of BPNN, GA-SVR, and LSTM were 16.53%, 17.14%, and 7.88%, respectively. This fact also shows the superiority of LSTM in daily natural gas forecasting. Similar to what happened in the Xi'an case, by introducing PCA, the computation time decreased by 48.27 s compared to that of LSTM, but the MAPE of PCA-LSTM was higher than that of LSTM, reaching 13.40%. The lowest forecasting errors were obtained by PCCA-LSTM. The MAE, RMSE, MARNE, and MAPE of the proposed model were 0.13, 0.04, 6.66%, and 7.29%, respectively, which was better than the forecasting performance obtained in [6]. Compared with LSTM, the MAPE of PCCA-LSTM reduced by 7.49%. Furthermore, the computation time of PCCA-LSTM was 139.31 s, which was lower than that of LSTM. It can be concluded that using PCCA to extract the components from the original dataset can reduce the computation time and improve the forecasting accuracy of LSTM. Compared with conventional PCA, PCCA is more robust and effective in improving the forecasting performance of forecasting models.

Table 4. The forecasting results in Athens case.

Items	BPNN	GA-SVR	LSTM	PCA-LSTM	PCCA-LSTM
Time(s)	176.39	303.01	154.34	106.07	139.31
MAE	0.29	0.3	0.14	0.26	0.13
RMSE	0.1	0.09	0.04	0.07	0.04
MARNE	14.02%	14.19%	5.56%	11.27%	6.66%
MAPE	16.53%	17.14%	7.88%	13.40%	7.29%
Evaluation	Good	Good	Highly accurate	Good	Highly accurate

4. Conclusions

In this work, the PCCA-LSTM model was proposed to provide accurate forecasting of the daily natural gas load. The proposed model combined PCCA and the LSTM network, simultaneously. PCCA can automatically extract the components that affect natural gas load and reduce the redundant factors existing in the original dataset. LSTM, as a useful DL model, was used to train the hybrid model. The robustness and effectiveness of the proposed model were verified by recent natural gas load data from Xi'an (China) and Athens (Greece).

In the Xi'an case, the best forecasting performance was obtained by PCCA-LSTM. The MAPE of the proposed model was 3.22%, a reduction of 51.91% compared with LSTM. By introducing PCCA, the computation time of LSTM decreased by 24.03 s. In the Athens case, compared with LSTM, the MAPE of PCCA-LSTM reduced by 7.49%, reaching 7.29%, and the computation time of PCCA-LSTM was 139.31 s, which is lower than that of LSTM (154.34 s). In both of those two cases,

the LSTM model outperformed BPNN and GA-SVR in terms of forecasting errors, which shows the superiority of LSTM in daily natural gas forecasting. Additionally, forecasting results showed that when PCA was introduced, the forecasting errors of PCA–LSTM were higher than that of LSTM. It can be concluded that using PCA to extract the components from the raw dataset will omit some minor components that affect natural gas load and ultimately worsen forecasting accuracy. Compared with conventional PCA, PCCA is more robust and effective in improving the forecasting performance of forecasting models.

The proposed intelligence hybrid model, PCCA–LSTM, has high flexibility, low requirements, and high accuracy for NGLF. It can be used by operators, market participants, decision makers, etc. Since NGLF is not an easy task, more efforts should be made in this field. Future challenges in NGLF could be related to the following: (a) combining the forecasting model with artificial intelligence technology and (b) improving the quality of the input dataset and focusing on data collection and preprocessing.

Author Contributions: Data curation, F.Z.; Formal analysis, N.W.; Funding acquisition, C.L.; Investigation, J.D. and J.L.; Writing—original draft, N.W.

Funding: This study was financially supported by the National Natural Science Foundation of China (Nos. 51674213, 51474184, 51604233, and 51504206), a subproject of the National Science and Technology Major Project of China (No. 2016ZX05028-001-006) and the China Scholarship Council (CSC) No.201808510209.

Conflicts of Interest: The authors declare that they have no competing financial interests.

References

1. Bunn, D.W. Short-term forecasting: A review of procedures in the electricity supply industry. *J. Oper. Res. Soc.* **1982**, *33*, 533–545. [CrossRef]
2. Tamba, J.G.; Essiane, S.N.; Sapnken, E.F.; Koffi, F.D.; Nsouandélé, J.L.; Soldo, B.; Njomo, D. Forecasting Natural Gas: A Literature Survey. *Int. J. Energy Econ. Policy* **2018**, *8*, 216–249.
3. Natgas Natural Gas Demand. Available online: <http://naturalgas.org/business/demand/#factorshort> (accessed on 20 September 2013).
4. DESFA Historical Data of Natural Gas Deliveries. Available online: <http://www.desfa.gr/en/regulatory-services> (accessed on 1 March 2018).
5. Soldo, B. Forecasting natural gas consumption. *Appl. Energy* **2012**, *92*, 26–37. [CrossRef]
6. Wei, N.; Li, C.; Li, C.; Xie, H.; Du, Z.; Zhang, Q.; Zeng, F. Short-term forecasting of natural gas consumption using factor selection algorithm and optimized support vector regression. *J. Energy Resour. Technol.* **2018**, *141*, 032701. [CrossRef]
7. Hubbert, M. Nuclear Energy and the Fossil Fuel. In *Drilling and Production Practice*; American Petroleum Institute: Washington, DC, USA, 1956.
8. Wang, J.; Jiang, H.; Zhou, Q.; Wu, J.; Qin, S. China's natural gas production and consumption analysis based on the multicycle Hubbert model and rolling Grey model. *Renew. Sustain. Energy Rev.* **2016**, *53*, 1149–1167. [CrossRef]
9. Ediger, V.S.; Akar, S. ARIMA forecasting of primary energy demand by fuel in Turkey. *Energy Policy* **2007**, *35*, 1701–1708. [CrossRef]
10. Brown, R.H.; Matin, I. Development of artificial neural network models to predict daily gas consumption. In Proceedings of the IEEE IECON 21st International Conference on Industrial Electronics, Control, and Instrumentation, Orlando, FL, USA, 6–10 November 1995; Volume 2, pp. 1389–1394.
11. Bai, Y.; Li, C. Daily natural gas consumption forecasting based on a structure-calibrated support vector regression approach. *Energy Build.* **2016**, *127*, 571–579. [CrossRef]
12. Lee, Y.S.; Tong, L.I. Forecasting time series using a methodology based on autoregressive integrated moving average and genetic programming. *Knowl. Based Syst.* **2011**, *24*, 66–72. [CrossRef]
13. Kaynar, O.; Yilmaz, I.; Demirkoparan, F. Forecasting of natural gas consumption with neural network and neuro fuzzy system. In Proceedings of the EGU General Assembly Conference, Vienna, Austria, 2–7 May 2010; pp. 221–238.

14. Divina, F.; Gilson, A.; Gómez-Vela, F.; García Torres, M.; Torres, J.F. Stacking Ensemble Learning for Short-Term Electricity Consumption Forecasting. *Energies* **2018**, *11*, 949. [\[CrossRef\]](#)
15. Zhou, X.; Yuan, Q.; Rui, Z.; Wang, H.; Feng, J.; Zhang, L.; Zeng, F. Feasibility study of CO₂ huff'n'puff process to enhance heavy oil recovery via long core experiments. *Appl. Energy* **2019**, *236*, 526–539. [\[CrossRef\]](#)
16. Yu, F.; Xu, X. A short-term load forecasting model of natural gas based on optimized genetic algorithm and improved BP neural network. *Appl. Energy* **2014**, *134*, 102–113. [\[CrossRef\]](#)
17. Zhu, L.; Li, M.; Wu, Q.; Jiang, L. Short-term natural gas demand prediction based on support vector regression with false neighbours filtered. *Energy* **2015**, *80*, 428–436. [\[CrossRef\]](#)
18. He, Y.; Liu, R.; Li, H.; Wang, S.; Lu, X. Short-term power load probability density forecasting method using kernel-based support vector quantile regression and Copula theory. *Appl. Energy* **2017**, *185*, 254–266. [\[CrossRef\]](#)
19. Zhang, X.; Wang, J.; Zhang, K. Short-term electric load forecasting based on singular spectrum analysis and support vector machine optimized by Cuckoo search algorithm. *Electr. Power Syst. Res.* **2017**, *146*, 270–285. [\[CrossRef\]](#)
20. Chen, H.; Engkvist, O.; Wang, Y.; Olivecrona, M.; Blaschke, T. The rise of deep learning in drug discovery. *Drug Discov. Today* **2018**, *23*, 1241–1250. [\[CrossRef\]](#) [\[PubMed\]](#)
21. Ghasemi, F.; Mehridehnavi, A.; Fassihi, A.; Pérez-Sánchez, H. Deep neural network in QSAR studies using deep belief network. *Appl. Soft Comput.* **2018**, *62*, 251–258. [\[CrossRef\]](#)
22. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems—Volume 1, Lake Tahoe, NV, USA; 2012; pp. 1097–1105.
23. Graves, A. *Supervised Sequence Labelling with Recurrent Neural Networks*; Springer: Berlin/Heidelberg, Germany, 2012.
24. He, W. Load Forecasting via Deep Neural Networks. *Procedia Comput. Sci.* **2017**, *122*, 308–314. [\[CrossRef\]](#)
25. Gensler, A.; Henze, J.; Sick, B.; Raabe, N. Deep Learning for solar power forecasting—An approach using AutoEncoder and LSTM Neural Networks. In Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, Banff, Canada, 5–8 October 2017; pp. 002858–002865.
26. Zaytar, M.A.; El Amrani, C. Sequence to sequence weather forecasting with long short term memory recurrent neural networks. *Int. J. Comput. Appl.* **2016**, *143*, 7–11.
27. Zhou, X.; Yuan, Q.; Zhang, Y.; Wang, H.; Zeng, F.; Zhang, L. Performance evaluation of CO₂ flooding process in tight oil reservoir via experimental and numerical simulation studies. *Fuel* **2019**, *236*, 730–746. [\[CrossRef\]](#)
28. Malvoni, M.; Giorgi, M.G.D.; Congedo, P.M. Photovoltaic forecast based on hybrid PCA–LSSVM using dimensionality reduced data. *Neurocomputing* **2016**, *211*, 72–83. [\[CrossRef\]](#)
29. Bianchi, F.M.; Santis, E.D.; Rizzi, A.; Sadeghian, A. Short-Term Electric Load Forecasting Using Echo State Networks and PCA Decomposition. *IEEE Access* **2015**, *3*, 1931–1943. [\[CrossRef\]](#)
30. Lecun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Jolliffe, I.T. Principal component analysis. *J. Mark. Res.* **2002**, *87*, 513.
32. Li, B.; Morris, J.; Martin, E.B. Model selection for partial least squares regression. *Chemom. Intell. Lab. Syst.* **2002**, *64*, 79–89. [\[CrossRef\]](#)
33. Harrou, F.; Kadri, F.; Chaabane, S.; Tahon, C.; Sun, Y. Improved principal component analysis for anomaly detection: Application to an emergency department. *Comput. Ind. Eng.* **2015**, *88*, 63–77. [\[CrossRef\]](#)
34. Zhu, M.; Ghodsi, A. Automatic dimensionality selection from the scree plot via the use of profile likelihood. *Comput. Stat. Data Anal.* **2006**, *51*, 918–930. [\[CrossRef\]](#)
35. White, D.A.; Goodlin, B.E.; Gower, A.E.; Boning, D.S.; Chen, H.; Sawin, H.H.; Dalton, T.J. Low open-area endpoint detection using a PCA-based T/sup 2/statistic and Q statistic on optical emission spectroscopy measurements. *IEEE Trans. Semicond. Manuf.* **2002**, *13*, 193–207. [\[CrossRef\]](#)
36. Taylor, R. Interpretation of the correlation coefficient: A basic review. *J. Diagn. Med. Sonogr.* **1990**, *6*, 35–39. [\[CrossRef\]](#)
37. Zhao, Z.; Chen, W.; Wu, X.; Chen, P.C.Y.; Liu, J. LSTM network: A deep learning approach for short-term traffic forecast. *IET Intell. Transp. Syst.* **2017**, *11*, 68–75. [\[CrossRef\]](#)
38. Zheng, H.; Yuan, J.; Chen, L. Short-Term Load Forecasting Using EMD-LSTM Neural Networks with a Xgboost Algorithm for Feature Importance Evaluation. *Energies* **2017**, *10*, 1168. [\[CrossRef\]](#)

39. Panapakidis, I.P.; Dagoumas, A.S. Day-ahead natural gas demand forecasting based on the combination of wavelet transform and ANFIS/genetic algorithm/neural network model. *Energy* **2017**, *118*, 231–245. [CrossRef]
40. Lewis, C. *Industrial and Business Forecasting Methods: A Practical Guide to Exponential Smoothing and Curve Fitting*; Butterworth Scientific: London, UK, 1982.
41. OECD. *OECD Urban Policy Reviews: China 2015*; OECD: Paris, France, 2015. [CrossRef]
42. Nations, U. Greece Population. Available online: <http://worldpopulationreview.com/countries/greece-population> (accessed on 24 September 2018).
43. Wright, C.; Chan, C.W.; Laforge, P. Towards Developing a Decision Support System for Electricity Load Forecast. Available online: https://www.intechopen.com/books/decision-support-systems_2012/towards-developing-a-decision-support-system-for-electricity-load-forecast (accessed on 17 October 2012).



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).