

# A two-stage supervised learning approach for electricity price forecasting by leveraging different data sources

Shuman Luo, Yang Weng\*

School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ 85287, USA

## HIGHLIGHTS

- A two-stage method is proposed for electricity price forecasting.
- The method diversifies the data resources improve the forecasting accuracy.
- The method is tested using the real-world data of Texas electricity market.
- Causal relationship combined with physical models guarantee better results.
- Deep neural networks show a great performance in electricity price forecasting.

## ARTICLE INFO

### Keywords:

Electricity price forecast  
Machine learning methods  
Renewables  
Wind power generation

## ABSTRACT

Over the years, the growing penetration of renewable energy into the electricity market has resulted in a significant change in the electricity market price. This change makes the existing forecasting method prone to error, decreasing the economic benefits. Hence, more precise forecasting methods need to be developed. This paper starts with a survey and benchmark of existing machine learning approaches for forecasting the real-time market (RTM) price. While these methods provide sufficient modeling capability via supervised learning, their accuracy is still limited due to the single data source, e.g., historical price information only. In this paper, a novel two-stage supervised learning approach is proposed by diversifying the data sources such as highly correlated power data. This idea is inspired by the recent load forecasting methods that have shown extremely well performances. Specifically, the proposed two-stage method, namely the rerouted method, learns two types of mapping rules. The first one is the mapping between the historical wind power and the historical price. The second is the forecasting rule for wind generation. Based on the two rules, we forecast the price via the forecasted generation and the first learned mapping between power and price. Additionally, we observed that it is not the more training data the better, leading to our validation steps to quantify the best training intervals for different datasets. We conduct comparisons of numerical results between existing methods and the proposed methods based on datasets from the Electric Reliability Council of Texas (ERCOT). For each machine learning step, we examine different learning methods, such as polynomial regression, support vector regression, neural network, and deep neural network. The results show that the proposed method is significantly better than existing approaches when renewables are involved.

## 1. Introduction

Since the industrial revolution, energy has become a key factor in everyday life [1]. Fossil fuels have become the most primary energy production in the world [1]. However, with the population growth and technological development, the current world is facing two vital problems, environmental pollution, and energy resource shortages [2]. One way to overcome problems is to improve efficiency and reduce emission

[3]. The other way is to develop alternate energy resources [2]. People draw their eyes to renewable resources for their properties of environmental-friendly and sustainability. The most competitive renewables include water, wind, photovoltaic energy, and biofuel. Many of them have been proved to be advanced in addressing environmental and energy issues [4,5]. Many renewables have been applied to the electricity market.

In the last few years, electricity market prices decreased a lot due to

\* Corresponding author.

E-mail addresses: [sluo31@asu.edu](mailto:sluo31@asu.edu) (S. Luo), [yang.weng@asu.edu](mailto:yang.weng@asu.edu) (Y. Weng).

**Table 1**  
Reference explanation.

Reference number	Contents
[1,2,4,3,5]	The reason why renewables entered the electricity market
[6–8]	The impact of renewables entering the electricity market on electricity prices
[9–11]	The difficulties to build a precise electricity price forecasting model
[13–15]	The difficulties to apply ‘hard computing techniques’ to build the electricity price forecasting model
[12]	The advantages to using ‘soft computing techniques’ to build the electricity price forecasting model
[16–18]	The high accuracy of the wind generation forecasting
[19,20]	The successful examples of employing machine learning methods into power systems

the close-to-zero marginal costs from renewable energies [6]. Therefore, the electricity market participants are seeking ways to be more competitive in the market. Many companies have adopted new electricity price plans [7], for example, time-of-use electricity price plans. These plans charge higher rates when demand is high, and lower rates when demand is low. This encourages customers to wisely decide their electricity usages and reduce on-peak energy usages [8]. This situation makes not only the producers but also the customers pursue more precise forecasts of the electricity market prices than ever. However, electricity price usually has complex features, such as highly volatile behavior and non-linearity, which makes it rather difficult to build a precise forecasting model [9–11].

In general, the electricity market price forecast has two classes of computing techniques. One is so-called ‘hard computing techniques’ [12], which can accurately predict the electricity prices if we know the exact model of the system. Time series models [13] and autoregressive integrated moving average (ARIMA) models [14] are two typical models. However, electricity prices are influenced by many factors, such as the volatile prices of generation resources, seasonal weather risks, the uncertain behavior of competitors in the market and so on [15]. These elements make it rather difficult to build the accurate model of the system. Besides, the solutions of hard computing techniques are solved according to physical regularity which needs high computation costs. Different from ‘hard computing techniques’, ‘soft computing techniques’ are proposed without needing to build the models of the systems [12]. This type of technique learns the mapping between the input and the output data, which needs less information and has higher computation efficiency [12].

Hence, we employ ‘soft computing techniques’, such as forecasting future real-time market (RTM) bus price using the historical bus price with different machine learning methods. However, this direct method from price to price has a relatively poor performance, no matter which learning method we try or if we do the validation step for hyper-parameters like training or testing data size. This is because the model only considers a single feature with limited information. In order to improve this method, we add another important data type, namely the wind power generation, which directly impact price variation. Additionally, we also redesign the forecasting model by leveraging the fact that wind generation forecasting is with very high accuracy, e.g., mean absolute percentage error is less than 5% [16–18]. Specifically, the proposed method learns two types of mapping rules. The first one is the mapping between the historical wind power generation and the historical price. The second is the forecasting rule for wind power generation. Based on the two rules, we forecast the price via the forecasted generation and the first learned mapping rule between power generation and price. We name the proposed method the rerouted method (two-stage method).

As a highlight, we examine the advantages and disadvantages of each machine learning method for both direct method (price-to-price method) and the rerouted method (two-stage method), so that we can select the best method with the best hyper-parameters for the benchmark. Specifically, we choose machine learning methods that are widely used in real-world applications [19,20], e.g., polynomial regression, support vector regression (SVR), neural network (NN), and

deep neural network (DNN).

For numerical validation, we use RTM bus price data and system-wide wind power generation data from the Electric Reliability Council of Texas (ERCOT). RTM bus price is the simple average of the time-weighted hub bus prices for each settlement interval in real-time, for each hub bus included in this hub. We preprocessed and removed some extreme data to make all the data in the normal range. The selected wind power generation contains the wind power generation all over the system. Simulation results show that the direct forecasting (price-to-price method) obtain its best testing accuracy when we employ polynomial regression. The rerouted method (two-stage method) obtain its best testing accuracy when we adopt deep learning. In general, the results show that the proposed method is significantly better than the direct forecasting (price-to-price method) when renewables are involved. All the references mentioned above can be better summarized up in Table 1.

Current research work indicates that we may obtain higher forecasting accuracy if we consider additional highly correlated data sources such as solar energy and biofuels. The NN and DNN used in this work are basic networks, future research can explore more of the network structure.

The rest of the paper is organized as follows: Section 2 formulates the forecasting problem. Section 3 describes the machine learning methods we use. Section 4 describes the simulation setup and the numerical results. Section 5 concludes the paper.

## 2. Problem formulation

In this section, we explain the direct method (price-to-price method) and the rerouted method (two-stage method) in detail using diagrams and mathematical formulas. To ensure an objective assessment of all the methods in this paper, we use the same dataset to test different approaches and models. The ideas are shown in Fig. 1.

### 2.1. Direct method (price-to-price method)

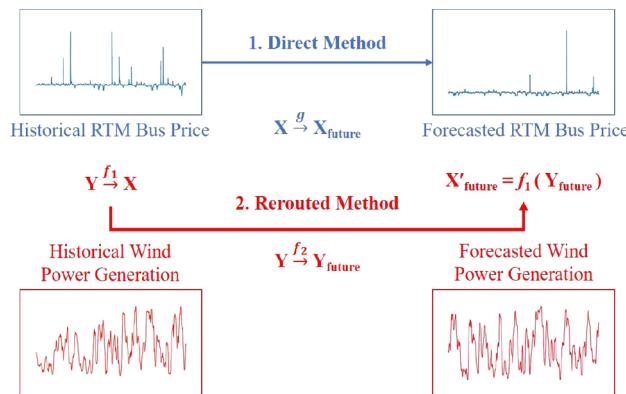
The problem is defined as: forecast the real-time market (RTM) bus price for the following month using the historical RTM bus price. Specifically, we first preprocessed and removed some extreme data to make all the data in the normal range. Then, let  $M$  be the size of the input data,  $N$  be the size of the output data, we adjust the size of  $M$  and  $N$  to find the best pairs that obtain the highest testing accuracy. The parameters are formulated as follows.

- **Input:** The input matrix is the RTM bus price from January 2016:  $X: M \times 1$ .

- **Output:** The output is the predicted RTM bus price for February 2016  $X_{\text{future}}: N \times 1$ , which is given by Eq. (1):

$$X_{\text{future}} = g(X), \quad (1)$$

where  $X_{\text{future}}$  is the prediction of future RTM bus price.  $g(\cdot)$  is the method chosen for forecasting.



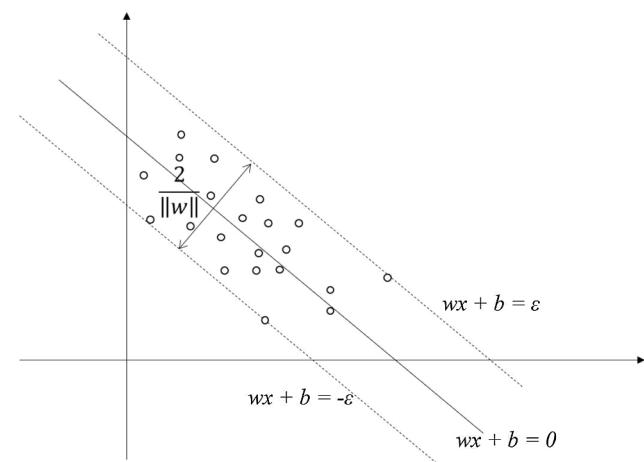
**Fig. 1.** The blue pictures represents the direct method (price-to-price method). We use function  $g$  to learn the mapping between  $X$  (Historical RTM Bus Price) and  $X_{\text{future}}$  (Forecasted RTM Bus Price). The red arrows and pictures form the rerouted method (two-stage method). The rerouted method (two-stage method) contains 3 steps. Step one is to use function  $f_1$  to learn the mapping between  $X$  and  $Y$  (Historical Wind Power Generation). Step two is to use function  $f_2$  to learn the mapping between  $Y$  and  $Y_{\text{future}}$  (Forecasted Wind Power Generation). Step three is to predict  $X_{\text{future}}$  using  $Y_{\text{future}}$  and the function  $f_1$  learned before. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

In order to get  $g(\cdot)$ , we use historical data ( $X$ ,  $X_{\text{future}}$ ) to learn the mapping. By adjusting the sizes of the historical data, we can determine the best mapping  $\hat{g}(\cdot)$  of the different methods presented in this paper.

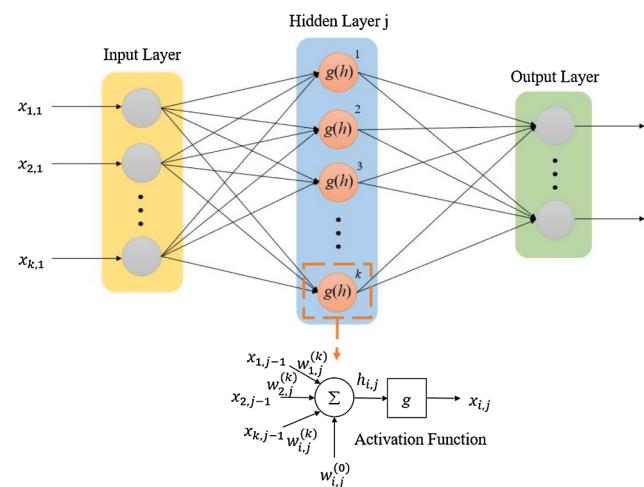
## 2.2. Rerouted method (two-stage method)

The problem is defined as: forecast the RTM bus price for the following month using the historical RTM bus price and the system-wide wind power generation. Specifically, the rerouted method contains 3 steps. The parameters are formulated as follows.

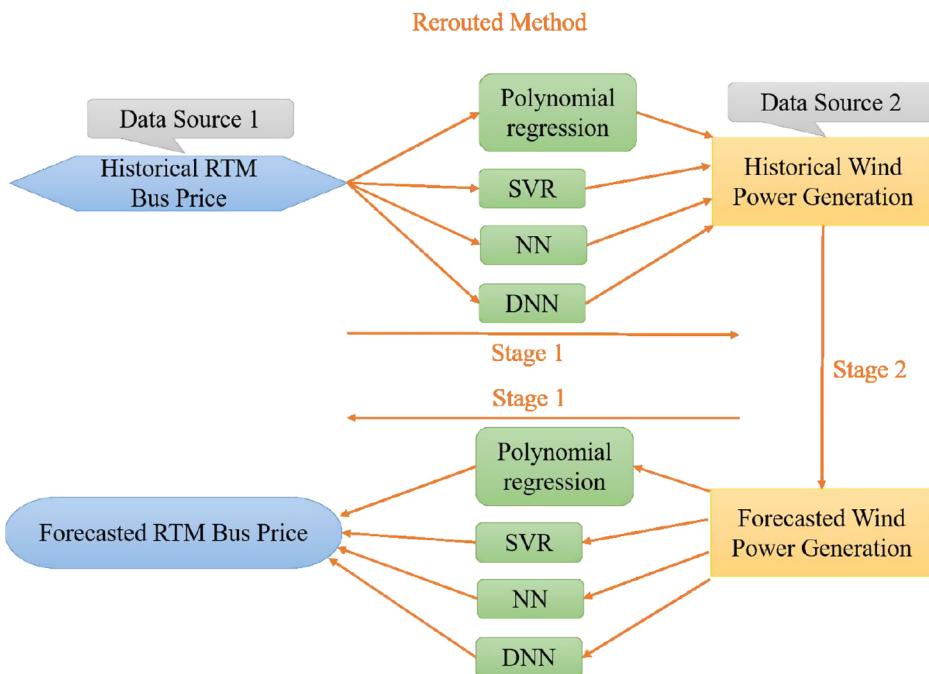
- **Input 1:** The input 1 matrix is RTM bus price from January 2016:  $X$ .
- **Input 2:** The input 2 matrix is system-wide wind power generation from January 2016:  $Y$ .



**Fig. 3.** SVR Diagram.



**Fig. 4.** Neural Network Diagram.



**Fig. 2.** The flow chart of the rerouted method.

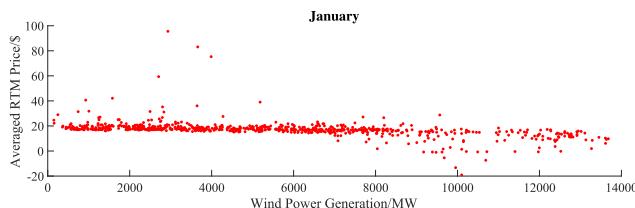


Fig. 5. Initial data distribution.

**Table 2**  
MSE comparison.

Method	Degree 1 polynomial regression	(a) Rerouted method		
		Degree 2 polynomial regression	Degree 3 polynomial regression	Degree 4 polynomial regression
Training	36.9341	36.5456	36.0694	35.9630
Testing	22.3945	23.0906	22.6684	22.2578
Method	NN with 30 hidden neurons	NN with 60 hidden neurons	Linear kernel SVR	Polynomial kernel SVR
Training	36.2109	33.4186	38.7031	37.9564
Testing	24.6695	26.23	25.6771	24.0645
Method	Gaussian kernel SVR	7 layers DNN	11 layers DNN	14 layers DNN
Training	38.1197	34.9439	36.8437	36.1379
Testing	23.7623	22.7658	23.2444	<b>20.5065</b>
Method	Polynomial regression	(b) Direct method		
		NN with 30 hidden neurons	NN with 60 hidden neurons	Linear kernel SVR
M	12	6	96	192
N	3	3	3	1
Training	40.6163	28.3315	23.4281	25.5157
Testing	<b>27.1534</b>	27.3736	30.0632	29.8577
Method	Polynomial kernel SVR	Gaussian kernel SVR	11 layers DNN	14 layers DNN
M	192	96	3	3
N	1	1	2	2
Training	25.5097	35.6567	23.3813	43.3681
Testing	29.865	37.1421	32.3289	32.1427

The bold values in table are the best testing results for each method.

- Step 1: We use historical data  $(Y, X)$  to learn the mapping function  $f_1(\cdot)$  between the historical system-wide wind power generation and the RTM bus price.
- Step 2: Let  $Y_{\text{future}}$  be the prediction of future system-wide wind power generation. We use historical data  $(Y, Y_{\text{future}})$  to learn the mapping function  $f_2(\cdot)$  between the historical system-wide wind power generation and the future system-wide wind power generation.

• **Output  $X_{\text{future}}$ :** (Step 3) We use the predicted wind power generation  $Y_{\text{future}}$  and the mapping function  $f_1(\cdot)$  learned in step 1 to predict the future price  $X_{\text{future}}$ . The output is given by Eq. (2):

$$X_{\text{future}} = f_1(Y_{\text{future}}), \quad (2)$$

where  $X_{\text{future}}$  is the prediction of RTM bus price for February 2016,  $f_1(\cdot)$  is the method chosen for forecasting.

Fig. 2 shows the flow chart of the rerouted method summarizing all

the processes.

### 3. Machine learning methods

In this section, we explain existing and popular machine learning methods for the proposed learning process in the last section.

#### 3.1. Methods overview

##### 3.1.1. Polynomial regression

In general, the polynomial regression model is given by Eq. (3):

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \cdots + \beta_m x_i^m + \epsilon_i, \quad i = 1, 2, \dots, n. \quad (3)$$

It can also be written as Eq. (4):

$$\vec{y} = X\vec{\beta} + \vec{\epsilon}, \quad (4)$$

where  $X$  is a design matrix,  $\vec{y}$  is a target vector,  $\vec{\beta}$  is a coefficient vector, and  $\vec{\epsilon}$  is a vector of random errors.

The vector of the estimated polynomial regression coefficient can be calculated using Eq. (5):

$$\vec{\beta} = (X^T X)^{-1} X^T \vec{y}, \quad m < n. \quad (5)$$

##### 3.1.2. Support vector regression (SVR)

SVR is a regression analysis of the data when we do the fitting [21]. It uses the idea of support vectors and the Lagrange multiplier. SVR constructs a hyperplane or set of hyperplanes in a high- or infinite-dimensional space by minimizing the margin on all the training data [22]. The support vector regression is obtained in Eq. (6).

$$\begin{aligned} \min \quad & F(w) = \frac{1}{2} \|w\|^2 \\ \text{s.t.} \quad & |y_i - (w^T x_i + b)| \leq \epsilon, \quad i = 1, 2, \dots, n, \end{aligned} \quad (6)$$

where  $x_i$  is a training sample with target value  $y_i$ .  $w^T x_i + b$  is the prediction for that sample and  $\epsilon$  is a free parameter that serves as a threshold. For example, all predictions have to be within an  $\epsilon$  range of the true predictions. The method can be better illustrated by Fig. 3.

The mapping of SVR to higher dimensions results in a serial of problems. It's hard to obtain the form of the mapping and to compute the coordinates of the data in that space. Hence, kernel methods are introduced to solve the problem. A kernel function can compute the dot product between the two mapping transforms in the feature space without knowing the mapping transform function itself. Assume  $X_i, X_j \in R_n$ , nonlinear function  $\Phi$  implements the mapping from input space  $X$  to feature space  $F$ , where  $F \subseteq R_m$ ,  $n \ll m$ . Refer to the kernel method, we have:

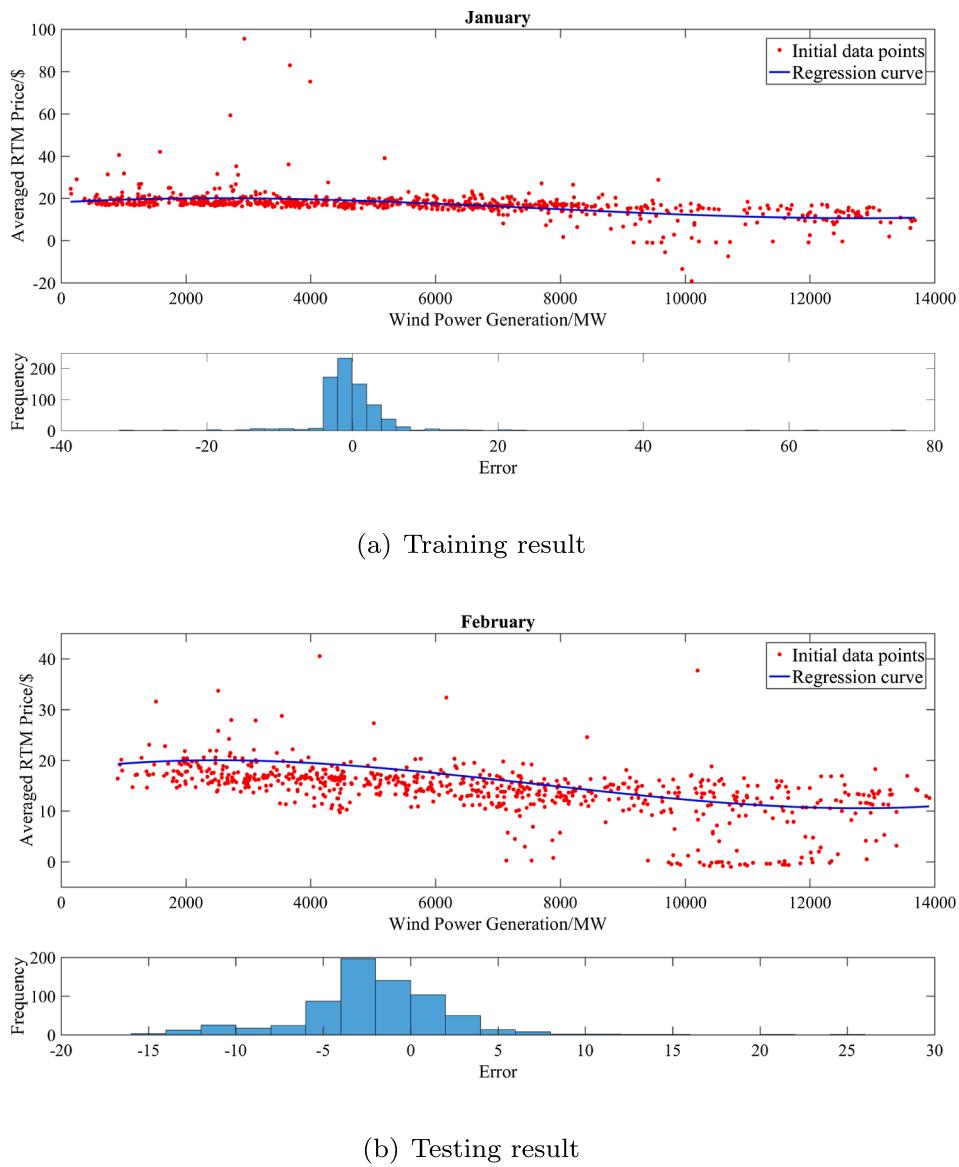
$$K(X_i, X_j) = \phi(X_i) \cdot \phi(X_j), \quad (7)$$

where  $K(X_i, X_j)$  is the kernel function. Commonly used kernel functions include linear kernel, polynomial kernel, and Gaussian kernel, also known as radial basis function (RBF) kernel.

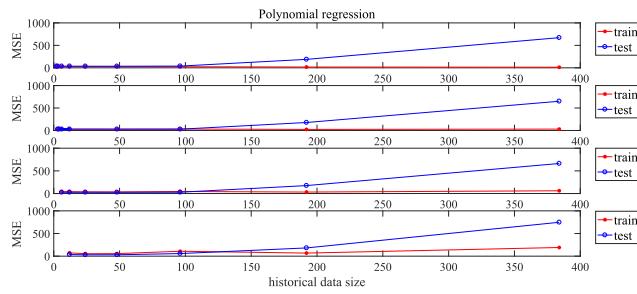
#### 3.1.3. Neural network (NN)

NNs are highly-interconnected-computing systems inspired by modern biology [23]. NNs are built up from a number of processing units, also called neurons. Each neuron is a weighted sum of the inputs formed by a linear function with a biased term [24]. The sum is then passed through a transfer function, also called an activation function, which is often a unit step, sigmoid and Gaussian [24]. Neurons can be grouped into layers. Typically, the first layer and the last layer of a basic NN is called the input layer and the output layer. The layers between the input and output layers are called the hidden layers.

NNs can be represented in Eq. (8):



**Fig. 6.** Rerouted method: degree of 3 polynomial regression.



**Fig. 7.** Direct method: data size determination.

$$x_{i,j} = g(h_{i,j}), \quad h_{i,j} = w_{i,j}^{(0)} + \sum_k w_{i,j}^{(k)} x_{k,j-1}, \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, m, \quad k = 1, 2, \dots, t, \quad (8)$$

where  $x_{i,j}$  is the input for the current layer,  $x_{k,j-1}$  is the input for the last layer,  $w_{i,j}^{(k)}$  is the weights of the  $k^{\text{th}}$  neuron,  $w_{i,j}^{(0)}$  is the biased term,  $g$  is the transfer function. The transfer function is introduced to increase the non-linearity. We conduct many experiments on different activation functions and find sigmoid function can achieve the highest accuracy. We also provide a diagram illustrating the structure of a basic NN

**Table 3**  
The MSEs of the polynomial regression.

Degree	(a) Rerouted method			
	1	2	3	4
Training	36.9341	36.5456	36.0694	35.9630
Testing	22.3945	23.0906	22.6684	<b>22.2578</b>
(b) Direct method				
$N$	1	2	3	6
$M$	48	96	12	24
Training	24.9384	31.8074	40.6163	54.1
Testing	35.2281	28.5926	<b>27.1534</b>	30.3579

The bold values in table are the best testing results for each method.

shown in Fig. 4.

Backpropagation (BP) is a method to calculate the gradient of the loss function (produces the cost associated with a given state) with respect to the weights in an artificial neural network (ANN) [25]. Backpropagation neural networks (BPNNs) have the ability to

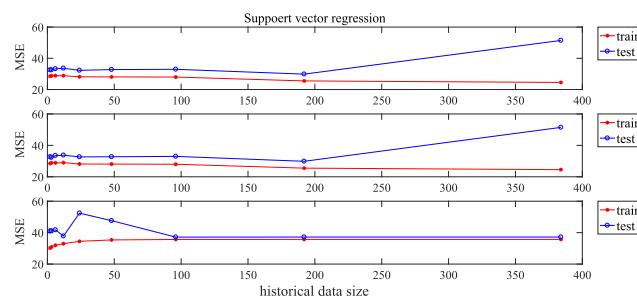


Fig. 8. Direct method: data size determination.

Table 4  
The MSEs of SVR.

(a) Rerouted method (two-stage method)			
MSE	Kernel		
	Linear	Polynomial	Gaussian
Training	38.7031	37.9564	38.1197
Testing	25.6771	24.0645	<b>23.7623</b>

(b) Direct method (price-to-price method)			
Kernel	Linear	Polynomial	Gaussian
<i>M</i>	192	192	96
Training MSE	25.5157	25.5097	35.6567
Testing MSE	<b>29.8577</b>	29.8650	37.1421

The bold values in table are the best testing results for each method.

implement any complex nonlinear mapping from input to output, to learn by themselves and to adapt to changes [26]. Furthermore, BPNNs have generalization ability and error tolerance. In the meanwhile, BPNNs have many shortcomings such as the local minimization problem. With different initializations of the weights, a BPNN will converge to different local minimums. So every time we train, we get a different result.

### 3.1.4. Deep learning

Deep learning is a class of machine learning algorithms that use multiple layers of nonlinear processing units for feature extraction and transformation [27]. Each successive layer uses the output from the previous layer as input. Most deep learning models nowadays are based on ANNs [28]. It is very time-consuming to train a model, and the validation is very complex and troublesome. However, a well trained deep learning model can be easily applied to other problems by doing some simple refinements.

### 3.1.5. Methods comparison

In the polynomial regression, all the features are determined by us, which may contain useless features. Hence, NNs and DNNs are brought up for not needing to decide how to construct the features. We directly input the raw data, if we achieve high accuracy, the model is useful. However, NNs and DNNs involve the random initialization of weights. So training on the same data may give different results. Besides, considerable parameters are set concerning the architecture of the ANNs as well as the learning algorithms. The optimizations of these parameters can only be carried out through a trial-and-error process which consumes much time and resources [29]. The training of an SVR is a convex quadratic optimization which has one unique solution and it does not involve the random initialization of weights like NNs and DNNs [30]. Any SVR with the same parameter settings trained on identical data will give the same results. This greatly reduces the number of training required to find the optimum.

### 3.2. Performance evaluation metric

The performances of all the methods are determined by the mean square errors (MSEs). Let  $K$  be the size of the output, the computational formula is defined as follows:

$$\text{MSE} = \frac{1}{N} \sum_{t=1}^N (\hat{Y}_{t,1} - y_{t,1})^2, \quad \hat{Y}_{t,1} = \begin{cases} \frac{1}{t} (\hat{Y}_{1,t} + \hat{Y}_{2,t-1} + \dots + \hat{Y}_{t,1}), & \text{if } t \leq K-1 \\ \frac{1}{K} (\hat{Y}_{t-K+1,K} + \hat{Y}_{t-K+2,K-1} + \dots + \hat{Y}_{t,1}), & \text{if } t > K-1 \end{cases} \quad (9)$$

where  $\hat{Y}_{t,1}$  is the forecasted price at hour  $t$ ,  $y_{t,1}$  is the real price at hour  $t$ , and  $N$  is the number of the total hours.

## 4. Numerical results

### 4.1. Data preparation

Electric Reliability Council of Texas (ERCOT) is an independent system operator managing about 90 percent of the states electric load. ERCOT made significantly large investments in the renewable energy sector, particularly in wind energy and continues to lead as the top wind production in the nation [31]. ERCOT has an adequate market and grid information which can be easily accessed and downloaded from its website [32]. If you need some specific range of data that is not available on the website, you can contact ERCOT by submitting an information request form [33]. ERCOT is eager to help and responds quickly. The raw data we get from ERCOT is excel files containing the information of all districts. We extract all the data we need and build vectors of RTM price and system-wide wind power generation that are hourly measured. To ensure the RTM price data in the normal range, let  $\mu$  be the mean of the data and  $T$  be the threshold, the normal range is defined as:  $\mu \pm T$  in our specific problem.

### 4.2. Benchmark

For rerouted method (two-stage method), we do the following simulations using electricity price data and the data of system-wide wind power generation. We use a  $744 \times 1$  vector of system-wide wind power generation as the input data and a  $744 \times 1$  vector of real-time market (RTM) bus price as the target data for training. Both data came from the same time slot that is January 2016 which contains  $31 \text{ (days)} \times 24 \text{ (hours)} = 744$  data points. The training data can be visualized in Fig. 5, the x-axis is the system-wide wind power generation from January 2016 and y-axis is the RTM bus price from January 2016.

For direct method (price-to-price method), let  $M$  be the input data size,  $N$  be the output data size. By adjusting these two hyper-parameters, we are able to find the best  $M$  and  $N$  that make the mapping between the prices reach the highest accuracy. In this method,  $M$  is chosen from 2, 3, 6, 12, 24, ..., 384; and  $N$  is chosen from 1, 2, 3, 6. We start from 2 for  $M$ , because if we only use one historical data to predict one or more data, the uncertainty is so huge that a high accuracy is hard to achieve. For other numbers, we let the latter number be twice the former to study the tendency of the testing accuracy.

General results show that the rerouted method can guarantee better accuracies when compared to the direct method, for all the machine learning methods we used. This statement can be confirmed by Table 2, where we compare the results of both methods. As we can observe in Table 2, the rerouted method gain its highest accuracy when 14-layer DNN is used. To ensure consistency, the direct method also employs the same machine learning methods. And the result shows that it obtains its highest frequency when using polynomial regression. The detailed results and comparisons will be listed in the following subsections.

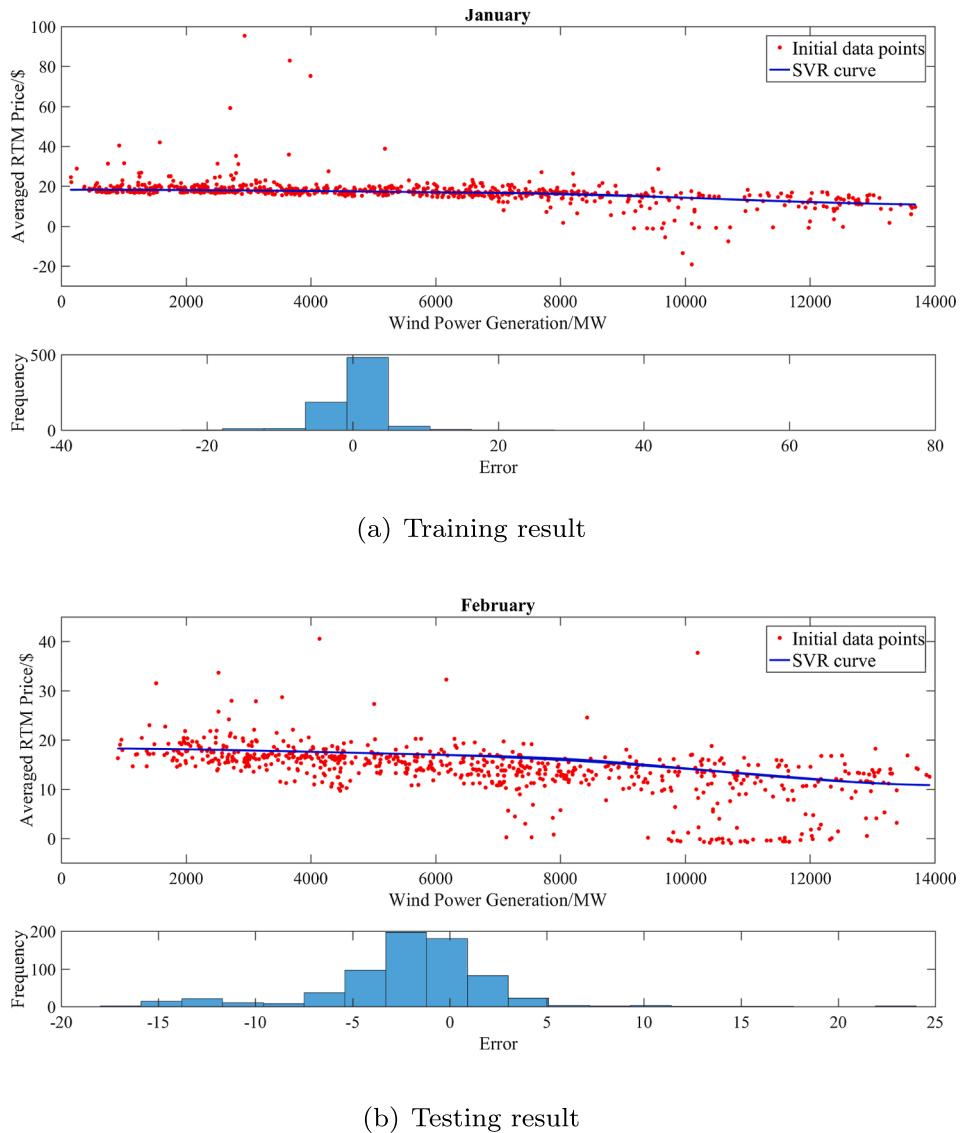


Fig. 9. Rerouted method: Gaussian kernel SVR.

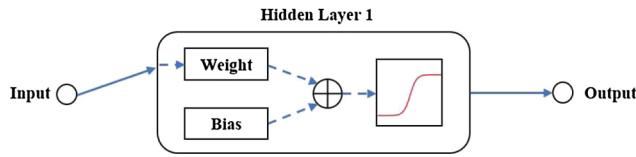


Fig. 10. The structure of the NN.

#### 4.2.1. Polynomial regression

For rerouted method (two-stage method), we use the system-wide wind power generation from January 2016 as the training data and that from February 2016 as the testing data. We vary the degrees of the polynomial model from 1 to 4 and pick degree of 3 as our example. Fig. 6 represents the training and testing regression curves along with the error histograms. The training and testing regression curves along with the error histograms of other polynomial models are shown in Appendix A. For direct method (price-to-price method), we fix the predicting data size  $N$  for each time and adjust the historical data size  $M$  to obtain the best testing results. The process of determining the size of

the data can be illustrated in Fig. 7. As shown in Fig. 7, the training mean squared error (MSE) is fluctuating around the minimum value when we increase the historical data size  $M$ , while the testing MSE becomes extremely large. This reveals the overfitting problem when we perform the training of the data set. The highest testing accuracy of each figure is determined by the gap between the training MSE and the testing MSE. We select the ones with the smallest gaps and merge them into Table 3(a) in order to be compared with the rerouted method. The MSEs for both methods are shown in Table 3.

As we can see from Table 3(a), when we increase the order of the polynomial regression the testing accuracy does not change much, which means that the data has great linearity as a whole. Combined with Fig. 7 and Table 3(b), the rerouted method has smaller testing MSE than any of the direct method no matter how you resize  $M$  and  $N$ . In this case, when we use the rerouted method and polynomial regression with degree 1 or 4, we can obtain good testing performance. As you can observe from Fig. 6, the polynomial regression with a reasonable degree has a poor behavior detecting outliers. Hence, we employ support vector regression (SVR) for our next step to see if it can reach a higher

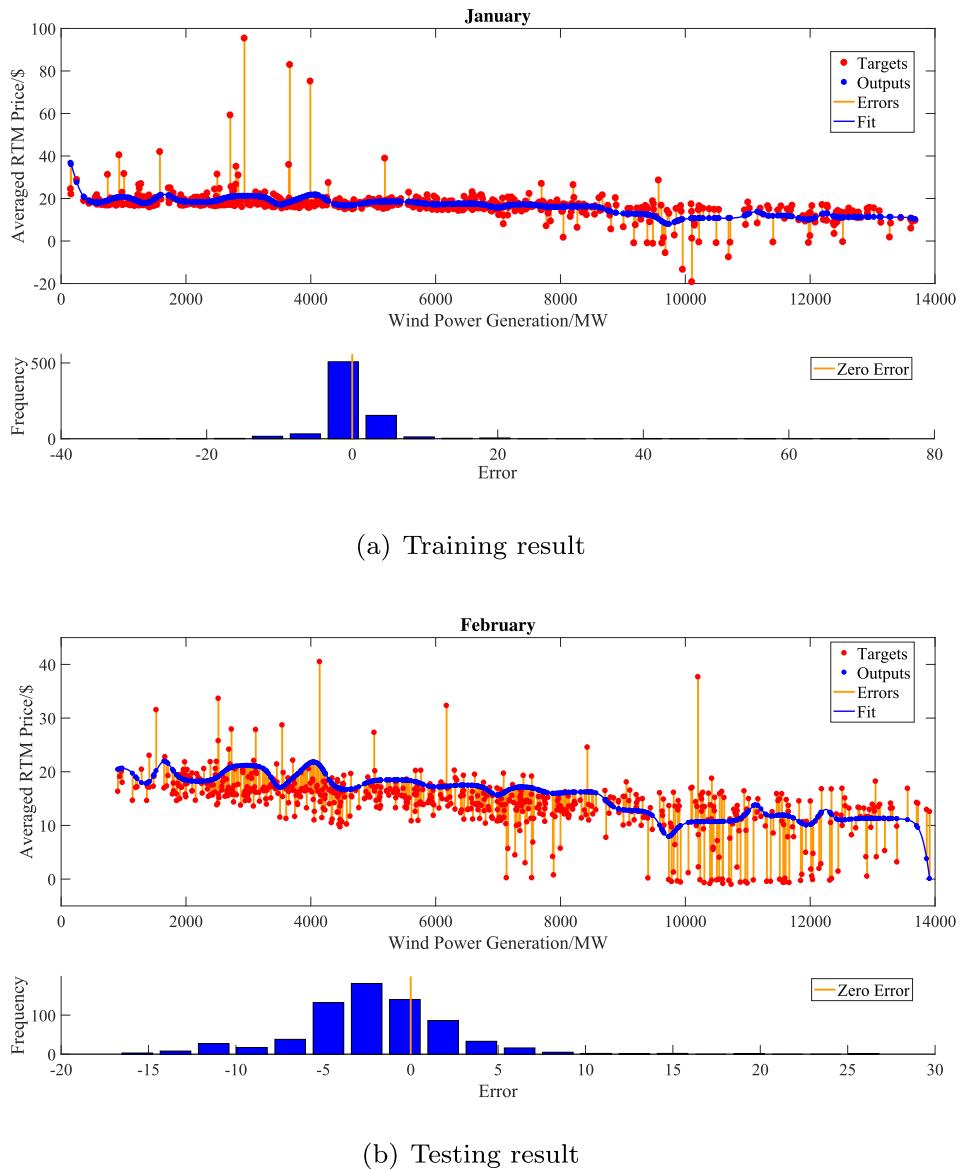


Fig. 11. Rerouted method: NN with 30 hidden neurons.

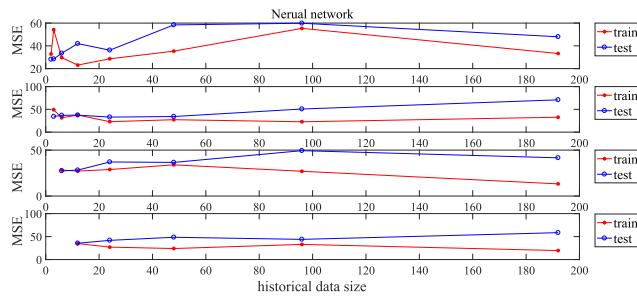


Fig. 12. Direct method: data size determination.

**Table 5**  
Direct method: the best testing results according to  $N$ .

N	1	2	3	6
M	2	24	6	12
Training MSE	32.8374	22.9490	28.3315	34.7806
Testing MSE	28.1420	32.9945	<b>27.3736</b>	35.4926

The bold values in table are the best testing results for each method.

accuracy when we map the data into higher dimensions.

#### 4.2.2. Support vector regression (SVR)

For both methods, we use command ‘fitrsvm’ in MATLAB and choose three different kernel functions: linear kernel, polynomial kernel and Gaussian kernel to do SVR. For the direct method (price-to-price method), the predicting data size  $N$  is fixed to be 1 and the historical data size  $M$  is adjustable. The procedure is displayed in Fig. 8. The determination of the best testing accuracy obeys the same rule stated in the last section. In order to be better compared with the rerouted method, the highest testing accuracy of each figure is merged into Table 4(b). For the rerouted method (two-stage method), we choose the SVR with Gaussian kernel as an example. The training and testing regression curves along with the error histograms are shown in Fig. 9. The results for other kernels can be found in Appendix A. The MSEs for both methods of SVR with 3 different kernels are shown in Table 4.

It's clearly shown in Table 4 that when we adopt the rerouted method, all the SVR with different kernels have similar training and testing MSEs and the testing MSE of the SVR with Gaussian kernel is slightly better than all the others. The rerouted method has smaller testing MSEs than any of the direct method. In this case, when we use

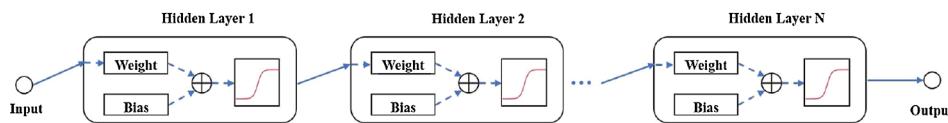
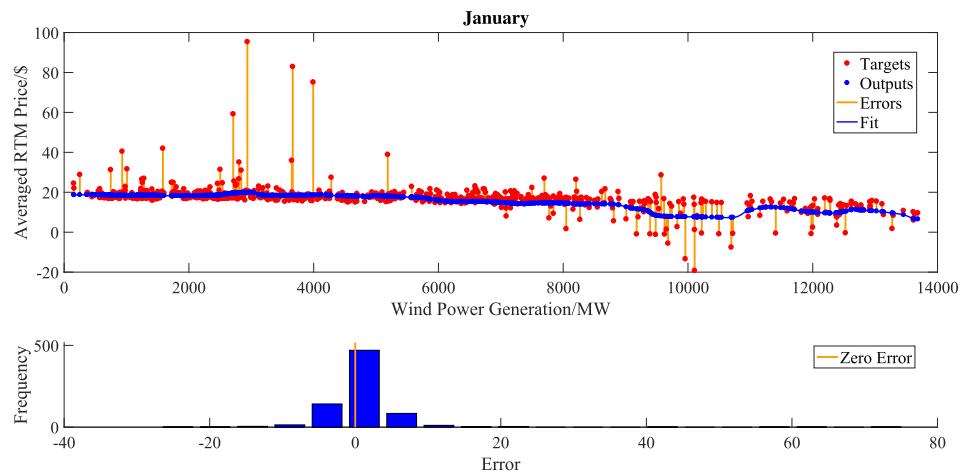
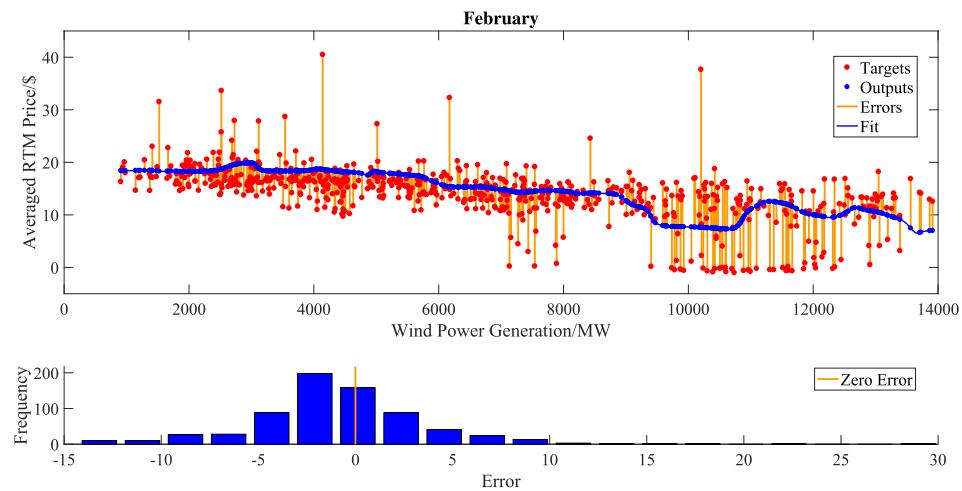


Fig. 13. The structure of the DNN.



(a) Training result



(b) Testing result

Fig. 14. Rerouted method: DNN with 14 layers.

**Table 6**  
The MSEs of the deep neural networks.

MSE	Layers		
	7	11	14
Training	34.9439	36.8437	36.1379
Testing	22.7658	23.2444	<b>20.5065</b>

The bold values in table are the best testing results for each method.

the rerouted method and SVR with Gaussian kernel, we can obtain the best testing performance. Compared to Table 3, the testing MSEs become worse indicating that SVR ignores outliers. Hence, we employ neural networks to make sure they capture the outliers.

**Table 7**  
The MSEs of other months.

Method	Index	Month			
		April	May	June	July
Direct	MSE	81.5258	49.5156	120.1935	159.2903
Rerouted	MSE	65.7889	41.4192	56.4694	53.2651

**Table 8**  
Model endurance towards noise.

Noise level	5%	10%	12%	15%
MSE (Feb)	22.7429	25.2443	26.0906	29.4196

#### 4.2.3. Neural network (NN)

Past researches have shown great success in forecasting electricity price using NNs [9,29,24]. Drawing on their ideas, we conduct comparisons between the direct method and the rerouted method based on the same NN. For both methods, we use command ‘nntool’ in MATLAB to build a simple NN with 3 layers (1 input layer, 1 hidden layer, and 1 output layer). The hidden layer size is 30 and the NN is trained by the Levenberg–Marquardt algorithm. The structure of the NN is shown in Fig. 10.

For rerouted method (two-stage method), the training and testing regression curves along with the error histogram are shown in Fig. 11. The training MSE is 34.8300 and the testing MSE is 24.0220. For direct method (price-to-price method), we fix the predicting data size  $N$  for each time and adjust the historical data size  $M$  to obtain the best testing results. The procedure is shown in Fig. 12 and the best results for each case are shown in Table 5.

As we can observe from Table 5, the rerouted method can reach a smaller testing MSE than any of the direct method no matter how you resize  $M$  and  $N$ . In this case, when we use the rerouted method and NN with 30 hidden neurons, we can obtain the best testing performance. Compared to the results of the polynomial regression, we can see that this neural network has a better training MSE, but a worse testing MSE. Therefore, we try to find out if we can gain a better testing result by increasing the hidden layer size from 30 to 60. The simulation result is shown in Appendix A. The MSE for training is 33.4186, the MSE for testing is 26.3436. The result shows that when we increase the hidden layer size, we get a better training MSE, but a worse testing MSE. This means that increasing the sizes of the hidden neurons does not help us to obtain a better testing MSE but results in an over-fitting in this specific problem.

#### 4.2.4. Deep neural network (DNN)

Owing to the poor performance of the simple neural networks, we employ DNNs to see if they can do better. DNNs are known for their powerful ability to learn the essential features of data sets from a small sample set. We find that there are not many papers make use of DNN to forecast electricity price. Therefore, we hold a detailed discussion here to present the advance of using DNN.

For rerouted method (two-stage method), we use ‘nntool’ in MATLAB to build DNNs with 3 different number of layers. The first DNN has 14 layers: 1 input layer, 13 hidden layers, and 1 output layer. The second DNN has 11 layers: 1 input layer, 10 hidden layers, and 1 output layer. The last DNN has 7 layers: 1 input layer, 6 hidden layers, and 1 output layer. All of them has a layer size of 30 for each hidden layer. The transfer function for the last hidden layer is ‘purelin’, the transfer function for all the other hidden layers is ‘tansig’. All of the DNNs listed above are trained by the Levenberg–Marquardt algorithm. The structure of the DNN is shown in Fig. 13.

We take the deep neural network with 14 layers as an example to illustrate the results. The training and testing regression curves along with the error histogram are shown in Fig. 14. Simulation results for the deep neural networks with 7 and 11 layers can be found in Appendix A. The MSEs of the deep neural networks are shown in Table 6.

As we can see from Table 6, when we increase the number of hidden layers, the training MSE doesn’t increase much but the testing MSE becomes smaller. This indicates that we may achieve better testing MSEs if we add more hidden layers. However, the training time will surely be longer if we increase the number of hidden layers. Hence,

there’s a trade-off between the number of layers and the time for training. Therefore, we should carefully choose the number of layers to get relatively good results.

#### 4.2.5. Additional discussion

In order to testify the effectiveness and the generalization ability, real-time market(RTM) price and wind power generation from other months are chosen to perform the verification. All the simulation results are shown in Table 7 and it is obvious that the proposed two-stage method is better than the direct method.

In addition, to testify the effectiveness and the generalization ability, we also make use of Monte Carlo tools to test the model stability against noise. As wind power generation will be influenced by the noise, we add different noise levels to test the model endurance towards the noise. The experimental results are shown in Table 8. Recalling previous results, the direct method can achieve the best testing accuracy of 27.1534. Compared to Table 8, we can draw the conclusion that the proposed method can endure at most 12% noise from the system and environment.

## 5. Conclusion

This paper develops a novel two-stage method to forecast the real-time market (RTM) price. This new method, namely the rerouted method (two-stage method), predict the future price using historical RTM bus price along with system-wide wind power generation. The main contributions of this work are the diversified input data sources such as highly correlated power data and the validation step to quantify the best training interval for different data set.

By conducting a comparison to the conventional method, namely direct method (price-to-price method), we confirm our conjecture that we can obtain higher accuracy if we diversify the data source. Furthermore, when we examine the relationship between the input and the output, we find that they are actually in a causal relationship. This causal relationship combined with some physical models can guarantee us with better results.

To verify the effectiveness and the generalization ability of the model, we conduct simulations over the other four months. The result shows that the proposed method is much more accurate than the direct method. To further explore the model stability against noise, we set up different noise levels of the wind power generation. The results show that the proposed model also has good stability towards the noise.

Other related subjects of interest for further research could be the improvement of the prediction accuracy by taking into consideration other renewable energies, solar energy most likely. These features should have causal relationships towards electricity price. In this paper, most of the methods are simple models that do not have a lot of parameters. We can develop more complex models to achieve better results.

## Acknowledgment

The authors would like to thank the Electric Reliability Council of Texas (ERCOT) for their supply of wind power generation data and real-time market price data. Special thanks go to the anonymous reviewers for their valuable comments and suggestions which have improved the quality of this paper.

## Appendix A. Detailed results for rerouted method

### A.1. Polynomial regression

Fig. A.15 shows the training regression curves and the error histograms of degree of 1, 2, 4.

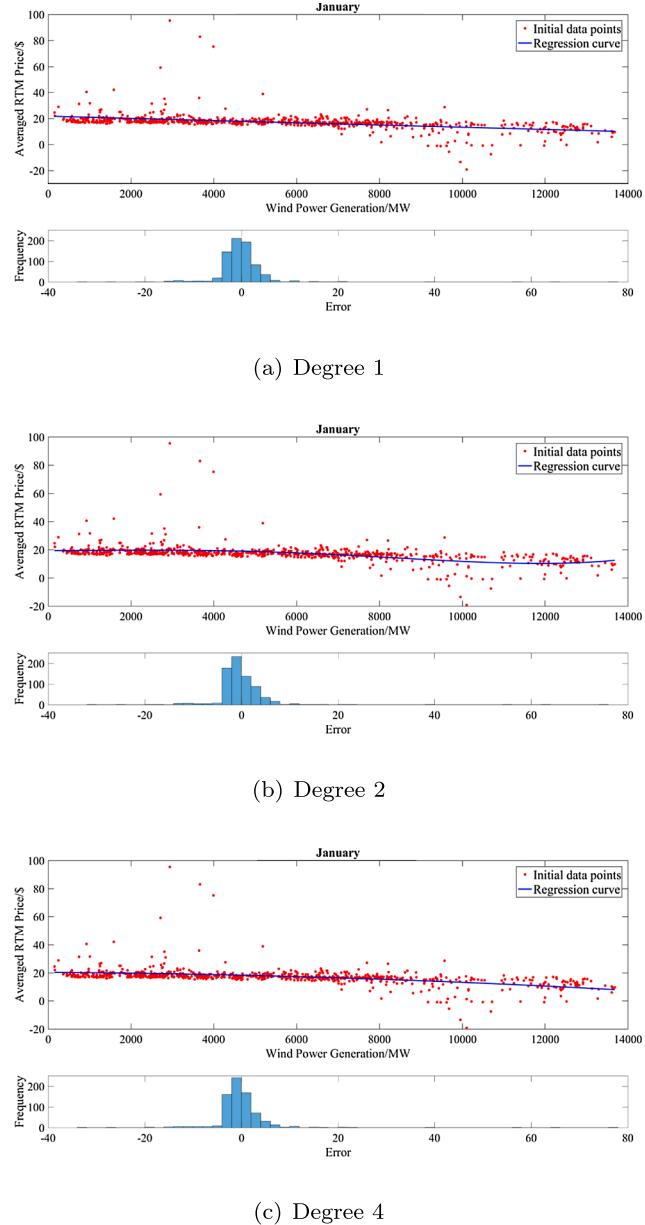


Fig. A.15. Training results of other orders polynomial regression.

Fig. A.16 shows the testing regression curves and the error histograms of degree of 1, 2, 4.

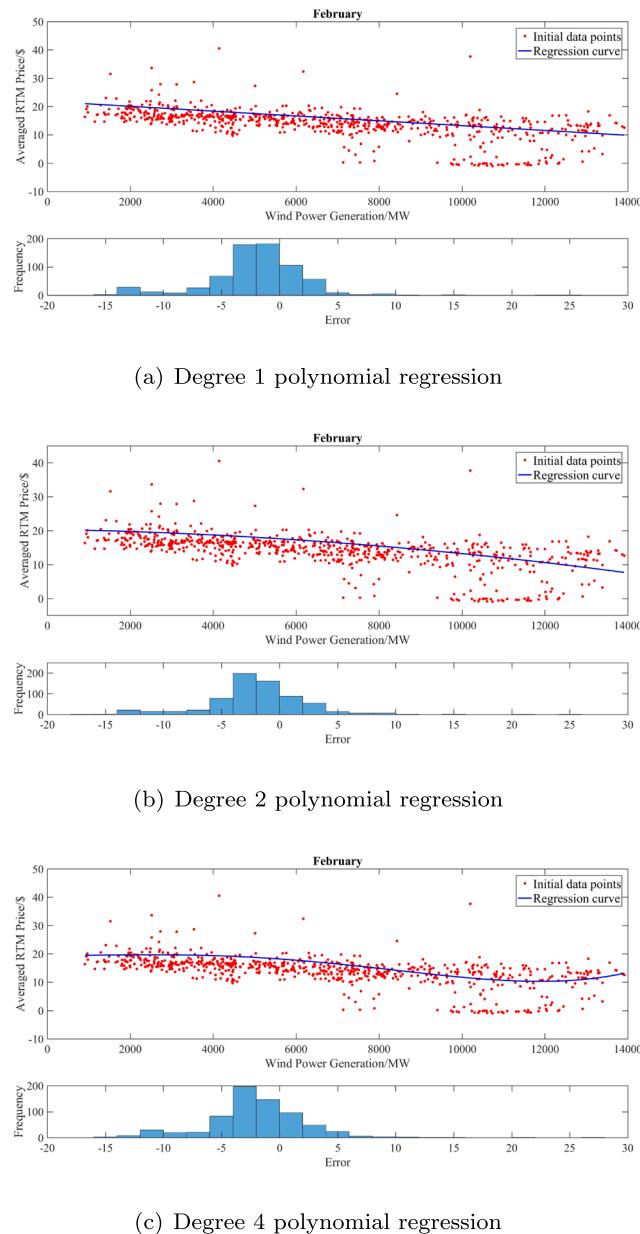
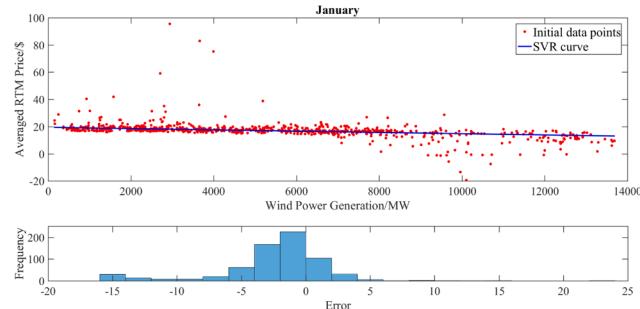


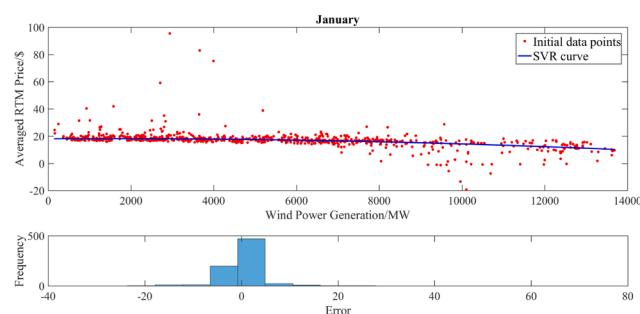
Fig. A.16. Testing results of other orders polynomial regression.

#### A.2. Support vector regression (SVR)

Fig. A.17 shows the training regression curves along with the error histograms of the SVR with linear kernel and polynomial kernel.



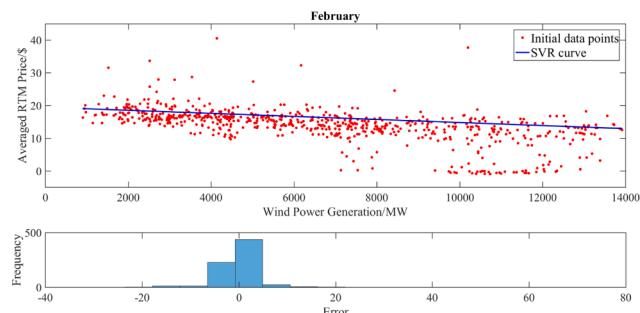
(a) Linear kernel SVR



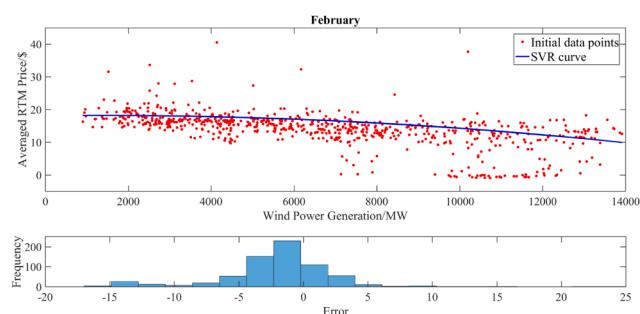
(b) Polynomial kernel SVR

Fig. A.17. Training results.

Fig. A.18 shows the testing regression curves along with the error histograms of the SVR with linear kernel and polynomial kernel.



(a) Linear kernel SVR

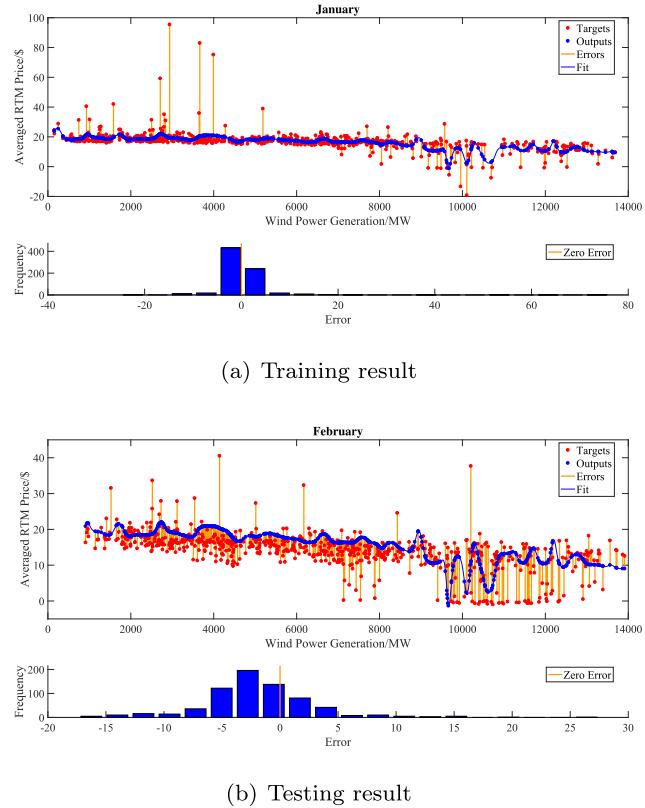


(b) Polynomial kernel SVR

Fig. A.18. Testing results.  
1509

### A.3. Neural network

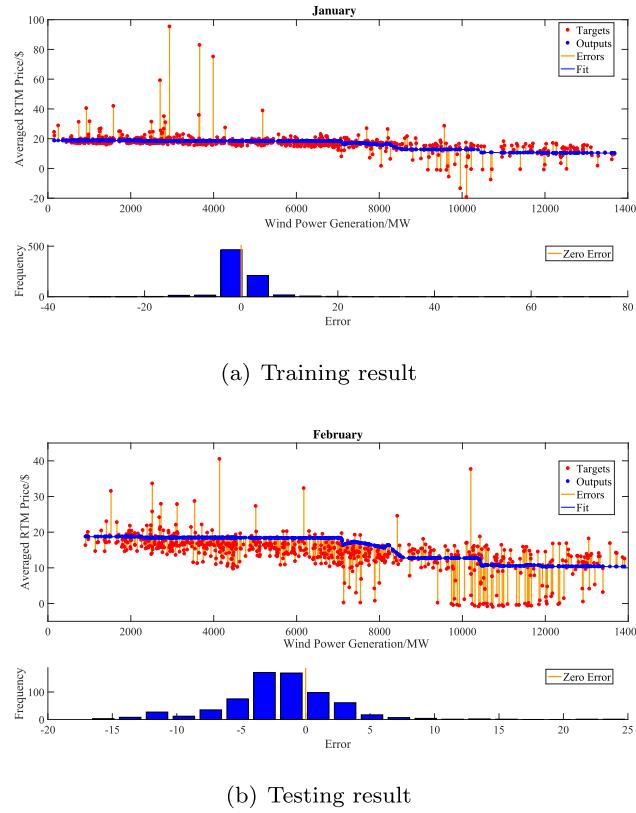
The neural network has 3 layers (1 input layer, 1 hidden layer, and 1 output layer). The hidden layer size is 60. The neural network is trained by the Levenberg–Marquardt algorithm. Fig. A.19 shows the training and testing regression curves along with the error histograms.



**Fig. A.19.** Rerouted method: NN with 60 hidden neurons.

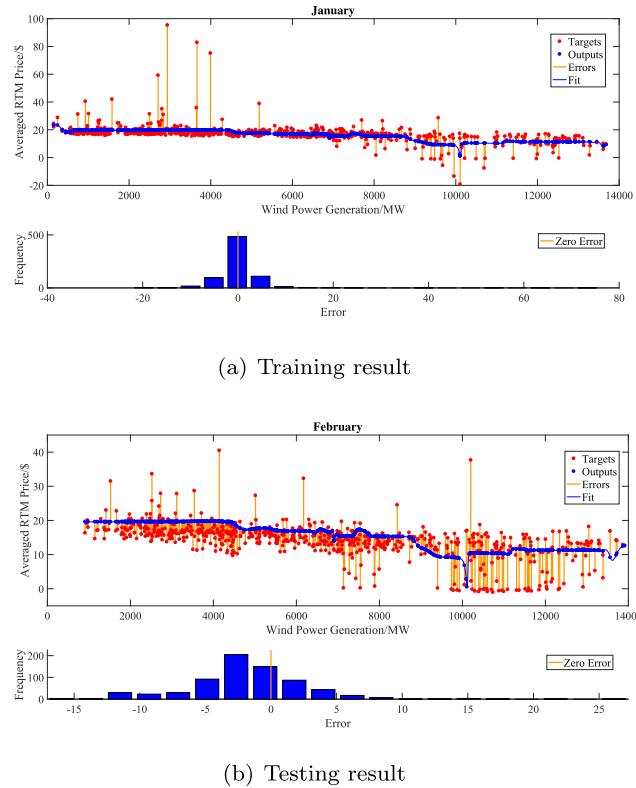
### A.4. Deep neural network

The deep neural network has 11 layers: 1 input layer, 10 hidden layers, and 1 output layer. The hidden layer size is 30. The transfer function for the last hidden layer is ‘purelin’, the transfer function for all the other hidden layers is ‘tansig’. The deep neural network is trained by the Levenberg–Marquardt algorithm. Fig. A.20 shows the training and testing regression curves along with the error histograms.



**Fig. A.20.** Rerouted method: DNN with 11 layers.

The deep neural network has 7 layers: 1 input layer, 6 hidden layers, and 1 output layer. The hidden layer size is 30. The transfer function for the last hidden layer is ‘purelin’, the transfer function for all the other hidden layers is ‘tansig’. The deep neural network is trained by the Levenberg–Marquardt algorithm. Fig. A.21 shows the training and testing regression curves along with the error histograms.



**Fig. A.21.** Rerouted method: DNN with 7 layers.

## References

[1] Lin Lin, Cunshan Zhou, Vittayapadung Saritporn, Xiangqian Shen, Mingdong Dong. Opportunities and challenges for biodiesel fuel. *Appl Energy* 2011;88(4):1020–31.

[2] Dhinesh B, Raj Y Maria Ambrose, Kalaiselvan C, KrishnaMoorthy R. A numerical and experimental assessment of a coated diesel engine powered by high-performance nano biofuel. *Energy Convers Manage* 2018;171:815–24.

[3] Vigneswaran R, Annamalai K, Dhinesh B, Krishnamoorthy R. Experimental investigation of unmodified diesel engine performance, combustion and emission with multipurpose additive along with water-in-diesel emulsion fuel. *Energy Convers Manage* 2018;172:370–80.

[4] Dhinesh B, Annamalai M. A study on performance, combustion and emission behaviour of diesel engine powered by novel nano nerium oleander biofuel. *J Clean Prod* 2018.

[5] Nomura Noboru, Inaba Atsushi, Tonooka Yutaka, Akai Makoto. Life-cycle emission of oxidic gases from power-generation systems. *Appl Energy* 2001;68(2):215–27.

[6] Sánchez de la Nieta AA, González V, Contreras J. Portfolio decision of short-term electricity forecasted prices through stochastic programming. *Energies* 2016;9(12):1069.

[7] Najafi A, Falaghi H, Contreras J, Ramezani M. Medium-term energy hub management subject to electricity price and wind uncertainty. *Appl Energy* 2016;168:418–33.

[8] Yang P, Tang G, Nehorai A. A game-theoretic approach for optimal time-of-use electricity pricing. *IEEE Trans Power Syst* 2013;28(2):884–92.

[9] Wang D, Luo H, Grunder O, Lin Y, Guo H. Multi-step ahead electricity price forecasting using a hybrid model based on two-layer decomposition technique and bp neural network optimized by firefly algorithm. *Appl Energy* 2017;190:390–407.

[10] García-Martos C, Rodríguez J, Sánchez MJ. Modelling and forecasting fossil fuels, co2 and electricity prices and their volatilities. *Appl Energy* 2013;101:363–75.

[11] Yang Z, Ce L, Lian L. Electricity price forecasting by a hybrid model, combining wavelet transform, arma and kernel-based extreme learning machine methods. *Appl Energy* 2017;190:291–305.

[12] Catalão JPdS, Mariano SJPS, Mendes V, Ferreira L. Short-term electricity prices forecasting in a competitive market: A neural network approach. *Electric Power Syst Res* 2007;77(10):1297–304.

[13] Nogales FJ, Contreras J, Conejo AJ, Espínola R. Forecasting next-day electricity prices by time series models. *IEEE Trans Power Syst* 2002;17(2):342–8.

[14] Contreras J, Espínola R, Nogales FJ, Conejo AJ. Arima models to predict next-day electricity prices. *IEEE Trans Power Syst* 2003;18(3):1014–20.

[15] Bello A, Bunn DW, Renes J, Muñoz A. Medium-term probabilistic forecasting of electricity prices: A hybrid approach. *IEEE Trans Power Syst* 2017;32(1):334–43.

[16] Ma J, Yang M, Han X, Li Z. Ultra-short-term wind generation forecast based on multivariate empirical dynamic modeling. *IEEE Trans Ind Appl* 2018;54(2):1029–38.

[17] Wan C, Xu Z, Pinson P, Dong ZY, Wong KP. Probabilistic forecasting of wind power generation using extreme learning machine. *IEEE Trans Power Syst* 2014;29(3):1033–44.

[18] Dowell J, Pinson P. Very-short-term probabilistic wind power forecasts by sparse vector autoregression. *IEEE Trans Smart Grid* 2016;7(2):763–70.

[19] Che J, Wang J. Short-term electricity prices forecasting based on support vector regression and auto-regressive integrated moving average modeling. *Energy Convers Manage* 2010;51(10):1911–7.

[20] Szkuta B, Sanabria LA, Dillon TS. Electricity price short-term forecasting using artificial neural networks. *IEEE Trans Power Syst* 1999;14(3):851–7.

[21] Drucker H, Burges CJ, Kaufman L, Smola AJ, Vapnik V. Support vector regression machines. *Advances in neural information processing systems*. 1997. p. 155–61.

[22] Chen Y, Xu P, Chu Y, Li W, Wu Y, Ni L, et al. Short-term electrical load forecasting using the support vector regression (svr) model to calculate the demand response baseline for office buildings. *Appl Energy* 2017;195:659–70.

[23] van Gerven M, Bohte S. Frontiers research topic: artificial neural networks as models of neural information processing, frontiers in computational neuroscience, Retrieved 2018-02-20, <https://www.frontiersin.org/research-topics/4817/artificial-neural-networks-as-models-of-neural-information-processing>.

[24] Bento P, Pombo J, Calado M, Mariano S. A bat optimized neural network and wavelet transform approach for short-term price forecasting. *Appl Energy* 2018;210:88–97.

[25] Goodfellow I, Bengio Y, Courville A. Deep learning. MIT Press; 2016.

[26] Panapakidis IP, Dagoumas AS. Day-ahead electricity price forecasting via the application of artificial neural network based models. *Appl Energy* 2016;172:132–51.

[27] Deng Li, Yu Dong. Deep learning: methods and applications. *Found Trends® Signal Process* 2014;7(3–4):197–387.

[28] Yoshua Bengio. Learning deep architectures for ai. *Found Trends® Machine Learn* 2009;2(1):1–127.

[29] Keles D, Scelle J, Paraschiv F, Fichtner W. Extended forecast methods for day-ahead electricity spot prices applying artificial neural networks. *Appl Energy* 2016;162:218–30.

[30] Sansom Damien C, Downs Tom, Saha Tapan K. Evaluation of support vector machine based forecasting tool in electricity price forecasting for australian national electricity market participants. *J Electr Electron Eng, Australia* 2003;22(3):227.

[31] Tao HYS, Srivastava AK, Pineda RL, Mandal P. Wind power generation impact on electricity price in ercot. 2012 IEEE Power and Energy Society General Meeting. IEEE; 2012. p. 1–7.

[32] <http://www.ercot.com/>, accessed 20-November-2018.

[33] <http://www.ercot.com/about/contact/inforequest>, accessed 20-November-2018.