Applied Energy 185 (2017) 1769–1777

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

Modeling a hybrid methodology for evaluating and forecasting regional energy efficiency in China

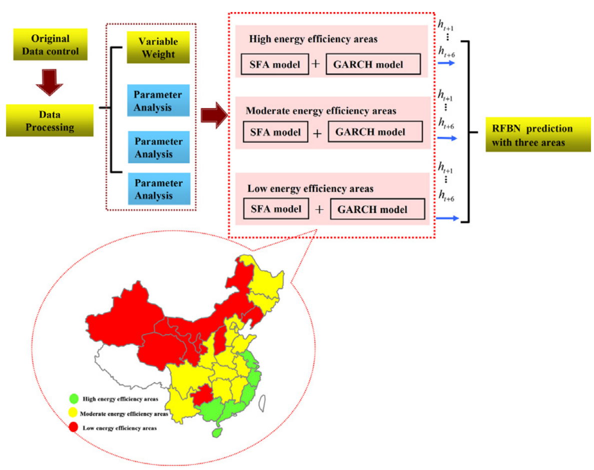
Ming-Jia Li, Ya-Ling He, Wen-Quan Tao *

Key Laboratory of Thermo-Fluid Science and Engineering of Ministry of Education, School of Energy and Power Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China

HIGHLIGHTS

- A new hybrid methodology consists of SFA-GARCH model and RBFN model is structured.
- Regional energy efficiency in China is measured during 2003–2014.
- Short-term forecast is examined without manual intervention from 2016 to 2020.
- The hybrid methodology avoids the superposition of errors of the individual forecasts.
- The 30 regions in China are clustered into high, moderate and low efficiency areas.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 14 August 2015
Received in revised form 6 November 2015
Accepted 26 November 2015
Available online 23 December 2015

Keywords:
Energy efficiency indicator
Cluster areas
Radial basis function neural
GARCH model
SFA model

ABSTRACT

This study proposes a new hybrid methodology for short-term prediction of energy efficiency. This new method consists of the stochastic frontier analysis-generalised autoregressive conditional heteroskedasticity (SFA-GARCH) model and the radial basis function neural (RBFN) model. The study finds that 30 regions (provinces and municipalities) in China have cluster-hetergeneity, and the different levels of industry structure, technology content and energy resources in the different regions lead to dissimilar energy saving quotas. In addition, through fair comparison between the traditional GARCH model and the new hybrid model, it is proved that the new hybrid model shows good performance and the results are reasonable. The energy efficiency indicators predicted by the hybrid model appear to be more reliable than the summation of the individual forecasts because it avoids the superposition of errors.

1. Introduction

1.1. Background

Due to the environmental pressure, energy efficiency and energy saving are becoming increasingly significant to government policy in China in response to a range of challenges including energy resource scarcity, shortage of energy supply and high energy price. Many of the high energy-consuming fixed asset investments have become enterprises in energy-intensive industries. Chinese energy consumption per unit output value is 2.3 times of the world average, and the energy efficiency is 10% less than the world average [1]. Continually increasing energy consumption and energy intensity have not only caused the alarm for Chinese energy security but also led to carbon emission

* Corresponding author.
E-mail address: wqtao@mail.xjtu.edu.cn (W.-Q. Tao).

http://dx.doi.org/10.1016/j.apenergy.2015.11.082
0306-2619/

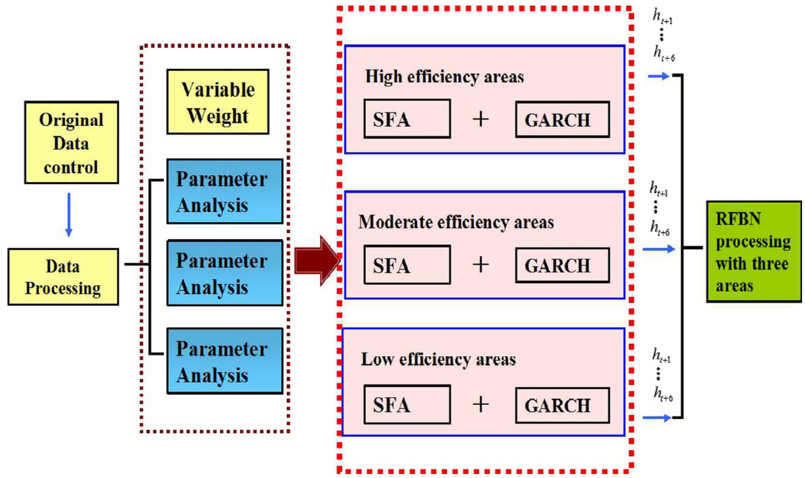


Fig. 1. Flowchart of the proposed methodology.

pressure in the post Kyoto-protocol era [2]. Therefore, to improve energy efficiency has become a rigorous matter in China. It is of great significance to investigate regional energy efficiency and analyse the influencing factors, because it can impact the strategy of scientific and technological development and the transformation of the mode of economic growth. Moreover, accurate forecasting of regional energy efficiency is vital for politic decision.

1.2. Previous studies of energy efficiency evaluation and prediction

There are mainly four types of models closely related to energy efficiency evaluation. First is the stochastic frontier analysis (SFA) model. It is mainly adopted to deal with the linear regression of energy efficiency. Boyd [3] and Boyd et al. [4] adopted the SFA model to investigate energy performance index. Then, the index data analysis (IDA) model is further extended based upon the SFA model. The SFA model can be used to investigate how the linear change of industrial structure impacts the total energy intensity. For example, Newell et al. [5] adopted it to analyse the relation between technical change and energy saving. Lin and Du [6] adopted the SFA model to estimate energy efficiency in 30 provinces, and they argued that the deficiency of this linear approach is ignoring the technology gaps across variable groups. Furthermore, the data envelopment analysis (DEA) model has been applied to overcome the disadvantage of the SFA model basically. Scholars employed the DEA model because they believed that energy efficiency should be put together with wide independent factors in order to evaluate outputs [1]. Wang et al. [7] applied the DEA model to do multi-directional efficiency analysis. Wang and Feng [8] adopted this model to evaluate the performance of environmental efficiency in China. They figured out that the tendency of environmental efficiency has begun to have an ascending path because of improved technologies. Recently, the DEA model has been implemented to examine the energy efficiency of coal-fired power units by Song et al. [9]. The obvious causal relationship could be obtained thanks to the linear characteristics of the DEA model. The DEA model can be easily applied to a multiple input–output black-box framework for estimating different indexes, especially for the decision making units of industries. However, the DEA model has some limitations. It does not overcome the summation of measurement errors, and evaluation of this method is easy to be influenced by its extreme value. Fourth, some studies employed the generalised autoregressive conditional heteroskedasticity (GARCH) models to estimate the volatility of assets, and they showed the short-run performance based upon in-sample forecasts [10]. The GARCH model mainly focuses on volatility analysis of time series data, without examining the underlying physical process. Ji and Guo [11] recently analysed the oil price volatility and its related issues via adopting the GARCH model for figuring out the reason of global financial crisis. Li et al. [12] further employed the GARCH model to examine the detailed causality relationship among different variables in thermal power plants, and then to figure out the most influential factors through non-manual intervention methodology.

On the other hand, the radial basis function neural (RBFN) model is a successful application for seasonal and time series forecast. It is able to structure frameworks for modeling a broad range of nonlinear issues. It can structure any type of relations with a high degree of accuracy. The most obvious advantage is that the RBFN model can universally approximate a large number of data, and no prior model needs to be built within the process. This power comes from the computation progress of the information from the data. Wedding and Cios [13] used the RBFN model and the Box–Jenkins model to generate certainty factors with normal output. Ginzburg and Horn [14] also employed a neural network to analyse time series prediction.

Recently, due to the limitation of each single model, hybrid methodology has become an important methodology to analyse comprehensive nonlinear systems. Conejo et al. [15] proposed a hybrid method based on the auto-regressive integrated moving average (ARIMA) models and the wavelet transform (WT) models to forecast day-ahead energy price of Spanish market. A hybrid model has been constructed by Li et al. [12] combining time series methods (WT model + RBFN model) and an adaptive evolutionary algorithm for day-ahead price. However, all the above methods ignore the effect of measurement errors and other statistical noises that may lead to the result errors.

There are many papers that focus on particular sectors of Chinese energy system but few studies pay attention on prediction of energy efficiency in medium-term (i.e., to 2020). The objective of this paper is to address this gap, by quantifying energy efficiency factors at a function level and by adopting the hybrid model. The new hybrid model will overcome the limitation of single model, and it combines the SFA-GARCH model and the RBFN model, for evaluating influencing factors and forecasting regional energy efficiency. The advantage of this method can be described as follows. When the amount of data is large, the RBFN model is presenting that the data match a training pattern with a high degree of accuracy and the results are reliable. When the amount of data is low, there are few training patterns and the forecast will be obtained in favour of the GARCH model. Therefore, the new hybrid model is more accurate than adopting RBFN model or GARCH model alone.

Using the proposed hybrid model, the investigations of this paper are twofold. The first one is to analyse the energy efficiency indicators (hereafter referred to as *EEI*) of the 30 regions (provinces and municipalities) in China from 2003 to 2014 by the new hybrid methodology. Unlike the usual regression approach, the hybrid method considers the heteroskedasticity, and results will not be easily biased. The second one is to forecast the *EEI* in the next 5 years. The results of prediction present three energy-efficiency cluster areas for which the government needs to provide helpful support with specific regional policies. To our knowledge, there has been rare study applying econometric models to forecast regional energy efficiency in China.

2. Proposed methodology

In this section, the proposed new hybrid method for short-term regional energy efficiency in China is described. This approach is a four-step procedure. First, the selected explanatory variables are tested whether data series are stationary or not. Stationary is premised of further calculation. In the second step, the SFA model is adopted to investigate the regional *EEI* using explanatory variables. In the third step, the GARCH model is used to evaluate the volatility among the *EEI* and influencing factors. Finally, the hybrid model is employed for estimating the short-term *EEI*. The results obtained with the proposed approach are compared with those from the traditional GARCH model. The prediction time-horison is 5-ahead years, from 2016 to 2020.

The proposed hybrid methodology is sketched in Fig. 1 and can be listed as follows:

(1) Structure a database including historical data of all variables that affect the short-term *EEI*s.
(2) Apply data processing for the explanatory variables to each sub-panel.
(3) Calculate each sub-panel by the SFA-GARCH model. This process is demonstrated in the red box of the flowchart.
(4) The RFBN model is adopted to forecast the nonlinear modeling of every panel.

2.1. Stationary test functions

In order to describe the unbiased estimators by the classical regression model, it is necessary to have a test of the time series in empirical studies for proving whether there is good evidence of a causal relationship. This section employs the augmented Dickey-Fuller (ADF) unit root test to examine whether data is stationary for further calculation or not. It includes extra lagged terms of the dependent variables for capturing the auto-correlation. The Lagrange multiplier (LM) test framed by Engle is more widely accepted in testing the presence of the autoregressive conditionally heteroskedasticity (ARCH) effect [16]. The LM test is employed for testing whether time series data have heteroscedasticity. The functions are as bellows.

2.1.1. Unite root test

$$\Delta y_t = \gamma y_{t-1} + \sum_{i-1}^{p} \beta_i \Delta y_{t-i} + \upsilon_t \tag{1}$$

$$\Delta y_t = \alpha_0 + \gamma y_{t-1} + \sum_{i-1}^{p} \beta_i \Delta y_{t-i} + \upsilon_t \tag{2}$$

$$\Delta y_t = \alpha_0 + \gamma y_{t-1} + a_2 t + \sum_{i-1}^{p} \beta_i \Delta y_{t-i} + \upsilon_t \tag{3}$$

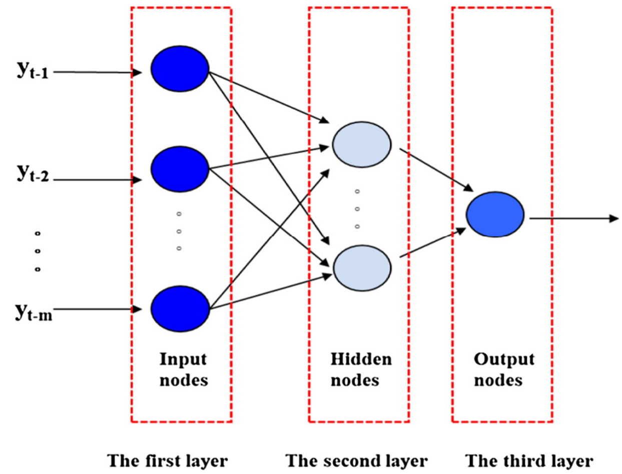


Fig. 2. A three-layer RBFN model.

where α_0 means constant, $a_2 t$ represents trend of equations. γ and β_i denote the short-run coefficients. The first equation demonstrates the ADF test with no constant and no trend in the series, whereas the second one represents the test as having a constant but no trend. The third equation shows that the test has both a constant and a trend.

2.1.2. LM test

This part sets lag number at 1 in order to test whether the ARCH effect exists. The LM test applies the null hypothesis that there is no serial correlation up to lag order *p*, where the lag is equal to 1 in this test.

$$\mu_t^2 = \gamma_0 + \gamma_1 \upsilon_{t-1}^2 + \ldots + \gamma_q \upsilon_{t-q}^2 + \mu_t \tag{4}$$

where μ_t represents squared distribution of Δy_t, which is obtained from Eq. (3). υ is data series, q presents the number of lags, and γ stands for the coefficient of correlation.

2.2. SFA model

2.2.1. Shephard energy distance function

As mentioned, the Shephard energy distance function is further employed in this section to define the *EEI* for investigating regional energy efficiency performance. The characteristic of this function is to include all variables in an aggregate production framework. The economy level, labour, technology level and investment level are all related variables. The function is as follows:

$$T = \{(PGDP, C, LP, IFA, EEI) : (PGDP, C, LP, IFA) \rightarrow EEI\} \tag{5}$$

where *PDGP* is per capita GDP, and it represents regional economic development; *C* is the proportion of coal in primary energy consumption, and it describes energy consumption characteristics; *LP* stands for labour productivity, and it represents technology level; *IFA* is fixed asset investment accounted for the proportion of GDP every year, and it represents investment level. In Eq. (5), economy level, labour, technology level and investment level are inputs. They are treated as independent variables. *EEI* is output and it is treated as a dependent factor. *T* includes all the independent and dependent vectors. In production theory, it is a closed and bounded set [10]. Moreover, the function normally assumes that:

$$(PGDP', C', LP', IFA', EEI') \in T \tag{6}$$

$$\text{if } (PGDP', C', LP', IFA') \geqslant (PGDP, C, LP, IFA) \text{ and } Y' \leqslant Y. \tag{7}$$

To investigate energy efficiency, the *EEI* is defined as the ratio of energy use to actual energy use, and it can be expressed as:

$$EEI = 1/D_E(PGDP, C, LP, IFA, EEI) \tag{8}$$

where the estimation of *EEI* equals 1, and it represents that the estimated figure of regions is located at the frontier of best practice.

2.2.2. SFA model

The Shephard energy distance function has to be analysed via a practical perspective. The Shephard energy distance function for region *i* can be represented as $D_E(PGDP_i, C_i, LP_i, IFA_i, EEI_i)$. The function is expressed in the form of the logarithms:

$$\ln D_E(PGDP_i, C_i, LP_i, IFA_i, EEI_i) = \alpha_0 + \beta_1 PGDP_i + \beta_2 C_i + \beta_3 LP_i + \beta_4 IFA_i + \varepsilon_t \tag{9}$$

where β_i means parameter *i* of region *i*, and it proves the impact of various factors from the perspective of empirical evidence. ε_t is a random statistical noise and approximation error.

2.3. GARCH model

The GARCH method is further adopted after the estimation of regional *EEI* by the SFA model. The objective is to investigate both long-term and short-term relationships among the volatility of *EEI*, *PDGP* and *C*, *LP* and *IFA*.

With the work of ARCH models pioneered by Engle, there are a large number of volatility models for precise evaluation [10]. GARCH (1,1) model is one of the GARCH (*p*,*q*) models, which is presented by Bollerslev in 1986 [16]. Where *p* is the order of the moving average ARCH term and *q* stands for the autoregressive GARCH term. The model equation is as bellows:

The traditional GARCH (1,1) model

$$R_{EEI} = c + \sum_{k \geqslant 1} a_k R_{t-k} + \sum_i b_i M_{i,t} + \varepsilon_t \tag{10}$$

$$\varepsilon_t \sim N(0, h_t) \tag{11}$$

$$h_t = \alpha_0 + \beta h_{t-1} + \alpha_1 \varepsilon_{t-1}^2 + \sum_i \gamma_i M_{i,t} \tag{12}$$

where R_{EEI} is the regional energy efficiency, and $M_{i,t}$ is the impact variable of the event *i*. The coefficients of Eq. (4) need to satisfy $a_k \geqslant 0, b_i \geqslant 0$, and $c > 0$, to ensure the conditional variance is positive. h_t explains the conditional variance of R_{EEI}, and it relies on $M_{i,t}$. The event of *i* stands for the time-varying changes of impact variable. α_0, β and α_1 have to be estimated in the variance to enable the past squared errors to determine the time-varying conditional variance.γ_i stands for the presence of the asymmetric effect.

In order to analyse specifically four influencing factors of regional *EEI*, the model is extended as bellows:

The extended GARCH (1,1) model

$$R_{EEI} = c + \sum_{k \geqslant 1} a_k R_{t-k} + b_{PGDP} M_{PGDP,t} + b_C M_{C,t} + b_{LP} M_{LP,t} + b_{IFA} M_{IFA,t} + \varepsilon_t \tag{13}$$

Table 2
Unit root test of four variables.

Variables	Automatic lag length	ADF statistic	5% Level of critical value	Inference
ln(*PGDP*)	0	−37.103	−3.382	Inf∼I(0)
ln(*C*)	0	−36.227	−3.416	Inf∼I(0)
ln(*LP*)	0	−42.491	−3.388	Inf∼I(0)
ln(*IFA*)	0	−30.743	−3.622	Inf∼I(0)

Table 3
LM test for ARCH of the variables.

Lags (p)	Prob. Chi-square	Obs * R-squared
1	0.0002	126.573

Note: The Prob. Chi-square is probability Chi-squared distribution, it is a continuous probability Distribution. The Obs * R-squared is observation R-squared distribution.

Table 4
Volatility relationship among four determinants of variables.

Variables	Coefficient		
	East region	Central region	West region
ln(*PGDP*)	−0.569***	−3.062***	0.147***
ln(*C*)	−0.706***	−0.101***	−0.093***
ln(*LP*)	0.482***	1.299***	−0.714***
ln(*IFA*)	−0.012***	−0.260***	−0.034***

Note: *** means figure is at significant level of 1%.

$$\varepsilon_t \rightarrow N(0, h_t) \tag{14}$$

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta h_{t-1} + \gamma_{PGDP} M_{PGDP,t} + \gamma_C M_{C,t} + \gamma_{LP} M_{LP,t} + \gamma_{IFA} M_{IFA,t} \tag{15}$$

where $M_{PGDP,t}$ stands for the per capita GDP each year, $M_{C,t}$ represents the proportion of coal in primary energy consumption, $M_{LP,t}$ stands for labour productivity, and M_{IFA} means fixed asset investment accounted for the proportion of GDP every year.

2.4. RBFN model

After the evaluation of regional *EEI*, the RBFN model is further adopted to forecast. It has been implemented to a variety of comprehensive problems because of its ability to solve non-linear relations between input and output variables. Neural forecasting network implicates that it is not only a vital candidate for time series data but also an accurate modeling tool for any types of causal relationship. There are many types of neural network models. One of the most popular methodologies is the multi-layer interceptions, which is used in this paper. It includes an input layer, hidden layers and an output layer as shown in Fig. 2. The input nodes are the first layer that collects data to each node of the second layers or hidden layers, and then the second layer represents a data cluster which is centred at a particular point. Finally the third layer has only one output node, and it is the sum of all the hidden nodes to lead to the decision value. Amjady [17] demonstrated that the weights

Table 1
Descriptive statistics of variables.

	Observation	Standard Deviation	Skewness	Kurtosis	Mean	Maximum	Minimum
PGDP	300	12748.708	0.681	3.075	16356.382	80376.600	2861.203
C	300	17.811	−0.064	3.573	64.272	96.500	10.100
LP	300	7.694	24.841	625.942	1.2712	195.000	0.000
IFA	300	2.656	3.152	11.159	2.950	15.393	0.090

Note: Full sample of variables are collected from January 01, 2003 to December 31, 2013.

Table 5
Energy efficiency values of provinces in China from 2003 to 2014.

	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Beijing	0.760	0.772	0.931	0.978	0.894	0.991	0.867	0.939	0.956	0.931	0.982	0.979
Tianjin	0.658	0.701	0.848	0.973	0.762	0.877	0.943	0.860	0.842	0.853	0.902	0.896
Hebei	0.780	0.711	0.719	0.858	0.755	0.840	0.833	0.865	0.902	0.851	0.899	0.875
Shanxi	0.720	0.628	0.687	0.729	0.696	0.908	0.825	0.896	0.910	0.903	0.842	0.886
Inner Mongolia	0.586	0.542	0.649	0.768	0.566	0.705	0.716	0.826	0.758	0.727	0.786	0.764
Liaoning	0.564	0.698	0.550	0.611	0.487	0.587	0.881	0.863	0.865	0.892	0.803	0.790
Jilin	0.487	0.576	0.687	0.758	0.560	0.852	0.756	0.861	0.859	0.855	0.831	0.901
Heilongjiang	0.620	0.633	0.528	0.662	0.648	0.846	0.846	0.743	0.856	0.901	0.931	0.899
Shanghai	0.856	0.879	0.901	0.937	0.872	1.000	0.987	0.964	0.915	0.958	0.987	0.973
Jiangsu	0.755	0.815	0.881	0.879	0.854	0.931	0.935	0.948	0.959	0.853	0.893	0.932
Zhejiang	0.785	0.800	0.915	0.944	0.879	0.956	0.897	0.902	0.936	0.895	0.849	0.896
Anhui	0.677	0.759	0.715	0.780	0.674	0.862	0.798	0.747	0.876	0.928	0.901	0.900
Fujian	0.862	0.877	0.914	0.963	0.928	0.993	0.786	0.864	0.880	0.855	0.869	0.893
Jiangxi	0.735	0.733	0.821	0.899	0.675	0.768	0.802	0.811	0.876	0.818	0.846	0.795
Shandong	0.540	0.569	0.612	0.685	0.624	0.711	0.780	0.726	0.697	0.708	0.799	0.843
Henan	0.550	0.536	0.535	0.542	0.673	0.589	0.672	0.617	0.550	0.543	0.653	0.663
Hubei	0.496	0.579	0.664	0.719	0.610	0.532	0.525	0.550	0.580	0.618	0.700	0.713
Hunan	0.557	0.568	0.609	0.657	0.660	0.583	0.652	0.579	0.610	0.669	0.785	0.721
Guangdong	0.875	0.910	0.932	0.960	0.922	0.900	0.897	0.848	0.921	0.900	0.933	0.967
Guangxi	0.657	0.663	0.699	0.719	0.688	0.670	0.731	0.684	0.782	0.699	0.778	0.729
Hainan	0.910	0.932	0.817	0.879	0.842	0.880	0.883	0.850	0.821	0.786	0.881	0.892
Chongqing	0.506	0.601	0.592	0.660	0.762	0.614	0.760	0.760	0.579	0.683	0.769	0.526
Sichuan	0.504	0.563	0.583	0.896	0.575	0.487	0.542	0.550	0.421	0.682	0.583	0.595
Guizhou	0.281	0.307	0.221	0.372	0.319	0.343	0.358	0.441	0.462	0.520	0.557	0.584
Yunnan	0.960	0.882	0.801	0.764	0.898	0.941	0.976	0.866	0.866	0.866	0.832	0.886
Shannxi	0.629	0.522	0.587	0.716	0.611	0.544	0.507	0.698	0.720	0.747	0.756	0.774
Gansu	0.479	0.357	0.435	0.479	0.438	0.342	0.474	0.479	0.593	0.593	0.532	0.550
Qinghai	0.363	0.354	0.460	0.371	0.381	0.245	0.261	0.220	0.288	0.245	0.437	0.445
Ningxia	0.366	0.347	0.428	0.383	0.495	0.440	0.587	0.391	0.391	0.358	0.427	0.551
Xinjiang	0.325	0.477	0.582	0.487	0.519	0.369	0.440	0.525	0.608	0.621	0.507	0.573

and biases of the neural network are appropriate to be supervised by the back-propagation in order to minimise the error between the actual outputs and the designed outputs. The input nodes are the lagged observations and the output nodes present the forecasting result. Hidden nodes are nonlinear transfer function to analyse information collected by input nodes.

The data set is usually divided into a training set and a test set. The training set is used to construct the neural network model and map the relationship between input patterns and output patterns. The training set is used to measure the predictive ability of the neural network. Once the neural network is properly trained, it can extrapolate patterns using a limited amount of input data.

According to the traditional method of RBFN model, 80% of the data are chosen to be training set. 20% of the data are chosen to be test set. It can be ensured that the prediction result still contains the tendency of historical data. The formula is as bellows:

$$y_t = \alpha_0 + \sum_{j=1}^{n} \alpha_j f\left(\sum_{i=1}^{m} \beta_{ij} y_{t-i} + \beta_{0j}\right) + \varepsilon_t \tag{16}$$

where n is the number of hidden nodes, and m is the number of input nodes. f means a vector of weights from the hidden nodes to output nodes: $\mathrm{f}(x) = \frac{1}{1+\exp(-x)} \cdot \{\alpha_j, j = 0, 1, \ldots n\}$, and $\{\beta_{ij,} i = 0, 1, \ldots, m; j = 1, 2, \ldots, n\}$ are vectors of weights from the input nodes to hidden nodes. α_0 and β_{ij} have values which are always equal to 1.

3. Explanatory variables selection

In the prediction methodology, modeling of energy efficiency is usually based on historical data, which has the relationship with other relevant factors, such as economic indicators, social indicators, environmental indicators and such. This paper sets influencing factors from following four perspectives:

(a) per capita GDP representing regional economic development level;
(b) the proportion of coal in primary energy consumption describing energy consumption characteristics;
(c) labour productivity demonstrating technology level;
(d) fixed asset investment accounting for the proportion of GDP every year, and it means investment level.

Due to the different sources, per capita GDP and the proportion of coal in primary energy consumption are composed of yearly data series, which are from China energy yearbooks. Labour productivity is chosen from China labour statistic yearbooks, and fixed asset investment accounted for the proportion of GDP is set from regional statistic yearbooks. Panel data area units include China's 30 provinces (autonomous regions and crown city, not including Tibet, Hong Kong, Macao and Taiwan area). The data set from January 01, 2003 to December 31, 2014, and the time period covers 11 years.

4. Empirical analysis

4.1. Unit root test of variables

The descriptive statistics of the volatility of the four independent variables for *PGDP*, *C*, *LP* and *IFA* are presented in Table 1. Before applying models, it is necessary to identify whether variables belong to a non-stationary series or not. The examination result is listed in Table 2.

The ADF test indicates that the series of all the explanatory variables are stationary at the 5% level of significance. The test demonstrates that the t-statistics value is lower than the critical value, and it means the series data do not have a unit root problem and they are stationary series. The data can be used to do further calculation.

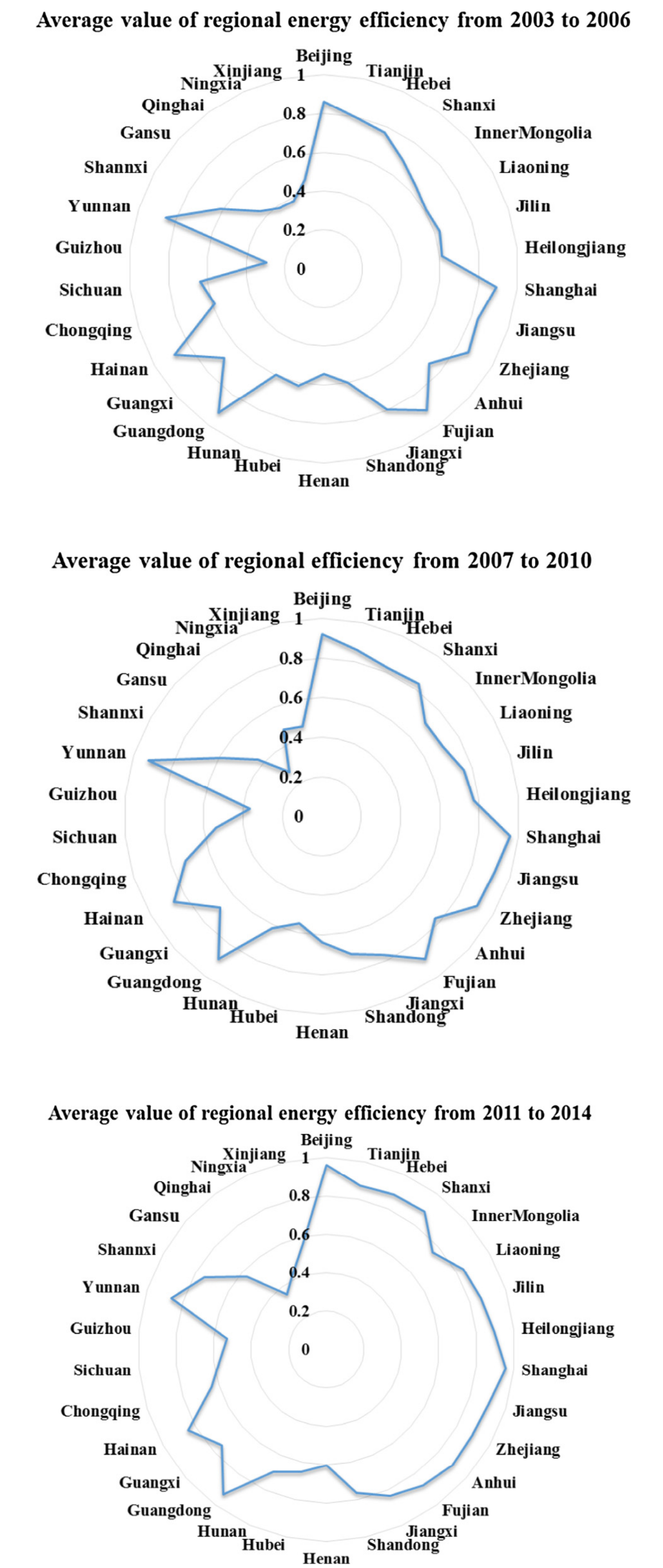


Fig. 3. Integrated average energy efficiency of 30 regions in China from 2003 to 2014.

4.2. LM test

The objective is to test how these variables play together to affect the output, and determine whether they can get the desired result. The estimated results of test are reported in Table 3. Assuming a null hypothesis that there is no serial correlation up to lag order p, where p is equal to 1 in this test. The LM test statistic value (126.573) should be compared with the critical value of Chi-Squared (1) value. The critical value of Chi-Squared (1) is selected as 3.82 from the statistical table. As 126.573 exceeds the critical value of 3.82, there is no doubt that the null hypothesis can be rejected. Therefore, a significant serial correlation exists between variables.

4.3. Fixed effect regressive analysis estimated by the GARCH model

The purpose of this section is to investigate the volatility coefficient among four variables. Table 4 summarises the results of the ARCH effect estimated by the GARCH (1,1) model. Due to the difference of resource allocation and economic development in every region, the following analysis is based upon eastern region, central region and western region, in order to examine the impact of various factors of regional differences.

The rapid economic growth (per capital GDP) has significant impact on energy efficiency of eastern region and central region of China. The changes in *PGDP* can have significant impacts on regional energy efficiency. However, the impacts of western region are not obvious because of the limited economic development. The volatility efficiency of *C* is highly influential with eastern cities due to the improving high clean coal technology in recent years. *LP* plays an important role in central regions. Therefore the energy efficiency in every province should be evaluated based upon the volatility of factors.

4.4. Application of hybrid model

How to use the new hybrid methodology for the implementation is another key issue. Due to the purpose of energy conservation, government should know the energy efficiency and the variation tendency. When government needs to make a decision on whether new policies should be approved, this hybrid model can be adopted to evaluate not only the national *EEI* but also the *EEI* in every province. The provincial energy efficiency evaluation will be implemented in this section.

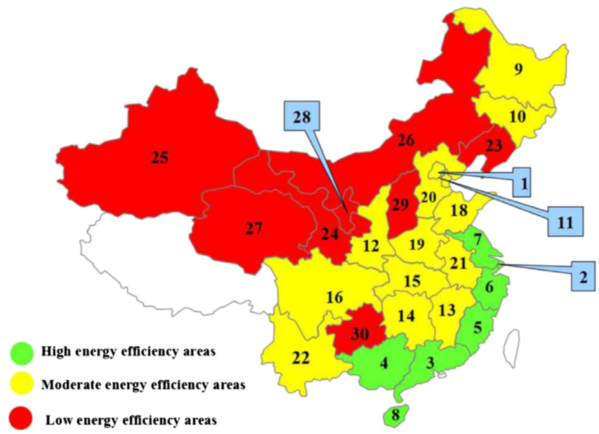
Implementation procedure is shown in Fig. 1, and it is listed as follows:

Step 1: A database is structured including provincial historical data of all variables that affect short-term *EEI*s. Then, apply data processing for the explanatory variables to each sub-panel. The unit root test and the LM test are employed for the unbiased estimators, and it is necessary to have a calculation of the time series in empirical studies for proving index of causal relationship.

Step 2: The Shephard energy distance function is shown in Eq. (2), and it is adopted and linearly simultaneous in energy efficiency. That is to say, the Shepard energy distance function will be increased by a certain proportion with the increasing of energy efficiency. It is also consistent with energy intensity. Then, calculate yearly energy efficiency of provinces by the SFA-GARCH model. This process is presented in the red box of flowchart.

Step 3: According to the perspectives of regional energy efficiency level, provinces can be divided into different *EEI* regions. Policy should be made based upon the regional cluster analysis. Furthermore, the RFBN model is adopted to forecast the nonlinear modeling of every panel.

The detailed estimation result of the hybrid methodology is presented as follows. Table 5 illustrates the empirical results of energy efficiency of China's 30 regions (provinces and municipalities). To obtain a more specific description and analysis of regional *EEI*, Fig. 3 presents the change of average value of energy efficiency in 30 regions from 2003 to 2014. It can be found out that: (i) Eastern regions, such as Beijing, Shanghai, Jiangsu and Zhejiang,



High energy efficiency area				Moderate energy efficiency area						Low energy efficiency area			
1	Beijing	6	Zhejiang	9	Hielongjiang	14	Hunan	19	Henan	23	Liaoning	27	Qinghai
2	Shanghai	7	Jiangsu	10	Jilin	15	Hubei	20	Hebei	24	Gansu	28	Ningxia
3	Guangdong	8	Hainan	11	Tianjin	16	Sichuan	21	Anhui	25	Xinjiang	29	Shanxi
4	Guangxi			12	Shannxi	17	Chongqing	22	Yunnan	26	Inner Mongolia	30	Guizhou
5	Fujian			13	Jiangxi	18	Shandong						

Fig. 4. The three cluster areas in China.

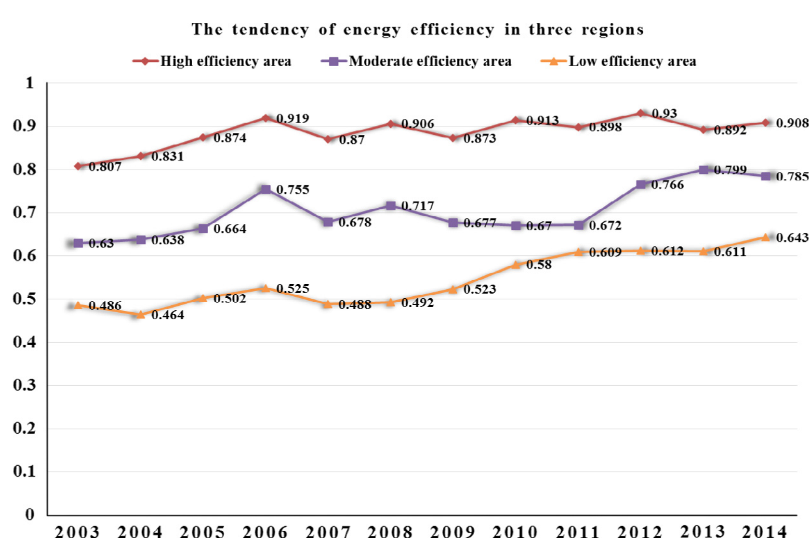


Fig. 5. The tendency of energy efficiency in three regions from 2003 to 2014.

have better performance than western and central regions. These regions have the most rapid economic growth in the past 3 decades, and their *PDGP* is around half of total GDP in China during the study period (from 2003 to 2014). Thanks to policies, a few high-consuming fixed asset industries and most service industries, as well as foreign technological investment are located in the eastern regions. (ii) The worst performance appears in Ningxia and Qinghai. Energy efficiency of Gansu is slightly increased in recent 5 years, because they have high resource reserves such as coal, oil, natural gas, and other minerals, and the historical development paid more attention on GDP than energy efficiency and environmental protection. These regions began to explore the transformation of energy consumption in recent years. (iii) The performance of other regions keeps stable during the study period. Because of

Table 6
5-year time-horison forecast.

	2016		2018		2020	
	GARCH	Hybrid	GARCH	Hybrid	GARCH	Hybrid
High efficiency areas	0.921	0.922	0.929	0.906	0.906	0.938
Moderate efficiency areas	0.722	0.705	0.816	0.816	0.816	0.903
Low efficiency areas	0.671	0.690	0.734	0.783	0.820	0.886

the promotion of "Twelfth Five year plan" of China, more western regions have obviously developed clean energy, such as wind energy and solar energy, in order to lessen the reliance on traditional fossil fuel energy resources. Inner Mongolia can be treated as a typical province in Fig. 3. The average value of its efficiency is increased because of lower proportion of *C* and the higher *IFA*. Moreover, it has inexhaustible wind power source and solar source. The government took several investment policies to develop technology in the past few years. Inner Mongolia has the most installed wind power capacity in China, which is 11.39 GW, and it is the most significant electrical supply base now.

Then the system cluster analysis is adopted, according to the perspectives of regional energy efficiency level. These provinces are divided into three areas: high energy efficiency area, moderate energy efficiency area and low energy efficiency area, as shown in Fig. 4. It not only demonstrates the geographic location of the regions, but also illustrates the efficiency level with different colors.

Based on the average values of energy efficiency in the different three areas, Fig. 5 shows three patterns of energy efficiency by areas (high efficiency, moderate efficiency and low efficiency). During the study period, changes of energy efficiency have obvious fluctuation. The overall trends of China present that energy efficiency was improved prior to 2006, then decreased after that, and increased gradually after 2009. Because energy saving target was promoted in the "Twelfth Five-year plan" of China, the *EEI* of low efficiency area has a steep upward trend while the change of *EEI* of high efficiency area is stable. The moderate efficiency area has the worst performance during these years.

Table 6 shows the proposed efficiency value for 2016, 2018 and 2020. The traditional GARCH model is further adopted to make a fair comparison for empirical estimation. The calculation results of prediction show that the *EEI* of high energy efficiency area will be stable in 5-ahead years, while those of moderate energy efficiency area and low energy efficiency area maintain a rising trend. The gap between high efficiency area and moderate efficiency area is smaller with the increase of prediction time-horison. The forecast results obtained by the hybrid model are quite close to the value proposed by the traditional GARCH model. It proves that the results estimated with the hybrid model are reasonable. The hybrid prediction methodology of *EEI* in three areas appears to be more reliable than the linear overall forecasts. When making prediction of every series data and adding them up to get a whole result, the residual error of each series will be also added up which leads to a bigger error. Moreover, following the same principle as mentioned above, the time horison of 5-year is appropriate to prediction management, the residual error may lead to biased results.

In a summary, the energy efficiency of China has been made a stable improvement during our study period. In 2006, the Chinese government announced a national objective that the energy intensity of China should be reduced by 20% until 2010, compared with the energy intensity in 2005. Moreover, the achievement is mainly contributed by a series of energy saving policies, regulations and programs since the published "Twelfth Five-year plan", and they primarily focuse on alleviation of energy shortage issues. This study proved that the series implementation of rigorous energy saving policies contributed to the better performance of energy efficiency in China. The tendency of prediction of *EEI* in every province will be held over time.

5. Conclusion

Using a combined SFA-GARCH model, and taking into account four explanatory inputs, this study evaluates the *EEI* of 30 regions (provinces and municipalities) in China from 2003 to 2014. Furthermore, it adopts the RBFN model for the short-term regional energy efficiency prediction based on the three cluster areas of China from 2016 to 2020.

Three conclusions can be obtained from the study. First, the paper presents a new four-stage hybrid methodology which combines the SFA-GARCH model and the RBFN model. It is straightforward to evaluate the *EEI* of different regions and to forecast the short-term performance without manual intervention. After the fair comparison, the predicting result is proved to be reasonable. Second, the new hybrid model of *EEI* prediction appears to be more appropriate than the summation of the individual forecasts because individual errors of series will be added up which results in greater errors. Third, there is significant cluster-heterogeneity among different regions, especially considering the four variables. High energy efficiency area plays a dominant role in comparing the *EEI* levels under frontier analysis, and it does behave better than the moderate efficiency area and low efficiency area of China from 2003 to 2009. The *EEI* gap between these three areas is narrowed during the "Twelfth Five-year plan".

Through the proposed hybrid model, this study has the following policy recommendation. First, government should apply different energy policies to associate different regions. The low efficiency area should be allocated with more energy saving quotas, and the energy policies should bias to them. For those regions in high energy efficiency area, such as Beijing, Shanghai and Guangdong, the leading industry has been changed from the secondary industry to the third industry. It is necessary for these regions to focus on controlling the scale of energy supply from other regions. For those regions in moderate energy efficiency area, such as Shannxi, Tianjin and Hebei, the proportion of coal in primary energy consumption is above the average level. It is significant for them to focus on technological improvement of production process. The application of energy saving technologies which are helpful for environment should be promoted and assisted by financial support. For those regions in low energy efficiency area, such as Liaoning, Ningxia and Inner Mongolia, energy resources outputs made huge financial support for them in past years. It is vital to control the allocation of energy resources, and extensive expansion of resource usage should be no longer permitted. Moreover, local government should create incentive for improving clean energy technology and reducing waste of energy. Second, the target of emission reduction is severe for the Chinese government in the next 5 years because the coal-burned power plants are still the major provider of electricity in China. It is suggested that the requirement level of carbon emission should be higher.

Acknowledgment

This work is supported by China National Energy Conservation Centre and the National Key Basic Research Program of China (973 Program) (2013CB228304).

References

[1] Hu JL, Wang SC. Total-factor energy efficiency of regions in China. Energy Policy 2006;34:3206–17.

[2] Andrews P. China's ongoing energy efficiency drive: origins, progress and prospects. Energy Policy 2009;37(4):1331–44.

[3] Boyd GA. Estimating plant level energy efficiency with a stochastic frontier. Energy 2008;29(2):23–43.

[4] Boyd G, Dutrow E, Tunnessen W. The evolution of the energy star, energy performance indicator for benchmarking industrial plant manufacturing energy use. J Clean Prod 2008;16:709–15.

[5] Newell RG, Jaffe AB, Stavins RN. The induced innovation hypothesis and energy-saving technological change. Quart J Econ 1999;114(3):941–75.

[6] Lin BQ, Du KR. Technology gap and China's regional energy efficiency: a parametric metafrontier approach. Energy Econom 2013;40:529–36.

[7] Wang K, Yu SW, Li MJ, Wei YM. Multi-directional efficiency analysis-based regional industrial environmental performance evaluation of China. Nat Hazards 2014.

[8] Wang ZH, Feng C. A performance evaluation of the energy, environmental, and economic efficiency and productivity in China: an application of global data envelopment analysis. Appl Energy 2015;147:617–26.

[9] Song CX, Li MJ, Zhang F, Tao WQ. A data envelopment analysis for energy efficiency of coal-fired power units in China. Energy Convers Manage 2015;102:121–30.

[10] Engle RF. Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. Econometrica 1982;50:987–1007.

[11] Ji Q, Guo JF. Oil price volatility and oil-related events: an Internet concern study perspective. Appl Energy 2015;137:256–64.

[12] Li MJ, Song CX, He YL, Tao WQ. Examining the volatility degree of mutual influence among different variables: thermal power plants. Energy Procedia 2014;61:1380–8.

[13] Wedding DK, Cios KJ. Time series forecasting by combining RBF networks, certainty factors, and the Box-Jenkins model. Neurocomputing 1996;10:149–68.

[14] Ginzburg I, Horn D. Combined neural networks for time series analysis. Adv Neural Inf Process Syst 1996;6:224–31.

[15] Conejo AJ, Plazas MA, Espínola R, Molina AB. Day-ahead electricity price forecasting using the wavelet transform and ARIMA models. IEEE Trans Power Syst 2005;20:1035–42.

[16] Bollerslev T. Generalized autoregressive conditional heteroskedasticity. J Econom 1986;31:307–27.

[17] Amjady N, Keynia F. Day ahead price forecasting of electricity markets by a mixed data model and hybrid forecast method. Elect Power Energy Syst 2008;30:533–46.