

# Regional Power Generation Prediction Model and Post-Evaluation Based on Coal Production and Economy Status

N.Q. Xiao, Q.M. Yang

State Key Laboratory of Industrial Control Technology, Department of Control Science and Engineering  
Zhejiang University  
Hangzhou, Zhejiang

**Abstract**--With the deepening of the reform of the electricity market and rapid development of power industry, advancement of economy and power generation is becoming more and more closely related to each other. Balance between supply and demand is always one of the core issues of regional electricity market research. This paper proposes an electricity generation prediction model in power plant combining statistical methods and neural network in order to improve predictive performance. On the basis of power generation influential factors analysis, first we adopt several statistical methods as ADF test, Granger Causality Model, Error Correction Model and Co-integration Test to investigate the relationship between several variables as GDP, industrial structure, technological progress, even the temperature and so on, aiming at the discovery of possible causes that lead to fluctuations of power generation. Then an ingenious predictive model using artificial intelligence method neural network is created in numerical prediction of future amount of electricity generation. At last, efficiency results of both operation scale and technology are given according to data envelopment Analysis (DEA) as post-evaluation. All the data is collected from website of National Bureau of Statistics, with time series in month up to 15 years (1998-2013). This study is of great significance to estimate the status of future electricity generation, and is of huge support to facilitate future intensification and optimization of resource allocation, and enhances operational efficiency.

**Keywords**--statistical methods; neural network; DEA; post-evaluation

## I. INTRODUCTION

Since 2006, coal production performed a U-shaped growth in China as economic growth slowed down affected by the financial crisis, and weaken electricity production with no doubt indicated reduced production for coal over time. Due to the economy recovery after 2008, direct and indirect consumption of coal resources is increasing promisingly, resulting in rapid growth of industrial and residential electricity consumption [1].

With the deepening of the reform of the electricity market and rapid development of power industry, advancement of economy and power generation is becoming more and more closely related to each other. Balance between supply and demand is always one of the core issues of regional electricity market research. In recent years, electric power provision in

China increases quickly, but not enough to catch up with the growth rate of regional production for electricity. The contradiction between supply and demand become conspicuous and may lead to severely restriction and impediment of economic development [1, 2, 3].

Thanks to the concept of integrated forecast, this paper manages to propose an electricity generation prediction model in power plant combining statistical methods and neural network in order to improve predictive performance.

On the basis of power generation influential factors analysis, first we adopt several statistical methods as ADF test, Granger Causality Model, Error Correction Model and Co-integration Test to investigate the relationship between several variables as GDP, industrial structure, technological progress, even the temperature and so on, aiming at the discovery of possible causes that lead to fluctuations of power generation. Then an ingenious predictive model using artificial intelligence method neural network is created in numerical prediction of future amount of electricity generation. At last, efficiency results of both operation scale and technology are given according to data envelopment Analysis (DEA) as post-evaluation. This study is of great significance to estimate the status of future electricity generation, and is of huge support to facilitate future intensification and optimization of resource allocation, and enhances operational efficiency.

The rest of this paper is organized as follow. Section 2 shows how to extract influential indicators of electricity production from various areas and predictive model shaped as multiple temporal units neural network is given in section 3. Typical simulation results of post-evaluation of predictive performance and persuasive DEA efficiency results of both operation scale and technology are demonstrated in section 4 and key conclusions are drawn in section 5.

## II. INFLUENTIAL FACTORS OF ELECTRICITY PRODUCTION

Power plant is a huge, complex system, and fluctuations in power generation are caused not only by their own inner factors, but also by many external factors like upstream industries, economics and environmental variables. How to extract and refine influential indicators of electricity production from various areas become very necessary to reduce complexity and ensure accuracy of the prediction model.

### A. Leading, Coincident and Lagging Indicators

If we choose power generation amount as benchmark, there will be 3 types of relationship between every single one of the rest variables and benchmark, called leading, coincident or lagging indicators [4].

In our circumstances, leading indicators are indicators that usually change before power generation amount as a whole changes. They are therefore essentially useful as short-term predictors of the economy.

Coincident indicators change at approximately the same time as the benchmark, thereby providing more information about the current status of the industry.

Lagging indicators are indicators that usually change after power generation amount as a whole does. Typically the lag is one or a few quarters of a year.

Fluctuations in power plant are usually monitored using these relationships to expand the analysis, which is of great importance when it comes to analysis on influential factors of power generation.

### B. Influential Indicator Filter

The scheme and procedures telling how to utilize statistical methods to filter variables from different areas to a few indeed influential indicators to power generation are shown in Figure 1, in which quite a few statistic methods are adopted.

After benchmark and alternative indicators are chosen, ADF test is adopted to determine whether the time series of one indicator is stationary or integrated, then all the indicators can be divided into corresponding  $n$ th ( $n=1,2,\dots$ ) integrated classifications. Variables only same order integration with benchmark are kept, following by co-integration test, Granger causality test and vector auto-regression model to determine which one is leading, coincidence or lagging indicator.

Hypothesis testing in regression analysis can inspect how much variable  $x$  is able to interpret variable  $y$ , and to what extent the enhancement of interpretation will be by introducing hysteresis values of variable  $x$ . That is to say, if the variable  $x$  is of help to predict variable  $y$  and correlation coefficient between two variables is statistically significant, the existence of Granger causality from variable  $x$  to variable  $y$  is sound.

At last, Vector Auto-Regression model (VAR) is applied to determine the period one specific variable is leading or lagging compared to benchmark [5].

In this paper, as benchmark chosen as power generation amount, alternative indicators taken into consideration contain multiple variables from different areas. By carrying out the proposed filter, there remain six indicators selected for electricity generation prediction, such as GDP, raw coal production, taxes turned over to state, industrial output value, above-scale industrial added value, etc.

Benchmark electricity production: has always been the most important indicator to measure the economic strength of electricity corporations;

Raw coal production: product amount and price of coal directly influence the beneficial results of power plant.

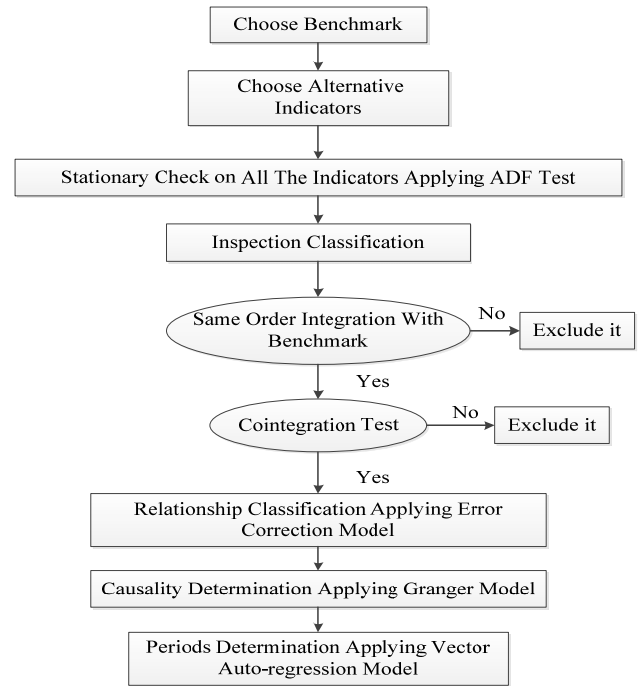


FIGURE 1. SCHEME OF INFLUENTIAL INDICATOR FILTER.

### III. ELECTRICITY PRODUCTION PREDICTION

Based on indicators selected by the above-mentioned filter, we propose a short-term power generation forecasting model applying artificial intelligent algorithm named supervised learning of neural network. 2 temporal models are developed sequentially for upgrading predicative performance. In the first part, MLP is applied to construct a reference model, followed by the second NN model we proposed utilizing the concept of divide-and-conquer.

#### A. Artificial Neural Network (Ann)

Artificial neural networks (ANNs) are computational models inspired by an animal's central nervous systems which are capable of machine learning as well as pattern recognition. An ANN is generally presented as a system of inter-connected "neurons" which can compute values from inputs, typically defined by three types of parameters:

- Interconnection pattern between the different layers of neurons
- Learning process for updating the weights of the interconnections
- Activation function that converts a neuron's weighted input to its output

What has attracted the most interest in neural networks is the possibility of learning. In supervised learning, given a set of example pairs  $(x, y)$  with  $x \in \mathbf{X}$  and  $y \in \mathbf{Y}$ , the aim is to find a function  $f: \mathbf{X} \rightarrow \mathbf{Y}$  in the allowed class of functions that matches the examples. In other words, we wish to infer the mapping implied by the data; the cost function is related to the mismatch between our mapping and the data and it implicitly

contains prior knowledge about the problem domain. A commonly used cost is the mean-squared error, which tries to minimize the average squared error between the network's output  $f(x)$ , and the target value  $y$  over all the example pairs.

In this sense, NNs maybe seen as multivariate, nonlinear and nonparametric methods, and they should be expected to model complex nonlinear relationships much better than the traditional linear models that still form the core of the forecaster's methodology.

### B. Prediction Model

A basic MLP is applied as a benchmark, which is a fully forward connected network usually with three layers named input layer, hidden layer and output layer, where hidden layer mixes information from input layer and generates new features for network learning. Each broken line between different layers has one weight on it and represents a parameter.

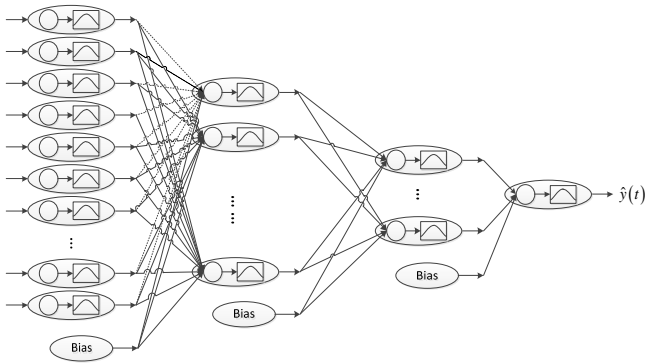


FIGURE II. PROPOSED PREDICTION MODEL BASED ON ANN.

One of specific functions of MLP is that hidden layer mixes information from input layer to generate novel features. Although the concept is straightforward to use all available information on hand to generate novel features (functioned by hidden neurons), the mix also blends all information together. This situation may make the learning ineffective, especially when input information is from various sources [6].

Proposed power generation prediction model based on ANN is shown in Figure 2, where concept of divide-and-conquer is utilized to break a complicated task into several pieces. Then each piece of the task is respectively trained. It is apparent that the input features, such as GDP, raw coal production, taxes turned over to state, industrial output value, above-scale industrial added value, are from several distinctive sources.

The proposed model handles these distinctive groups of information respectively, which is the process of division. Then each distinctive group of information is trained, which is the step of conquest.

It should be noted that the third layer is designed to describe different groups, named feature layer. One vital characteristic in feature layer is that weights are partially connected instead of fully connected between input and feature layers. After that, the training results are integrated together to obtain final prediction.

## IV. EMPIRICAL PREDICTION RESULTS AND EVALUATION

Samples used for training and testing are described as follows.

First, all the data is collected from website of National Bureau of Statistics, with time series in month up to 15 years (1998~2013). Second, in our study, training samples for model construction must be ample to show the features of the input set, so data of 12 years seems reasonable to realize the pattern of electricity consumption. Third, since electricity production is dynamic, the concept of moving window data learning method is adopted to keep including recent data and excluding the same amount of early data [7].

### A. Prediction Results

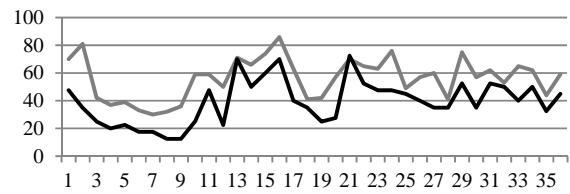


FIGURE III. COMPARISON BETWEEN MLP AND PROPOSED MODEL IN TERMS OF MSE MLP (GREY) AND PROPOSED MODEL (BLACK).

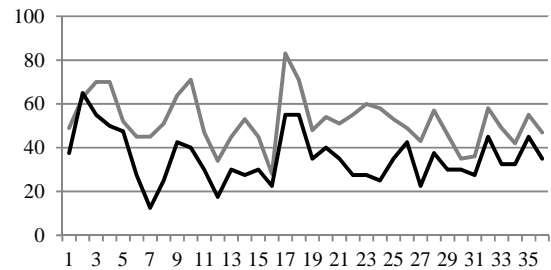


FIGURE IV. COMPARISON BETWEEN MLP AND PROPOSED MODEL IN TERMS OF MAPE.

MLP (grey) and Proposed Model (black).

Mean Square Error (MSE) and Mean Absolute Percentage Error (MAPE) are adopted for performance evaluation. Equation 1 and 2 show the formulas.

$$MSE = \frac{1}{n} \sum_{t=1}^n (y(t) - \hat{y}(t))^2, \quad MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|y(t) - \hat{y}(t)|}{y(t)}$$

The proposed network structures are applied to thirty-six testing experiments. Figs. 5 and 6 show the comparison between MLP and proposed model in terms of MSE and MAPE respectively. It can be seen that proposed model outperforms MLP in most testing experiments, no matter from MSE or MAPE.

In addition, it is more importantly to see the actual improvements. Averagely speaking, proposed model obtains 7.1% improvement of MSE and 4.3% improvement of MAPE in comparison with MLP. Although proposed model adopts four

extra parameters in comparison with MLP, it is worth because predictive performance is improved significantly.

**B. Post-Evaluation**

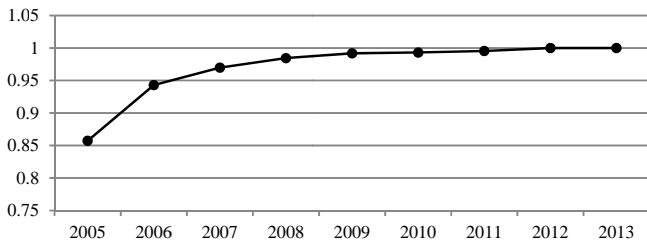


FIGURE V. DEA EFFICIENCY VALUES OF POWER GENERATION EACH YEAR.

	2005	2006	2007	2008	2009	2010	2011	2012	2013
Power Consumption Rate%	1.2725	0.4863	0.2076	0.1341	0.0916	0.0821	0.0583	0	0.6072
Coal Consumption(g/kwh)	57.1922	22.6025	10.2974	7.306	5.3468	4.7946	3.9781	0	0.6939
Fuel/t	967.8672	393.569	70.5376	20.2189	12.389	14.0534	10.5516	0	0
Boiler efficiency%	0.3437	0.3054	0.2441	0.1416	0.4504	0.5126	0.184	0	0
Power generation/10 <sup>4</sup> kwh	0	0	0	0	0	0	0	0	0
Industrial added value /10 <sup>4</sup> RMB	0.2108	0.2199	0.5358	0	0	0	0	0	0

FIGURE VI. INVALID SCALE OF INDICATORS IN POWER GENERATION OF A POWER PLANT.

Data envelopment analysis (DEA) efficiency results of both operation scale and technology under power generation prediction model. Establishment of the index system is the prerequisite and basis for prediction or post-evaluation studies [8].

This paper carried out in-depth research and consulting, and determined the following 5 indicators to be used in practical application: power consumption rate, coal consumption, fuel, boiler efficiency, power generation, and industrial added value.

Due to the limited length of the article, detailed procedures of DEA method is omitted here [8]. DEA efficiency values of power generation in both operation scale and technology are shown in Figure 5. Results shows that: 1) Overall efficiency of the generator set is gradually getting to a relatively high level; 2) However, the trend of growth is slowing down, that is to say, investment of generator sets appears to be redundant. Invalid scales of each indicator from 2005 to 2013 are shown in Figure 6, which actually quantifies the degrees for each indicator to improve power generation under this inadequate situation.

**V. CONCLUSIONS**

This paper proposes an electricity generation prediction model in power plant based on coal production by combining statistical methods and neural network. Core idea of proposed model is providing information at suitable places instead of mixing all input information together at the learning phase. The promising results show that elaborate designs of neural networks structure do improve predictive performance. This

study is of great significance to estimate the status of future electricity generation as a result of development of digital mines.

**ACKNOWLEDGEMENT**

This work is supported by National High Technology Research and Development Program of China (863) under Grant 2012AA062201.

**REFERENCES**

- [1] Crompton P, Wu Y. Energy consumption in China: past trends and future directions. *Energy Economics*, 2005, 27(1): 195-208.
- [2] Ozturk I. A literature survey on energy-growth nexus. *Energy policy*, 2010, 38(1): 340-349.
- [3] Payne J E. A survey of the electricity consumption-growth literature. *Applied energy*, 2010, 87(3): 723-731.
- [4] Granger C W J. Causality, co-integration, and control. *Journal of Economic Dynamics and Control*, 1988, 12(2): 551-559.
- [5] Goebel R, Roebroek A, Kim D S, et al. Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modeling and Granger causality mapping. *Magnetic resonance imaging*, 2003, 21(10): 1251-1261.
- [6] Tsai T H, Lee C K, Wei C H. Neural network based temporal feature models for short-term railway passenger demand forecasting. *Expert Systems with Applications*, 2009, 36(2): 3728-3736.
- [7] Morantz, BHWhalen, T. & Zhang, G. P. A weighted window approach toneural network time series forecasting. In G. P. Zhang (Ed.), *Neural networks inbusiness forecasting*. USA: IRM Press, 2004.
- [8] Cook, Wade D., and Larry M.Seiford. Data envelopment analysis (DEA)—Thirty years on. *European Journal of Operational Research*, 2009, 192, no. 1: 1-17.