Energy Conversion and Management 87 (2014) 439–454

Contents lists available at ScienceDirect

Energy Conversion and Management

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

Density prediction and dimensionality reduction of mid-term electricity demand in China: A new semiparametric-based additive model

Zhen Shao [a,b,*], Shan-Lin Yang [a,b], Fei Gao [a]

[a] School of Management, Hefei University of Technology, Hefei, Anhui 230009, PR China
[b] Key Laboratory of Process Optimization and Intelligent Decision-Making, Ministry of Education, Hefei, Anhui 230009, PR China

ARTICLE INFO

Article history:
Received 21 October 2013
Accepted 14 July 2014
Available online 2 August 2014

Keywords:
Mid-term demand forecast
Density prediction
Dimensionality reduction
Semiparametric additive model
Piecewise smooth
Variable simulation

ABSTRACT

Accurate mid-term electricity demand forecasting is critical for efficient electric planning, budgeting and operating decisions. Mid-term electricity demand forecasting is notoriously complicated, since the demand is subject to a range of external drivers, such as climate change, economic development, which will exhibit monthly, seasonal, and annual complex variations. Conventional models are based on the assumption that original data is stable and normally distributed, which is generally insignificant in explaining actual demand pattern. This paper proposes a new semiparametric additive model that, in addition to considering the uncertainty of the data distribution, includes practical discussions covering the applications of the external variables. To effectively detach the multi-dimensional volatility of mid-term demand, a novel piecewise smooth method which allows reduction of the data dimensionality is developed. Besides, a semi-parametric procedure that makes use of bootstrap algorithm for density forecast and model estimation is presented. Two typical cases in China are presented to verify the effectiveness of the proposed methodology. The results suggest that both meteorological and economic variables play a critical role in mid-term electricity consumption prediction in China, while the extracted economic factor is adequate to reveal the potentially complex relationship between electricity consumption and economic fluctuation. Overall, the proposed model can be easily applied to mid-term demand forecasting, and produce accurate and stable forecasts.

1. Introduction

Along with the fast advance in intelligent control technology, sensing measurement technology, and computing technology, electric power system has become an indispensable component of modern society in which its operation quality has a significant impact on the economic development and human civilization. Over the last decades, the long-standing central problem of energy shortage, fundamental to the economic development, is now still far from solved. To coordinate an effective response to this emergency, governments have started to expand the use of renewable energy such as wind energy, solar energy and hydropower. Furthermore, the increasing pressure on energy demand forces governments to capitalize rigorously on enhancing the ratio (efficiency) of energy utilization.

Nowadays, China has become the world's largest consumer and producer of energy, which relies mostly on its own strength to develop energy, and its energy self-sufficiency ratio has reached 90% mark. The report states that by 2015 China's non-fossil energy will add to 11.4% in the national total primary energy consumption, and the CO_2 emission per unit of GDP will decrease by about 3.9% per year in the next few years [1]. These developments, in addition to promoting the continued growth of China's economic, were significant factors in the progress of society from one characterized by relatively environmental polluted community into a resource conserving and sustainable nation. It is worth noting that the traditional energy resources, such as coal and hydropower, are so essential to the survival of China' economic will set to remain the main power source in China for a long time. Furthermore, the energy resources which are scattered widely across China are not evenly distributed. According to the report, coal resources are mainly occurs in north and northwest of China, hydraulic resources are mainly distributed in southwest of China [2,3]. However, most of the energy shortage areas are located in the costal and southeast areas where the economy is well developed. To handle this (reverse distribution of supply and demand) problem caused by

* Corresponding author at: School of Management, Hefei University of Technology, Hefei, Anhui 230009, PR China. Tel.: +86 055162904935; fax: +86 05512905263.
E-mail address: shaozhen@mail.hfut.edu.cn (Z. Shao).

http://dx.doi.org/10.1016/j.enconman.2014.07.038
0196-8904/

the difference of location between the producers and the consumers, China' electric power must be transport over long distances from the west to the southeast. Electric power transportation and dispatch present new challenges for planners, who now are forced to address them. Guangzhou and Suzhou, which are two of the mega-cities in southern and eastern China, for one thing their structural power shortages remain unsolved, for another thing their electricity consumption remained at a high growth rate [4]. A precisely estimation of electric power consumption, is of fundamental importance to the local economic development. Considering the case of electric demand forecasting, a power supply plan maker must decide how to allocate power resources among the brisk demand and the finite resource.

According to the difference of forecasting horizon, electric load forecasting could be mainly classified into three types: short-term forecast [5–7], mid-term forecast [8–10], long-term forecast [11–13]. Relative to mid-term and long-term forecast, literature more focused on short-term forecast. In this paper, we devote attention to the prediction of mid-term load which is the foundation and premise of electric power system dispatch, planning and management [10,14]. Compared with the short and long-term forecast, mid-term forecast has much more complex characteristics in its configuration spaces. The characteristics of short-term and long-term fluctuation together constitute the spatial mesh relations of mid-term sequence. Therefore, it's necessary to take into account the cyclical trends and longitudinal trends when we are forecasting. However, unlike short-term forecast, for which related theories have been well established, studying on the volatility, especially on the multi-dimensional volatility of mid-term forecast is still relatively limited. Moreover, traditional theory of forecasting is far from satisfactory if it is applied to this much complicated reality cases.

More recently, modeling dependence with statistics theory [15–17], Grey model [18,19] and artificial neural network [20,21] has attracted increasing attention. However, the use of traditional statistics forecasting models that assume linear relationships between observations and electric consumption might be problematic and lead to poor forecast accuracy if the sequence actually exhibit nonlinear patterns [22]. Grey model, which requires only a finite number of data and can obtain accurate prediction results, only depicts a monotonously increasing or decreasing process with time as exponential law [23]. Artificial neural network models are black-box models and require a large scale of data for model establishing [24]. Nevertheless, mid-term electricity consumption is characterized by multivariable, nonlinear, wide operation range, which is too complex to be described by a traditional model. Meteorological and other variables that exhibit significant time variation are hard to predict. Meanwhile, as the number of potential predictors could be large, it is still difficult to find out the fluctuations features of electric consumption which have long-term validity and can be implemented in different situation.

Motivated by the above, this work investigates a new framework for semiparametric additive statistical model which extends the principles described in [25–27]. By developing a novel framework for combining semiparametric model, ARIMA model, and piecewise smooth method, this paper makes three key contributions: (1) A new semi-parametric model composed of stationary time series smoothing is proposed for mid-term load forecasting, which does not impose any constraints on the assumption of the model distribution and enables the use of external variables for multi-step forecasting. (2) A novel semiparametric additive model based on piecewise smooth method is established. This model is designed to eliminate the multidimensional characteristic by a feature recognition method which could provide accurate prediction for multidimensional and cyclical series. (3) The bootstrap estimation method for proposed model is developed to provide a robust estimation of parameter and density forecasts of the future demand. This method could improve the small sample (characteristics of mid-long term load forecasting) estimation performance without reducing computational efficiency.

The rest of the paper is organized as follows. Section 2 reviews current literature on mid-term electricity demand forecasting, including its scope and methods. Section 3 proposes a detailed analysis of semiparametric additive model and describes the implementation of the proposed forecasting approach. Section 4 presents the forecasting results and discussions, and Section 5 concludes the paper.

2. Literature review

The literature on electric power consumption prediction is vast, and a very complete theoretical support can be found and referenced in [28,29]. The increasing gap between electric demand and supply has attracted more and more attention to the quality of electric demand forecasting in China [2]. Mid-long term load forecasting is essential for electric system planning, scheduling and controlling. Given the objective of this paper, we focus on the methodologies of mid-term load forecasting.

Widely adopted mid-long term electric consumption prediction techniques include regression techniques, time series analysis, machine learning techniques, etc. In general, the major research perspective can be distinguished into two types: one is weather insensitive approach that only employs historical load data to forecast future electricity load, the other is based on the cause effect relationships between electric demand and relevant independent variables (weather, economic, etc.). As a traditional theory, time series model which uses historical data to forecast future consumption has been extensively applied to electric prediction for its relatively efficiency and availability. Over past decades, different powerful time series models have been suggested to capture the historical fluctuation, including Exponential smoothing, Autoregressive (AR) model, Autoregressive moving average (ARMA) model, Grey model, etc. Since external factors can significantly affect electricity demand consumption, it is essential to consider the impact factors which reflect the basic trends and fluctuation for mid-term and short-term demand. However, traditional time series models which require assumptions such as normal (or exponential) distribution, linear relationships between past observations and future sequences may be problematic and lead to poor performance if the data actually exhibit a different pattern.

There has been a definite tendency to utilize the external factors for the mid-long term forecasting, in which higher accuracy and forecasting stability are desirable [30–33]. In addition to conventional factors (temperature, GDP, CPI, etc.), many newly found factors [6,34] (stock index, demand side management data, etc.) have been proven the effective in reducing forecasting error. In that case, a large and growing number of literature have been applied to investigate the influence of impact factors on electric power consumption, it can be addressed by using regression model [35,36], artificial neural networks [37,38], support vector machines [39,40]. These methods mentioned above are based on the univariate and multivariate analysis of external factors affecting electric consumption, but the potential factors can be large in real world, especially in the processing of complex changes in electric power system, where the traditional way of variables selection cannot be satisfied.

More recently, semiparametric model has been gaining popularity in economic and energy forecasting due to its apparent advantages in modeling dependence between variables. The application of semiparametric model could be traced back to the work of Engle et al. [25]. They presented the model to study the highly