Applied Energy 170 (2016) 455–465

Contents lists available at ScienceDirect

Applied Energy

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

Day-ahead hourly electricity load modeling by functional regression

Yonghan Feng [a,*], Sarah M. Ryan [b]

[a] Sears Holdings Corporation, Hoffman Estates, IL 60179, USA
[b] Industrial and Manufacturing System Engineering Department, Iowa State University, Ames, IA 50011, USA

HIGHLIGHTS

- A new hourly electricity load model is developed by considering weather and temporal dependence.
- Functional approximation is applied to model the complex nonlinear relationships between load and independent variables.
- Resulting forecast is less biased and more precise than regression-based forecasts.

ARTICLE INFO

Article history:
Received 11 November 2015
Received in revised form 31 January 2016
Accepted 20 February 2016
Available online 17 March 2016

Keywords:
Short-term load model
Forecasting
Day-ahead scenario
Epi-splines

ABSTRACT

Short-term load forecasting is important for power system generation planning and operation. For unit commitment and dispatch processes to incorporate uncertainty, a short-term load model must not only provide accurate load predictions but also enable the generation of reasonable probabilistic scenarios or uncertainty sets. This paper proposes a temporal and weather conditional epi-splines based load model (TWE) using functional approximation. TWE models the dependence of load on time and weather separately by functional approximation using epi-splines, conditional on season and area, in each segment of similar weather days. Load data are transformed from various day types to a specified reference day type among similar weather days in the same season and area, in order to enrich the data for capturing the non-weather dependent load pattern. In an instance derived from an Independent System Operator in the U.S., TWE not only provides accurate hourly load prediction and narrow bands of prediction errors, but also yields serial correlations among forecast hourly load values within a day that are similar to those of actual hourly load.

1. Introduction

Electricity load forecasting significantly influences planning and operation of power systems. Different models are used for long-term [1], medium-term [2], and short-term forecasting [3,4]. Short-term load forecasting (STLF) applies to horizons from an hour to a day ahead, and forms the basis for unit commitment, economic dispatch, maintenance plans for generators, and electricity price forecasting for utilities and independent system operators [4–6]. Because of its essential role in power systems operations, inaccurate load forecasting can cause high operational and generation cost, equipment failure, or even system blackout. As emphasis on efficient and robust scheduling of thermal generators with uncertain load prediction increases, stochastic programming and robust optimization have been extensively studied for use in the operation of power systems [7–14]. In the context of stochastic/robust unit commitment or economic dispatch by utilities or independent system operators, a regional short-term load model must provide not only accurate hour-by-hour point forecasts but also intervals and correlations to generate reasonable probabilistic scenario trajectories or ranges from which to form uncertainty sets.

A broad literature on STLF has developed for power systems of various scales. This paper focuses on regional short-term load forecasting. At the regional level, in addition to the nonstationarity in mean and variance as well as seasonal patterns, a number of external factors such as weather, time, economic activity, and social habits complicate the SLTF problem with nonlinearity. Various modeling approaches have been proposed to address these difficulties. Most of the short-term load modeling approaches can be categorized as either classical statistical methods or machine learning (ML) methods. The former often presume that load is a function of several explanatory variables; e.g., previous load values, temperature and humidity, and then estimate parameters of the specified function. In contrast, ML methods do not restrict

* Corresponding author.
E-mail addresses: yhfeng.isu@gmail.com (Y. Feng), smryan@iastate.edu (S.M. Ryan).

http://dx.doi.org/10.1016/j.apenergy.2016.02.118
0306-2619/

themselves to specified functions. The ML methods include artificial neural networks [15,16] and support vector machines [17–19]. See [4,20,21] for broad reviews of ML methods. Hybrids of statistical and ML methods have also been applied recently [22–25]. Similar day methods are often used by system operators and utilities to forecast future load and wind energy production [26,27]. In these methods, historical days with weather and day type similar to the target day are identified, and the actual load of a similar day is taken as a forecast. Such methods are often embedded in a more complicated load model, as in [16], because the methods do not sufficiently capture complex load features if used alone. Regarding smaller scale power systems including residential and commercial areas, as well as the building level, hybrids of multiple regression, time series, and ML based methods are often used [28–30].

Among statistical methods, time series and regression methods are widely used to build short-term load models. The autoregressive moving average (ARMA) model is one of the most frequently used time series methods [31–36]. In addition to a Box–Jenkins time series model, Uri improved the short-term load prediction accuracy by taking into account price and weather change [31]. Amjady identified different ARMA load models for hot days and cold days while considering effects from weekdays and weekends on daily load patterns, as well as the temperature effect on load in [32]. Huang and Shih presented a modified univariate ARMA short-term load model by considering a non-Gaussian process in [33]. Taylor first established a univariate time series load model which considered within-day and within-week seasonality by using exponential smoothing in [34]. He then improved the load prediction by accounting for the intraday, intraweek, and intrayear effects of short-term load in Britain and France without external factors [35]. Liu et al. improved upon the previous work by estimating the effects of temperature, hour-of-day and type-of-day on load level using nonparametric regression, and employed a univariate ARMA model to model the load residuals caused by the nonparametric regression [36]. In general, time series methods predict well for the immediate future, but they may suffer from imprecision in the multi-step ahead predictions because of the accumulated prediction errors.

Regression on various explanatory variables is another major direction of short-term load modeling. Usually, regression models for load forecasting include weather conditions, day types, holidays, economic conditions, and social habits. The early literature considered only impacts of temperature and holidays on load [37,38]. Haida and Muto applied transformations to include load changes in recent days and seasons, as well as annual load growth in a linear regression model [39]. Charytoniuk et al. considered load forecast as a local average of observed past loads within the local neighborhood and the specific weights on the loads defined by a multivariate product kernel [40]. Aldo et al. in [5] classified daily load curves by functional clustering, and then developed a family of functional linear regression models on the basis of obtained groups. Hong presented a multiple linear regression model that considered load as a polynomial function of temperature for each hour in each day type in each month [4]. Black also applied multiple linear regression to examine the influence of weather on load in [41], but focused on summer weekdays in the region served by the Independent System Operator of New England (ISO-NE). In the most recent published results of competitive energy forecasting, Charlton and Singleton in [42] refined the multiple linear regression model of [4], and achieved small prediction errors and weighted root mean square error (WRMSE) simultaneously by taking into account multiple weather stations, day-of-season effects and smoother temperature forecasts. Fan and Hyndman proposed a semi-parametric additive model [43], which applied cubic splines to estimate the relationship between load and temperature as well as previous loads. Most of the multiple regression methods mentioned here achieved encouraging point predictions by establishing a load model for each hour (or each half hour) in a day but, with the exception of [43], did not take into account or assess the distributions of hourly forecasts.

Although a number of load backcasting studies have presented encouraging matches to actual loads using actual weather records as input, the associated load models may not be satisfactory for predicting a day-long trajectory of hourly load or for constructing probabilistic load scenarios or uncertainty sets for day-ahead unit commitment. Even if the model accurately captures the effects of actual weather on electricity demand, the imperfect day-ahead weather forecast could distort the corresponding load forecast. To ensure that the load forecast error distribution would reflect the imprecision in both weather forecast and the modeled relationship between weather and load, models were constructed based on the day-ahead weather forecast that is available to utilities and independent system operators [44,45]. These models formed the basis for probabilistic scenarios as daylong trajectories of hourly loads [46]. Similarly to [44,45], we propose a regional day-ahead 24-h load prediction model based on the next day hourly weather forecast, but we assess the hourly non-weather dependent load component by means of an error-minimizing procedure instead of using average hourly load as in [46]. Due to its improved accuracy, the resulting load forecast is more appropriate for creating practical load scenarios.

The main contribution of this paper is a short-term load model that starts from a weather forecast and uses functional approximation rather than ordinary regression to model the complex nonlinearity between weather (i.e., temperature and dew point temperature) forecast and electricity consumption levels in multiple geographic zones. By capturing the temporal load patterns in segments of enriched data, it not only accurately predicts hourly load values but produces serial correlations among them that are similar to those of actual hourly load trajectories. The model is supported by segmenting the input data according to seasonal calendar effects on electricity consumption and by enriching the data in each segment to enhance the prediction capability. Calendar months are grouped into seasons according to monthly load-temperature patterns using *k*-means clustering. Hourly load data for a specified day type are enriched using linear transformation among different day types. In a case study derived from ISO-NE, when applied to historical weather forecast data the proposed short-term load model leads to a narrower and less skewed distribution of prediction errors for each zone in each season, compared to two recent regression based short-term load models [4,42].

The rest of the paper is organized as follows. After an overview of the data processing and modeling steps, Section 2 develops a weather conditional short-term load model by applying an epi-spline approximation technique. To demonstrate its regional application, Section 3 describes the partitioning of data into several segments according to calendar effects (day types and seasons), and enrichment by transformation to a reference day type. The numerical results in Section 4 indicate that our model captures the complex nonlinear relationship between electricity consumption level and weather conditions, and attains narrower and less skewed distributions of prediction errors, compared to those of two benchmark models. Section 5 concludes this paper.

2. Day-ahead hourly load model

In this paper, day-ahead hourly load models are developed for multiple zones within a region based on their corresponding weather forecasts. The temporal relationship between weather forecast and electricity consumption changes during the year