Applied Energy 205 (2017) 654–669

Contents lists available at ScienceDirect

Applied Energy

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

Short-term residential load forecasting: Impact of calendar effects and forecast granularity

Peter Lusis[a,b], Kaveh Rajab Khalilpour[a,c,*], Lachlan Andrew[a], Ariel Liebman[a,c]

[a] Faculty of Information Technology, Monash University, Australia
[b] Technical University of Denmark, Anker Engelunds Vej 1 Bygning 101A, 2800 Kgs. Lyngby, Denmark
[c] Monash Energy Materials and Systems Institute (MEMSI), New Horizons Research Centre, Monash University, Australia

HIGHLIGHTS

- Investigation of the impact of data resolution and calendar effects.
- Support vector regression yields a higher accuracy for a day-ahead load forecast.
- The forecast error can be reduced by using coarser forecast granularity.
- Calendar effects added to the model as dummy variables have little predictive power.
- One year of historical data is sufficient to develop a load forecast model.

ARTICLE INFO

Keywords:
Short-term load forecasting
Residential load
Calendar effects
Granularity
Distributed generation and storage management
Disaggregated load

ABSTRACT

Literature is rich in methodologies for “aggregated” load forecasting which has helped electricity network operators and retailers in optimal planning and scheduling. The recent increase in the uptake of distributed generation and storage systems has generated new demand for “disaggregated” load forecasting for a single-customer or even down at an appliance level. Access to high resolution data from smart meters has enabled the research community to assess conventional load forecasting techniques and develop new forecasting strategies suitable for demand-side disaggregated loads.

This paper studies how calendar effects, forecasting granularity and the length of the training set affect the accuracy of a day-ahead load forecast for residential customers. Root mean square error (RMSE) and normalized RMSE were used as forecast error metrics. Regression trees, neural networks, and support vector regression yielded similar average RMSE results, but statistical analysis showed that regression trees technique is significantly better.

The use of historical load profiles with daily and weekly seasonality, combined with weather data, leaves the explicit calendar effects a very low predictive power. In the setting studied here, it was shown that forecast errors can be reduced by using a coarser forecast granularity. It was also found that one year of historical data is sufficient to develop a load forecast model for residential customers as a further increase in training dataset has a marginal benefit.

1. Introduction

Electricity demand depends on several factors including weather, time, and socio-economic constraints [1]. Load forecasting considers these factors to facilitate the decision-making process of unit commitment, economic dispatch, and power system operation [2]. At low voltage level, demand forecasting improves optimal load control and circuit switching [3]. The traditional centralized power generation from conventional power plants involved little uncertainty. Utilities focused on the statistical accuracy of a cluster of loads rather than a single household [4]. This is changing due to the transition to distributed energy generation from intermittent energy sources, the decentralization of the electricity market, and the rising number of demand side control systems. The scale of management has been moved down to microgrid level and single households [5]. Around 30% of the global electricity demand is related to the power consumption in the

* Corresponding author at: Faculty of Information Technology, Monash University, Australia.
E-mail addresses: lusis.peteris@gmail.com (P. Lusis), kr.khalilpour@monash.edu (K.R. Khalilpour), lachlan.andrew@monash.edu (L. Andrew), ariel.liebman@monash.edu (A. Liebman).

http://dx.doi.org/10.1016/j.apenergy.2017.07.114
Received 28 March 2017; Received in revised form 29 June 2017; Accepted 26 July 2017

residential sector [6]. Photovoltaic (PV) systems are the most widespread distributed generation technology in the residential sector by far, reducing the demand reliance on conventional power plants and providing the peak load shaving [7].

Rooftop PV systems in the residential sector require new operational strategies to maximize households' self-sufficiency and minimize the negative effect of the afternoon PV feed-in spikes on the grid [8]. To mitigate the adverse effects of load fluctuations and voltage instability, as well as store excess PV output, the PV system can be combined with battery storage [9]. The battery is charged when the surplus energy from PV panels is available. When the battery is full and the local demand is met, the surplus power is fed into the grid. Considering that this process occurs in multiple households, a power feed-in peak occurs requiring the grid operator to pursue load balancing procedures. As there is a growing solar power generation, it is necessary to find a trade-off between the amount of the solar energy fed into the grid and optimal PV-battery system operation from the household perspective [8].

Households' demand changes quickly as appliances for domestic chores are turned on and off. While the system-wide load forecast takes advantage of the load smoothing effect from multiple households, at the individual household level rapid fluctuations cannot be avoided, making the load prediction more challenging [10]. Until the recent deployment of smart meters, there was a shortage of high resolution data from individual households [3]. A rapid increase in distributed generation and research in demand side management created a need for disaggregated data with a high sample rate enabling research in household load forecasting.

The scheduling of batteries is strongly influenced by errors in input variables [11]. Whether the household PV-battery system management is carried out locally or in the cloud environment by a third party, load forecasting has an important role [5]. The forecast accuracy is even more critical for single stand-alone systems or microgrids, where unforeseen, high-frequency irradiance fluctuations can result in severe voltage fluctuations [12]. A better voltage control can be achieved by including the battery state of charge in the microgrid operation model as this allows the battery to absorb some of the forecast error [13].

PV-battery system scheduling for residential customers is optimized with complex mathematical programming algorithms or relying on more efficient geometrical methodology. Case studies for PV-battery system optimization are often based on historical data rather than solar irradiation and load forecasts [14]. Since household load is less predictable than the overall system forecast, excluding load forecast errors from PV-battery management gives a perfect, but unrealizable solution. This paper examines how various calendar effects, forecast granularity, and forecasting strategies affect the load forecast accuracy with different techniques. The aim is to select a short-term load forecast model that should be deployed in the optimization process of household distributed generation and storage systems, including PV-battery.

2. Progress in household load forecasting

The load of any single residential customer is less predictable than a more aggregated load [3]. New forecasting approaches, which arise from the smoothing effect of household load aggregation, have been proposed. Humeau et al. [15] analyzed the load consumption of 782 households and found that the normalized forecast error decreased with the growing number of households in the cluster. Gajowniczek et al. [10] proposed a blind source separation approach for households with a similar load pattern; an improved forecast was achieved by decomposing the original forecasts into a set of independent components and classifying and eliminating some of the noise [10].

Historical load and weather data are at the heart of load forecast models. Several studies have focused on developing new or advanced features in order to improve the forecast model. Beccali et al. [16] introduced the "humidex index" that accounts for heating and cooling demand due to the thermal discomfort felt by household residents. Soliman et al. [17] defined "wind chill index" for winter months based on wind speed and air temperature. Taieb et al. [18] addressed weather and electricity demand uncertainty by proposing probabilistic forecasts based on quantile regression. Rodrigues et al. [19], in contrary, avoided the uncertainty of weather forecast by using only household physical and demographic data in combination with calendar effects. Sandels et al. [20] generated realistic household electricity consumption patterns combining behavioral models of residents with their electricity, hot water and space heating usage. Javed et al. [21] obtained a higher forecast accuracy when adding socio-economic factors such as the number of occupants, the age of occupants and the hours of day spent at home. Tascikaraoglu et al. [22] proposed a spatio-temporal approach considering the correlation between the energy usage in the target house and the houses surrounding it. Jain et al. [23] studied the accuracy of a step-ahead forecast in residential buildings examining multiple spatial levels within a building and various temporal granularities.

It is a common practice to apply multiple forecasting techniques and compare them against a perfect-forecast model, which is based on observed load values [12]. This benchmark is used to assess the accuracy of the proposed model. An alternative benchmark is a persistence model that takes an advantage of the fact that the load remains relatively constant for a short period. Therefore, it is often difficult to beat a persistence model in a short-term.

Traditional load forecasting techniques are based on time series or regression analysis. Time series models such as exponential smoothing [21], autoregrssive integrated moving average (ARIMA) models [24], and seasonal ARIMA [22,25], largely rely on correlation between the load and its past values. Other traditional load forecasting techniques include statistical methods such as regression trees [26], and multiple linear regression [5,22]. To tackle the non-linear and highly dynamic load fluctuations of residential customers, artificial intelligence techniques have become popular in load forecasting [27]. The main techniques include neural networks (NN) [16,19,21] and support vector machines (SVM) [15,23,27–29]. While artificial intelligence model tends to provide slightly better forecast [30], this comes at a cost of longer computational times. The optimal number of layers and neurons in neural network model has to be determined empirically [25]. A more accurate solar generation forecast model was achieved by [31], when combining several forecasting algorithms into a single forecasting technique. Stephen et al. [32] demonstrated that a forecast ensemble consisting of neural networks, Gaussian load profile, ARIMA, persistent and flat forecast models significantly outperforms the forecast model built on a single forecasting technique for aggregated load forecasting. Liu et al. [33] proposed a hybrid short-term load forecasting model with parameter optimization and focused on the model's implementation in the microgrid management.

Calendar effects comprise any changes in load consumption related to calendar periods. The use of calendar effects in load forecasting captures weekly and seasonal energy consumption patterns [34] and facilitates the prediction of peak demand [35]. Fewer studies have focused on the interaction between the residential load and calendar variables. Various approaches can be found in the implementation of calendar features into forecasting models. To capture the similarities in load variation in different time periods, dummy explanatory variables are used to represent each day of week [29], split weekdays and weekends or assign variables for different parts of day [5]. Instead of using dummy variables, the historical data can be split into subsets according to the same day of week [36] or by certain hours of day. It is not clear which of two approaches will yield a higher forecast accuracy. Similarly, seasonal variability (winter-summer) can be addressed through any of these approaches.

Past studies have built forecasting models using datasets from 60 days [30] to two years [25]. A common approach is to use all available data or, if the dataset is incomplete, select the longest period with complete information to build a forecasting model [36]. As the