Solar Energy 158 (2017) 236–246

Contents lists available at ScienceDirect

Solar Energy

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

Net load forecasts for solar-integrated operational grid feeders

Yinghao Chu[a,b], Hugo T.C. Pedro[a], Amanpreet Kaur[a,c], Jan Kleissl[a], Carlos F.M. Coimbra[a,*]

[a] Department of Mechanical and Aerospace Engineering, Jacobs School of Engineering, Center of Excellence in Renewable Resource Integration and Center for Energy Research, University of California, 9500 Gilman Drive, La Jolla, CA 92093, USA

[b] Current address: Environmental Engineering Center, China Nuclear Power Technology Research Institute CO., Ltd, 1620 Shenzhen Science and Technology Building, 1001 Shangbuzhong Road, Futian, Shenzhen, PR China

[c] Current address: SolarCity, 3055 Clearview Way, San Mateo, CA 94402, USA

ARTICLE INFO

Keywords:
Net load forecasts
Sky imaging
Support vector machines
Artificial neural networks
Solar integration

ABSTRACT

This work proposes forecast models for solar-integrated, utility-scale feeders in the San Diego Gas & Electric operating region. The models predict the net load for horizons ranging from 10 to 30 min. The forecasting methods implemented include hybrid methods based on Artificial Neural Network (ANN) and Support Vector Regression (SVR), which are both coupled with image processing methods for sky images. These methods are compared against reference persistence methods. Three enhancement methods are implemented to further decrease forecasting error: (1) decomposing the time series of the net load to remove low-frequency load variation due to daily human activities; (2) segregating the model training between daytime and nighttime; and (3) incorporating sky image features as exogenous inputs in the daytime forecasts. The ANN and SVR models are trained and validated using six-month measurements of the net load and assessed using common statistic metrics: MBE, MAPE, rRMSE, and forecast skill, which is defined as the reduction of RMSE over the RMSE of reference persistence model. Results for the independent testing set show that data-driven models, with the enhancement methods, significantly outperform the reference persistence model, achieving forecasting skills (improvement over reference persistence model) as large as 43% depending on location, solar penetration and forecast horizons.

1. Introduction

Uncertainties in electric loads require management of operating reserves and dispatchable ancillary generation, increasing the overall costs for utilities, customers, system operators, and other market participants (Ortega-Vazquez and Kirschen, 2006). Short-term load forecasts play a key role in mitigating the uncertainty of loads and are essential to decrease these costs (Kaur et al., 2014). The earliest studies on electrical demand/load forecasts date back to 1960 (Gross and Galiana, 1987). Since then various techniques such as time series analysis and regression (Hagan and Behr, 1987; Moghram and Rahman, 1989), data-driven learning and artificial intelligence (Alfares and Nazeeruddin, 2002; Metaxiotis et al., 2003), hybrid or ensemble models (Kaur et al., 2014; Hahn et al., 2009; Alamaniotis et al., 2012; Matijaš et al., 2013), have been developed and comprehensively reviewed in the literature. A detailed review of recently methods can be found in Suganthi and Samuel (2012).

Despite its long history, interest in forecasting electricity net demand is on the rise due to the increasing penetration of variable renewable generation fomented by favorable net energy metering tariffs and other incentives (Kaur et al., 2013). The impact of solar penetration on the net load profile is demonstrated in the relevant literature (Inman et al., 2013; Azadeh et al., 2009; Nguyen et al., 2016). During the daytime, variable rooftop solar generation adds substantial uncertainty to power demand at the substation level (Denholm and Margolis, 2007). For operating feeders with high solar penetration levels, the variability in solar power production propagates into the net load (load minus solar generations) profile and increases the error of net load forecasts if the model is not prepared to accept the added variability (Kaur et al., 2013). Given the nature of the additional variability we can anticipate that net load forecast models can be enhanced by integrating techniques used in the forecasting of solar irradiance (e.g. using image features from sky cameras as exogenous inputs to the model (Schmidt et al., 2016)). However, to the best of our knowledge, this approach has not been explored for utility-scale net load time series and, more importantly, the high level of solar penetration (larger than 50% in one of the feeders) studied in this work.

To accurately forecast the net load with the influence of distributed solar production, two popular stochastic models are employed in this work: Artificial Neural Network (ANN) and Support Vector Machine

* Corresponding author.
E-mail address: ccoimbra@ucsd.edu (C.F.M. Coimbra).

http://dx.doi.org/10.1016/j.solener.2017.09.052
Received 3 September 2016; Received in revised form 14 September 2017; Accepted 24 September 2017

(SVM). These two models are developed using data from operational grid feeders and are implemented for 10-, 20-, and 30-min forecast horizons. The ANN/SVR models are enhanced using three proposed methods: time series detrending, daytime/nighttime model training, and using sky image features as exogenous inputs to the stochastic models. Performance of the proposed models is assessed in terms of common statistical metrics and compared against reference persistence models.

The major contributions of this work are: (1) the development of appropriate real-time stochastic forecasting models for utility-scale net load time series of multiple operational grid feeders with significant solar penetration levels, (2) the evaluation of the effectiveness of enhancement methods based on data collected from the operational feeders. Ultimately, the proposed forecasting models and the recommended enhancement methods can be used to improve the accuracy and robustness of the net load forecasts for public utilities and system operators.

2. Data

In this work, forecasts up to 30 min are developed and implemented for four feeders, denoted as A, B, C, and D, located in the operating region of the San Diego Gas & Electric (SDG & E) utility company. Information about the four feeders is presented in Table 1 (CSS, 2015; Nguyen et al., 2015). In the table the solar penetration for a feeder is defined as the installed rooftop solar generation capacity divided by peak load demand (CSS, 2015). The solar penetration in the feeders ranges from 4% to 58%.

For all feeders, load data averaged in 10-min bins are collected from Oct. 2014 to Mar. 2015 (22550 time instances). Both daytime and nighttime data (when solar generation is zero) are considered in this work in order to develop the proposed 24-h operational load forecasts.

Data availability limits the analysis presented in this work to a 6-month period. In order to determine if these data properly describes the variability at the four locations, we estimate the probability density function (PDF) for different sizes of the datasets, starting with 2% of the data (≈ four days) and augmenting the dataset in 2% steps until all data points are included. The PDFs are estimated using the kernel density estimation technique (Bowman and Azzalini, 1997). Additionally, the convergence of the PDFs is tested by computing the relative mean deviation between two consecutive PDFs. The results for the estimated PDFs and the respective convergence are shown in Fig. 1 for the four feeders. The color scale in the figures indicates how much data was used in the PDF estimation. The figures indicate that the PDFs for all locations converge when considering all the data available to us. This analysis shows that, the 6 months of data here considered are well representative of the data variability, thus ensuring that the results presented below provide a general view of the forecasting performance under these conditions. Furthermore, because the core conclusions of this work relate to comparisons between forecasting models, 6 months of data provide enough data points to corroborate our conclusions.

Data for each feeder are divided into two disjointed datasets: the training dataset (the first three weeks of each month) for model training/optimization, and the testing dataset (the last week of each month) for model validation.

Table 1
Details about four SDG & E feeders for which the net load forecasting is implemented and tested (CSS, 2015; Nguyen et al., 2015).

SDGE feeders	Feeder A	Feeder B	Feeder C	Feeder D
Distance from the coast (km)	50	< 5	30	20
Feeder length (km)	34.9	39.6	177.8	51.5
Number of loads	471	584	1733	468
Average load (MW)	1.27	3.37	3.46	1.32
Peak load (MW)	2.60	6.39	5.93	2.47
Solar penetration	4%	15%	21%	58%

The distributed solar power generation in these regions depends on ground-level solar irradiance. However, solar irradiance measurements are not available for these sites. Solar irradiance highly depends on weather conditions, particularly the cloud cover (Chu et al., 2016; Li et al., 2016; Inman et al., 2016), which can be analyzed using ground-based sky imagers. Therefore, after considerations of costs and field restrictions, two UCSD Sky Imagers (USI) (Yang et al., 2014) were installed near Feeders A and B to provide cloud cover information. The USI employs an upward-facing charge-coupled device (CCD) image sensor to capture images (2048 × 2048 pixels), which are transferred via FTP to a UCSD server. Deployment of sky imager and other solar observation equipment at locations with higher level of solar penetration will be considered in future work.

The sky images obtained with the USI are processed to obtain numerical image features, which are used as exogenous inputs to the forecasts. Load forecasts for feeders without sky imagers are developed based on endogenous inputs. These consist of 10-min averaged net load values computed on a one hour sliding window (six values) that precedes the forecasting issuing time. The forecast outputs of this work are the 10-min averaged net load from 10 to 30 min ahead in steps of 10 min. For example, the output of 20-min forecast is the averaged load in the 10 to 20-min window that follows the forecasting issuing time.

3. Methods

3.1. Stochastic models

Both ANN and SVR are used in this work to implement the net load forecasts. Artificial Neural Networks (ANN) and Support Vector Regression (SVR) are popular data-driven tools for pattern recognition, data classification and regression, and have proven to be useful for nonlinear input/output mapping (Inman et al., 2013; Chang and Lin, 2011). The weights and parameters of both ANN and SVR are estimated using the training dataset. Inputs to both stochastic models are normalized before the training process.

The basic processing elements of the ANN are the neurons, which are interconnected and placed in layers. The layers between the first input layer and the last output layer are called hidden layers. Neurons of both hidden and output layers take in weighted sum of inputs Xj from previous layers. Each neuron produces one output using an activation function. The output from each neuron of the hidden layers is used as inputs to all neurons of the subsequent layer. The output from each neuron of the output layer is considered as one ANN output. In this work, the output layer has three neurons producing three ANN outputs (10-, 20-, and 30-min ahead predictions). The ANNs used in this work are feedforward networks, that is, only forward connections between the neurons are allowed. Mathematically, each neuron of the ANNs can be represented as:

$$Y = f\left(\sum_{j=1}^{M} (w_j X_j + \beta)\right), \tag{1}$$

where Y is the output from this neuron, f is an activation function (sigmoid function in this work), w_j are the weights for the j-th input X_j from previous layer that consists of M neurons, and β is a bias term of the current layer. In this work, we use the Bayesian regularization process with Levenberg-Marquardt optimization (Chu et al., 2013). The training procedure can be summarized as (Rumelhart and Zipser, 1985; Chu et al., 2016):

1. Set the number of layers, number of neurons in each layer, and a tolerance parameter $\varepsilon > 0$.
2. Initialize the values of MLP parameters weight vector (w_j^i,$i = 0,1,2,...,L$) randomly.
3. Calculate the neuron outputs $X_j^i = f((w_j^i)^T X^{(i-1)})$ and final output