Energy Conversion and Management 75 (2013) 701–709

Contents lists available at ScienceDirect

Energy Conversion and Management

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

Impact of onsite solar generation on system load demand forecast

Amanpreet Kaur, Hugo T.C. Pedro, Carlos F.M. Coimbra *

Department of Mechanical and Aerospace Engineering, Jacobs School of Engineering, Center for Energy Research and Center of Excellence in Renewable Resource Integration, University of California, San Diego, La Jolla, CA 92093, USA

ARTICLE INFO

Article history:
Received 19 June 2013
Accepted 18 August 2013

Keywords:
Load forecasting
Solar forecasting
High solar penetration
Error distribution
Photovoltaic farms

ABSTRACT

Net energy metering tariffs have encouraged the growth of solar PV in the distribution grid. The additional variability associated with weather-dependent renewable energy creates new challenges for power system operators that must maintain and operate ancillary services to balance the grid. To deal with these issues power operators mostly rely on demand load forecasts. Electric load forecast has been used in power industry for a long time and there are several well established load forecasting models. But the performance of these models for future scenario of high renewable energy penetration is unclear. In this work, the impact of onsite solar power generation on the demand load forecast is analyzed for a community that meets between 10% and 15% of its annual power demand and 3–54% of its daily power demand from a solar power plant. Short-Term Load Forecasts (STLF) using persistence, machine learning and regression-based forecasting models are presented for two cases: (1) high solar penetration and (2) no penetration. Results show that for 1-h and 15-min forecasts the accuracy of the models drops by 9% and 3% with high solar penetration. Statistical analysis of the forecast errors demonstrate that the error distribution is best characterized as a *t*-distribution for the high penetration scenario. Analysis of the error distribution as a function of daily solar penetration for different levels of variability revealed that the solar power variability drives the forecast error magnitude whereas increasing penetration level has a much smaller contribution. This work concludes that the demand forecast error distribution for a community with an onsite solar generation can be directly characterized based on the local solar irradiance variability.

1. Introduction

California generated approximately 20% of its in-state power using renewable energy resources in 2011 [1]. This percentage must increase substantially by 2020 if the state is to meet the guidelines mandated by its Renewable Portfolio Standard (RPS), which stipulates that a minimum of 33% of in-state electricity must originate from renewable resources like solar, wind, tidal and small hydroelectric power plants. To achieve this aggressive goal of RPS, California's AB 920 assembly bill allows customer-generators to receive a financial credit for the power fed back into the grid by their renewable generation system. Due to the inherent variability of renewable resources, most notably solar and wind, the increase in renewable energy penetration results in additional variability and uncertainty in the power put into the electric grid [2,3]. This situation can also result in additional variability and uncertainty in customer demand if the onsite power generation is not enough to meet the customer's demand. Given that the concept of onsite renewable energy generation is relatively new, its impact on customer demand and load forecast is unclear. This impact needs to be well understood to ensure a reliable grid operation as the planning of resource allocation depends greatly on demand load forecast. Accurate load forecasts are important given that, according to Western Electricity Coordinating Council (WECC) [4], the power systems must maintain an operating reserve to balance against the forecasting error and other unexpected power source failures in the electric grid. To fill this gap, we present a comprehensive case study of UC Merced campus with 1 MW of onsite solar generation plant.

Through a Power Purchase Agreement UC Merced campus contracted a single-axis tracking solar farm that annually produces 3–54% of the daily daytime campus power demand, making the campus a good proxy to study the impact of onsite solar generation. The centralized HVAC (Heating, Ventilation and Air Conditioning) system for the campus includes a Thermal Energy Storage (TES) water tank that operates at night when electricity prices and ambient temperature are lowest. For these reasons the load profile of UC Merced has a lower power demand during the day and higher power demand during the night which, generally speaking, is inverted with respect to the usual load profile for similar facilities

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

0196-8904/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.enconman.2013.08.026

Nomenclature

$\hat{\cdot}$	forecast of $\cdot$
$A(q)$	polynomial of order n_a
$C(q)$	polynomial of order n_c
d	given day
e	disturbance
$Eavg$	mean forecast error
L_g	load with onsite generation
L_{ng}	load with no onsite generation
L_{PV}	solar power generation
N	number of 15-min time steps
P	polynomial fit
q	shift operator
R^2	coefficient of determination
s	forecast skill
$S_{PV}(d)$	daily solar penetration (%)
V_{Lg}	step changes in load with onsite generation
V_{Lng}	step changes in load without onsite generation
$V_{PV,N}$	solar power variability between N steps
AR	Autoregressive Model
ARMA	Autoregressive Moving Average Model
GHI	Global Horizontal Irradiance
HVAC	Heating, Ventilation and Air Conditioning
kNN	k Nearest Neighbors Model
MAE	Mean Absolute Error
MBE	Mean Bias Error
P	Persistence Model
RMSE	Root Mean Square Error
SP	Smart Persistence Model
TES	Thermal Energy Storage

[5] – see Fig. 1. With current efforts to integrate more solar energy into the power grid, such as the net energy metering, we expect that such load profiles will become more common in the future. The additional variability in the demand load together with diurnal and annual solar cycles results in bigger challenges in load forecasting, balancing the power grid and managing ancillary systems [2,6,7]. Accurate forecasts of both demand and supply profiles are being pursed to mitigate these issues and guarantee adequate supply of electricity and reliable grid operation.

The earliest efforts to forecast electrical demand go back to the 1960s [8]. Several reviews on load forecasting methods have been published [9–18] and there is continuous research to develop better methods [19–31]. Because load profiles are expected to change with onsite renewable energy generation especially with intermittent solar generation [6]. It is very important to analyze the forecast model performance for such scenarios. Therefore, two cases were studied for 15-min and 1-h forecasting horizon: onsite generation and no onsite generation. Onsite generation case represents the campus power demand from the grid after consuming all the solar power produced on campus. No generation refers to the power demand that campus would have extracted from the electric grid if there was no solar generation on the campus. As seen in Fig. 1 the former is greatly affected by the variability of the solar resource. Several well established Short-Term Load Forecast (STLF) methods were applied to predict these two time-series: persistence models, regression based models and machine learning models. To make it simple and rely less on inputs, methods with non-exogenous inputs were applied in line with some of our previous work [32]. Like previous works that have studied the error distribution for wind forecasts [33–36], we characterized the error distribution of our predictions in order to understand the impact of additional variability in forecast accuracy.

The data used for this study is presented in Section 2. The models are described in Section 3. Results and discussion are presented in Section 4, where the accuracy of the forecasting models is evaluated and compared using standard statistical error metrics. The forecast error distribution is presented and characterized for the two scenarios, and the impact of onsite solar generation on forecast error is analyzed for different solar power variability levels. The main conclusions of this study are presented in Section 5.

2. Data

In this work we used two datasets: UC Merced demand load from the grid (campus demand after consuming all the solar power produced on campus) which represents the onsite generation case (L_g) and total UC Merced power demand (demand that the campus would extract from the grid if there was no solar power plant), which represents the no onsite generation case (L_{ng}). L_{ng} was obtained by adding the power consumed from the grid and the solar farm power output, that is $L_{ng}(t) = L_g(t) + L_{PV}(t)$ where L_{PV} represents the solar power generation on the campus. For all cases the data points consisted of 15 min backward averages.

2.1. Preprocessing data

The time-series was decomposed by removing the daily trend calculated over the whole year. A 6th order polynomial (P) was fit to both cases as shown in Fig. 2. The detrended demand loads can be represented as no onsite generation $L^{dt}_{ng}(t) = L_{ng}(t) - P_{ng}(t)$ and onsite generation $L^{dt}_{g}(t) = L_g(t) - P_g(t)$.

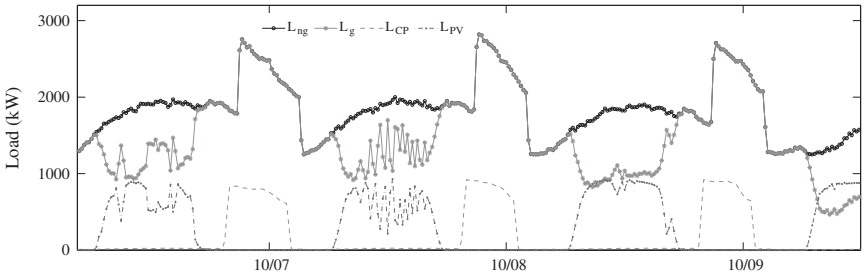

Fig. 1. UC Merced load profile for 06-Oct-2010 through 09-Oct-2010. The UC Merced campus has a unique load shape because the majority of the HVAC load (L_{CP}) has been shifted to the night time using Thermal Energy Storage (TES). The total energy consumed by the campus (L_{ng}), comprise of the energy from the solar farm and energy from electric grid, is comparatively smoother than the demand load (L_g) that is affected by onsite solar generation.

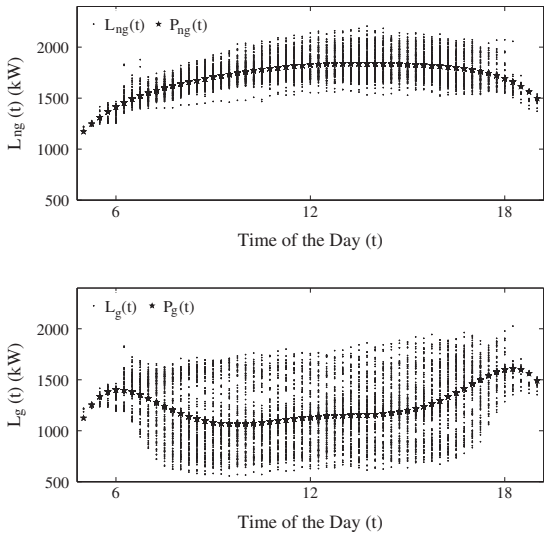

Fig. 2. Daily demand load profile for the UC Merced campus with no onsite power generation (*top*) and onsite solar power power generation (*bottom*). The trend in the daily load profile is computed by fitting a 6th order polynomial (P) to daily load profile over the whole year.

2.2. Training and validation datasets

Throughout this study we assumed the following:

1. At UC Merced, the solar power produced is consumed at the time of production (there is no storage). Therefore, given that the focus of this study is to analyze the impact of onsite solar generation, only daytime hours were considered.
2. The TES load at UC Merced is a deterministic load, that is, it is always known beforehand. Therefore, the load demand due to TES is removed from the datasets. Since the TES is operated at night this assumption is included in the previous one.
3. Like any office or school building UC Merced has different load profiles for weekdays, weekends and holidays. Because we wanted to isolate the impact of onsite generation from the other factors (such as school occupancy) we only considered work days for the year 2010 in our dataset.

After taking into account the above assumptions the resulting time-series was further separated into two disjoint datasets:

1. *Training or model evaluation set*: it consists of data from January 2010 to December 2010 minus a week from each month.
2. *Validation set*: data used to test the models performance and robustness for different times of the year. This dataset consists of a week from each month (the data not included in the training set).

3. Forecasting models

Forecasting models for time-series often use the series' past values as inputs. Because of that, it is useful to introduce the forward shift operator (q^N) and backward shift operator (q^{-N}) [37]

$$q^{\pm N}L(t) = L(t \pm N) \tag{1}$$

where $L(t)$ represents the load demand time-series for both cases, and N is the number of 15-min time steps (because the time-series consists of 15-min time averages).

3.1. Persistence (P) Model

The persistence model is one of the simplest forecasting model. It is based on the assumption that the current state of the system persists between the present time and the time of the forecast. Mathematically, the forecasted load ($\hat{L}$) can be represented as

$$\hat{L}(t) = q^{-N}L(t) \tag{2}$$

where N is 1 for the 15-min and 4 for the 1-h forecast horizons, respectively.

3.2. Smart Persistence (SP) Model

The performance of the persistence model can be improved by using information about the time-series trend. Such a model is termed Smart Persistence (SP) model. The additional information can be added in two ways:

1. SP1: this model assumes that the difference, with respect to the trend, persists to the future. In mathematical terms it can be represented as
$$\hat{L}(t) = q^{-N}L(t) - q^{-N}P(t) + q^{N}P(t). \tag{3}$$
2. SP2: a second variation of the persistence model assumes that the ratio between the time-series and its trend remains unchanged in the future. This persistence model can be represented as
$$\hat{L}(t) = \frac{q^{-N}L(t)}{q^{-N}P(t)}q^{N}P(t). \tag{4}$$

3.3. Autoregressive (AR) Model

Another model tested in this work is the autoregressive model. This model is a generalization of the persistence model in the sense that it may involve any lagged past values of the time-series. Mathematically, it can be presented as [37]

$$L(t) + a_1L(t-1) + \cdots + a_{n_a}L(t-n_a) = e(t). \tag{5}$$

Introducing the polynomial $A(q)$ which contains lagged values given by

$$A(q) = 1 + a_1q^{-1} + \cdots + a_{n_a}q^{-n_a}, \tag{6}$$

the AR-model becomes

$$A(q)L(t) = e(t) \tag{7}$$

where $e(t)$ is the disturbance in the time-series which cannot be characterized by the lagged past values of the time-series.

3.4. Autoregressive Moving Average (ARMA) Model

In ARMA models, information about the lagged past values of the disturbance term, $e(t)$ is added to the model. The mathematical expression for ARMA models is

$$L(t) + a_1L(t-1) + \cdots + a_{n_a}L(t-n_a) = e(t) + c_1e(t-1) + \cdots + e_{n_c}L(t-n_c), \tag{8}$$

which, with the introduction of the operators $A(q)$ and $C(q)$

$$A(q) = 1 + a_1q^{-1} + \cdots + a_{n_a}q^{-n_a}, \tag{9}$$

$$C(q) = 1 + c_1q^{-1} + \cdots + c_{n_c}q^{-n_c}, \tag{10}$$

can be rewritten as

$$A(q)L(t) = C(q)e(t). \tag{11}$$

3.5. k Nearest Neighbors (kNN) Model

The kNN algorithm is one of the simplest methods among the machine learning algorithms. It is a pattern recognition method for classifying patterns or features [38]. The classification is based on the similarity of a pattern of current values with respect to training samples in the feature's space. For the purpose of forecasting time-series, the kNN model looks into the series' history and identifies the timestamps that resemble the current conditions most closely. Once the best matches are found the forecast is determined using values from the time-series subsequent to the matches.

In this work the patterns, or features, consisted of the past hour values. The features for time t are assembled in the vector $\vec{p}(t)$ with components p_j. The features for historical data are assembled in a matrix A_{ij} whose rows correspond to the vector of features for each time in the historical dataset. Once the matrix A_{ij} is assembled, we compared the vector of features p_j against all the rows in A_{ij}

$$e_i = \sqrt{\sum_j (A_{ij} - p_j)^2}. \tag{12}$$

The list of errors was sorted in ascending order and the indices corresponding to the first N elements (where $N = 10$ in this case) of the sorted list were taken as the best matches. With the sets of best matches determined, the forecast for time t is calculated as

$$\hat{L}(t) = \frac{\sum_{i=1}^{N} w_i L_{train,i}}{\sum_{i=1}^{N} w_i}, \tag{13}$$

which involves only data in the training set. Different forecasted values can be obtained depending on the weights w_i. In this work we used two distributions for w_i.

3.5.1. kNN1

In the first one, every term $L_{train,i}$ is weighted equally that is $\vec{w} = \{1, 1, \ldots, 1\}$.

3.5.2. kNN2

For the second kNN forecast the weights vary linearly, with the highest weight attributed to the term associated with smallest error (first term of $L_{train,i}$) and the lowest weight attributed to term with the largest error (the last term), that is, $\vec{w} = \{10, 9, \ldots, 2, 1\}$.

3.6. Error metrics

The different models presented above were compared quantitatively with standard statistical analysis of the forecast error, which is calculated as the difference between the forecasted value and the measured value. In this work we used: Mean Absolute Error (MAE), Mean Bias Error (MBE), Root Mean Square Error (RMSE) and Coefficient of determination (R^2). The MBE is a measure of systematic errors (or bias), the RMSE is a measure of random errors and the MAE is the magnitude of the forecasting error. The coefficient of determination R^2 measures the level of dispersion about the 1:1 line in a scatter plot of measured vs. forecasted values. A coefficient of determination of 1 means a perfect forecast. In addition to these error metrics, a new metric known as forecast skill (s) [39] is used to compare the accuracy of the forecasting models with respect to the persistence model

$$s_{\text{Model}} \approx 1 - \frac{\text{RMSE}_{\text{Model}}}{\text{RMSE}_{\text{Persistence}}}. \tag{14}$$

4. Results and discussion

The forecasting models explained above were implemented in Matlab. The persistence models and the kNN models are fully specified with the information supplied above. On the other hand, the actual form of the AR and ARMA models (the values of the coefficients in the polynomials $A(q)$ and $C(q)$) was computed using the System Identification toolbox in Matlab. For this process, following the parsimony principle, only 4 parameters were used in these models. For AR, the coefficients of the polynomial $A(q)$ are listed in Table 1 and for ARMA, the coefficients for $A(q)$ and $C(q)$ are listed in Table 2.

These results were obtained with the detrended training dataset. However, for the error analysis that follows next, the trend was added to the results and the error was calculated using actual and forecasted demand loads.

4.1. 1-h forecasts

The results for various statistical error metrics for this forecasting horizon are tabulated in Table 3. From the results listed for no onsite generation we can conclude that the best models are the Smart Persistence models and the regression models (AR and ARMA) which perform similarly in terms of R^2. On the other hand, the kNN2 model performs the worst with a negative forecast skill. From this table we conclude that in terms of MAE and RMSE, SP1 and SP2 models perform the best. These observations show that for no onsite generation case (smoother load curves), simple models based on persistence can perform very well with the highest forecast skill of 0.24.

In the case of onsite solar power generation case, for the 1-h demand forecast, the regression based models AR and ARMA both capture the relationship in the time-series better than the other models in terms of R^2 (see Table 3). In terms of forecast skill the AR model performs the best with $s = 0.1$. However, in terms of MAE and RMSE, the error is smallest in case of the SP2. As in the previous case, the kNN models perform the worst with negative forecast skill, $s = -0.003$. The reason for such poor performance could be related to the absence of similar patterns in the training set.

4.2. 15-min forecasts

The results for various statistical error metrics for this forecast horizon are tabulated in Table 4. For this horizon and no onsite

Table 1
The coefficients of the polynomial $A(q)$ for AR model for the two cases: no onsite generation (no) and onsite solar generation (yes).

Forecast horizon	Case	$A(q)$
1-h	No	$1 - 0.72q^{-1} - 0.07q^{-2} - 0.12q^{-3} - 0.06q^{-4}$
	Yes	$1 - 0.89q^{-1} + 0.03q^{-2} - 0.07q^{-3} - 0.01q^{-4}$
15-min	No	$1 - 0.79q^{-1} - 0.11q^{-3} - 0.06q^{-4}$
	Yes	$1 - 0.92q^{-1} + 0.05q^{-2} - 0.08q^{-3}$

Table 2
The coefficients of the polynomials $A(q)$ and $C(q)$ for ARMA model for the two cases: no onsite generation (no) and onsite solar generation (yes).

Forecast horizon	Case	$A(q)$	$C(q)$
1-h	No	$1 - 0.93q^{-1} - 0.06q^{-2}$	$1 - 0.21q^{-1} - 0.14q^{-2}$
	Yes	$1 - 1.8q^{-1} + 0.85q^{-2}$	$1 - 0.96q^{-1}$
15-min	No	$1 - 1.14q^{-1} + 0.15q^{-2}$	$1 - 0.35q^{-1} - 0.13q^{-2}$
	Yes	$1 - 0.95q^{-1}$	$1 - 0.03q^{-1} - 0.08q^{-2}$

Table 3
Statistical error metrics for the 1-h demand load forecasting with no onsite generation and with onsite solar generation.

Model	Without onsite generation					With onsite solar generation				
	MAE (kW)	MBE (kW)	RMSE (kW)	R^2	s	MAE (kW)	MBE (kW)	RMSE (kW)	R^2	s
P	38.18	11.21	49.66	0.92	0	117.79	15.05	178.62	0.73	0
SP1	28.45	−5.67	37.52	0.95	0.24	110.73	−6.75	166.29	0.77	0.07
SP2	28.93	−6.04	38.05	0.95	0.23	108.29	−9.17	161.35	0.78	0.09
kNN1	39.40	−17.81	56.26	0.90	−0.13	127.48	−33.25	177.68	0.73	0.01
kNN2	39.89	−18.31	57.13	0.89	−0.15	128.28	−33.37	179.30	0.73	−0.003
AR	31.02	−10.36	40.60	0.95	0.18	116.21	−19.75	158.34	0.79	0.11
ARMA	30.58	−9.41	40.02	0.95	0.19	116.18	−18.92	159.04	0.79	0.10

Table 4
Statistical error metrics for 15-min demand load forecasting without onsite generation and with onsite solar generation.

Model	Without onsite generation					With onsite solar generation				
	MAE (kW)	MBE (kW)	RMSE (kW)	R^2	s	MAE (kW)	MBE (kW)	RMSE (kW)	R^2	s
P	20.08	1.93	26.54	0.98	0	62.81	8.82	101.12	0.92	0
SP1	18.86	−0.11	24.71	0.98	0.06	60.58	2.28	98.24	0.92	0.03
SP2	18.96	−0.13	24.79	0.98	0.06	59.91	1.98	97.48	0.92	0.03
kNN1	28.84	−11.41	45.10	0.94	−0.69	68.23	−9.62	103.87	0.91	−0.02
kNN2	28.77	−11.14	44.64	0.94	−0.71	68.25	−9.16	103.58	0.91	−0.02
AR	18.97	−1.69	24.87	0.98	0.06	62.27	−0.33	96.16	0.93	0.04
ARMA	20.09	−0.92	26.18	0.98	0.01	61.58	−0.28	96.75	0.93	0.04

generation case, it is harder to beat the persistence model than in the previous case. The highest forecast skill was found to be 0.06 for SP1 and SP2. However, in terms of MAE, the AR model shows the best performance. From these results, it can be inferred that for the 15-min forecast with no onsite generation, there is very low variability and the assumption about persistence holds. For this reason P, SP1 and SP2 models perform equally well as compared to the other models.

For the onsite solar generation case, AR and ARMA models perform better than all the other models with $s = 0.04$. In this case, due to high solar penetration, the basic assumption behind the persistence model does not hold because of variability in solar power output for the 15-min time scale and the persistence model (P) does not perform as well as in the previous case.

From the various error metrics for the scenarios discussed above, it can be seen that AR models perform the best and exhibit the highest forecasting skill for all the studied cases. For the no onsite generation case, persistence models also perform well but the performance degrades substantially once the variability of the solar resource becomes a factor. As expected, it can be observed that as the variability increases, the performance of the models degrades. The same is true for the 1-h case with the only difference that the performance decrease is not as pronounced. However, in order to fully characterize the forecast error we need to study its distribution, which we do in the next section.

4.3. Error distribution

In order to analyze the forecast error distribution for the two cases, we started by computing the mean forecasting error *Eavg*, that is the average of the error for all the models. With this approach the resulting differences in *Eavg* depend solely on the variability of the time-series and the forecast horizon. Given that we have two cases and two forecast horizons, this operation results in four different error time-series. The mean (μ), standard deviation (σ), skewness (s) and kurtosis (k) for each case are tabulated in Table 5. As expected the σ, s and k values are much higher for the high solar penetration scenario.

Table 5
Statistics of average error distribution for no onsite generation and onsite solar generation case for 1-h and 15-min lead times.

Moment	$Eavg_{ng,1hr}$	$Eavg_{g,1hr}$	$Eavg_{ng,15min}$	$Eavg_{g,15min}$
μ	−7.05	−13.27	−2.94	−0.79
σ	33.88	137.12	23.04	83.58
s	−0.06	−0.46	−0.25	−0.38
k	4.92	6.01	3.66	9.15

Another way of studying the forecasting error distribution is to determine which statistical distribution model best describes it. The normal or Gaussian distribution is one of the most popular distributions to characterize the distribution of a random variable in terms of mean and standard deviation. It is not well-suited for distributions with heavy tails and high skewness which are often observed in forecast error distributions. Heavy tails are due to the presence of high magnitude errors in the forecast and skewness is caused by the forecast's bias. For these reasons we can anticipate that the normal distribution will not characterize the forecasting error distribution properly for onsite solar generation case.

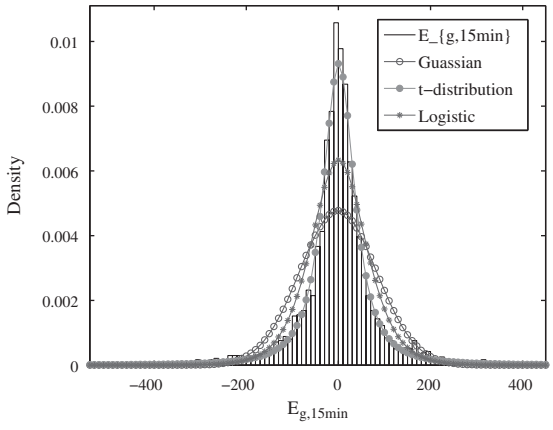

Fig. 3. Error distribution for 15-min load forecasting with solar penetration. The bins for the dataset are defined using Freedman–Diaconis rule. With high solar penetration the error distribution is better characterized by *t*-distribution.

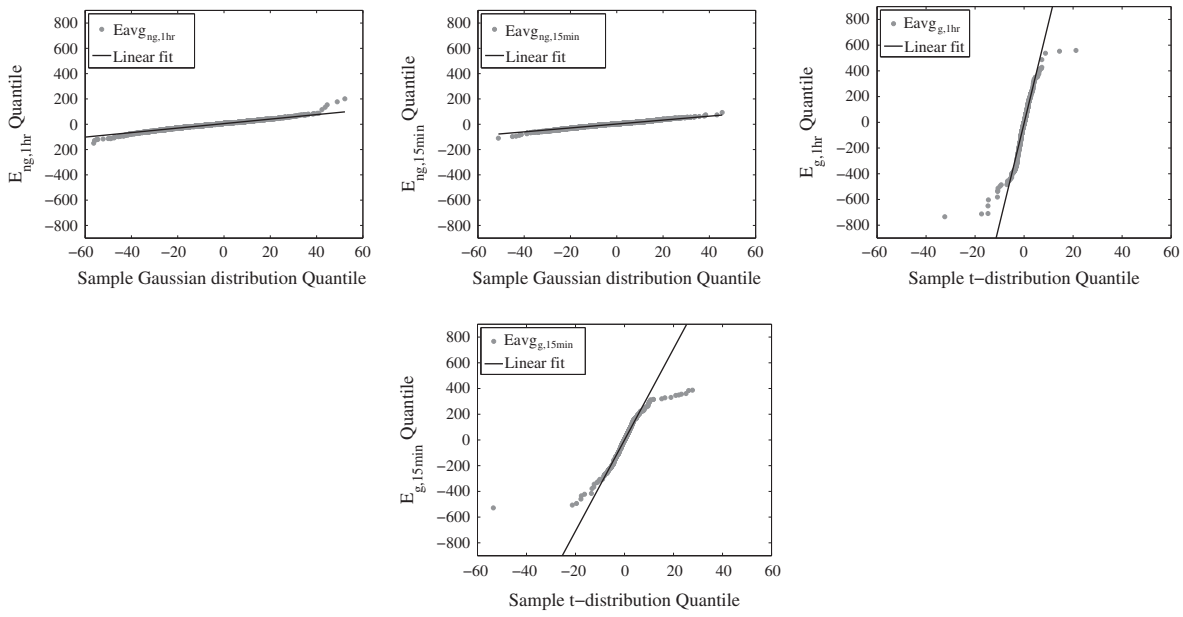

Fig. 4. Error distribution for 1-h and 15-min load forecasting with and without solar penetration. With no solar penetration the error distribution can be perfectly characterized by the Gaussian distribution as shown in the first and second plot. Whereas with solar penetration the error distribution can be better characterized by *t*-distribution. Third and fourth plot shows the sample *t*-distribution Q–Q plots for 1-h and 15-min cases with $\upsilon = 2.6$ and $\upsilon = 1.8$.

In order to identify the best matching distribution, various distributions were compared with the error distribution as shown in Fig. 3. We used a Q–Q plot (where Q stands for "Quantile") [34,40] to analyze the distribution fit more accurately. In the Q–Q plots, scattered points originated from two distributions are plotted against each other. If the two distributions are linearly related they will lie on the same straight line.

Fig. 4 shows the Q–Q plots of the error distribution for the 2 datasets and the 2 forecasting horizons. The sample Gaussian quantile for each error distribution is computed using μ and σ of the *Eavg* for each case, and Degrees of Freedom (υ) for the sample *t*-distribution quantile were computed based on the Maximum Likelihood principle.

These plots allow us to observe that with no onsite generation, the error distribution quantile has a linear relationship with a sample Gaussian quantile for both forecast horizon, which means that the error distribution can be characterized by a normal distribution. However, the same does not hold for onsite solar generation case. As shown in Fig. 3, *t*-distribution captures the tails and the peak of the distribution for high solar penetration much better than the Gaussian and the Logistic distribution. This observation is validated by the Q–Q plots in Fig. 4. These findings are in agreement with the results published for wind error forecasting distribution [33–35,40].

From this analysis we can observe the effect of the solar power on the distribution of the forecasting error. This effect can also be explained by analyzing the Cumulative Distribution Function (CDF) of normalized 15-min step change in load with (V_{Lg}) and without (V_{Lng}) onsite solar generation as shown in Fig. 5. It can be observed that the probability for $V_{Lng} > 100$ kW/15 min is 0.03 whereas in case of V_{Lg} it is 0.2. These large step changes are more challenging to forecast resulting in heavy tails in error distribution of load demand with solar generation.

4.4. Effect of solar penetration on the forecast error

As stated above, the main goal of this work is to explain the impact on the load forecast accuracy when we include highly variable solar resource in the energy budget. This is very important because

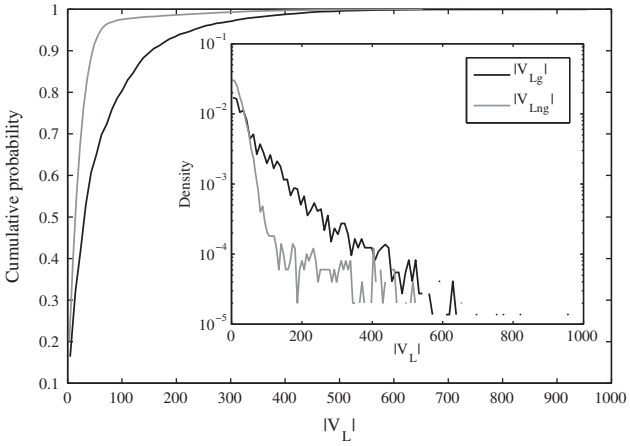

Fig. 5. Cumulative Distribution Function of absolute value of step change in load demand with (V_{Lg}) and without (V_{Lng}) onsite solar generation. The data is divided into 100 bins. The probability of step change in load for more than 100 kW/15 min is 0.03 for no generation case and its 0.2 for onsite generation case. The inset plots shows the Probability Density Function of the absolute value of step changes in load. It can be noticed that the PDF of step changes in load with onsite generation is wider than the PDF of step changes in load with no site generation which implies that the probability of larger step changes in L_g is higher than L_{ng}. Due to these higher step changes the error in load forecast is larger for onsite solar generation system.

the solar variability is always mentioned as one of the major challenges in promoting higher levels of solar penetration [2]. For this purpose we defined daily solar penetration ($S_{PV}(d)$) and solar power variability ($V_{PV,N}$) as

$$S_{PV}(d)(\%) = \frac{\sum_{t\in DT} L_d(t)}{\sum_{t\in DT} L_{PV}(t)}(\%) \tag{15}$$

where *t* is any time instant and it belongs to *DT* which is a set of all the daytime hours when solar power is produced for a given day (*d*).

$$V_{PV,N} = \sqrt{\frac{1}{N}\sum \left(\Delta L_{PV}(t) - \overline{\Delta L_{PV}(t)}\right)^2} \tag{16}$$

where $N = 4$ for the 1 h forecast and $N = 1$ for the 15-min forecast. $L_{PV}(t)$ is net solar power produced at time instant t and $\Delta L_{PV}(t)$ is the difference between $L_{PV}(t)$ and $L_{PV}(t-1)$ and $\Delta L_{PV}(t)$ is the mean of the difference.

The net error, i.e., the absolute difference between the load forecast error with and without onsite solar generation can be used to assess the effect of solar penetration. Fig. 6 shows the distribution of net error as a function of increasing solar penetration with different levels of solar variability for the 15 min horizon. The solar power variability for 1 h horizon varies between 0.01 kW/h to 399.58 kW/h and for 15 min horizon it varies from 0 kW/15 min to 800 kW/15 min. The smaller range for the 1 h horizon is due to the fact that as the time horizon increases solar variability reduces due the averaging. For the 1-h onsite solar generation case the average error values vary between −153.65 kW and 200.11 kW and for 15 min they vary between −109.35 kW and 91.92 kW. The daily solar penetration varies between 3% and 54%. The variability is also different for each day and its daily range depends on the daily penetration. For low (<10%) and high (>48%) penetration levels, solar variability is small. This is expected, given that these penetration levels are associated with overcast or clear days where fast ramp rates on the PV power are rare. For medium values of daily penetration, large ramping events lead to high variability. The dashed line in Fig. 6 shows the maximum magnitude of solar

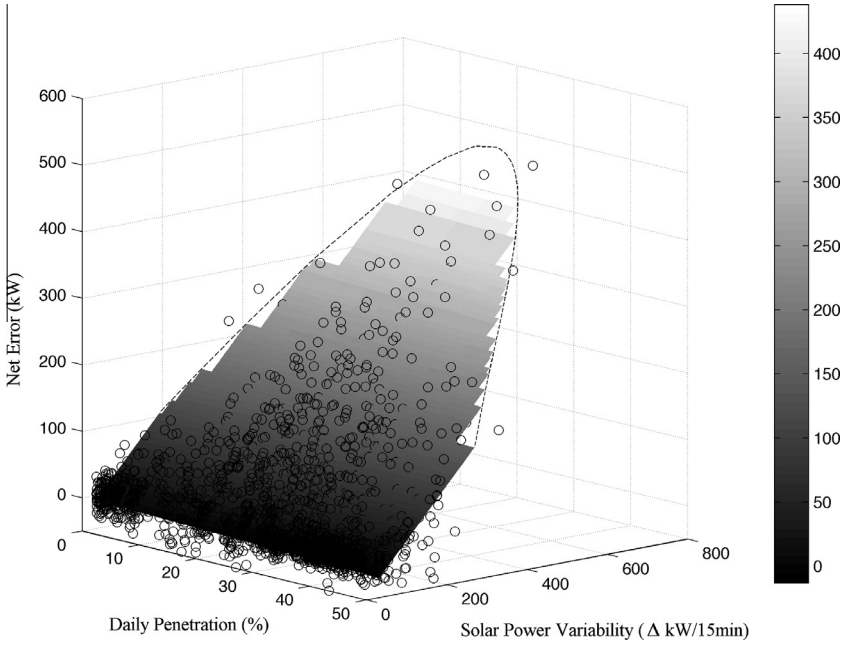

Fig. 6. Error distribution as a function of penetration with different levels of variability. The dashed line shows the maximum variability that can happen for a given daily penetration. The black markers shows the actual error for 15 min forecast and the surface plot shows the computed surface fit between the error, daily solar penetration and solar power variability for 15 min time steps. It shows that the error is directly related to solar variability. High magnitude error occurs during medium penetration level (20–40%) which is caused by big ramps events due to clouds.

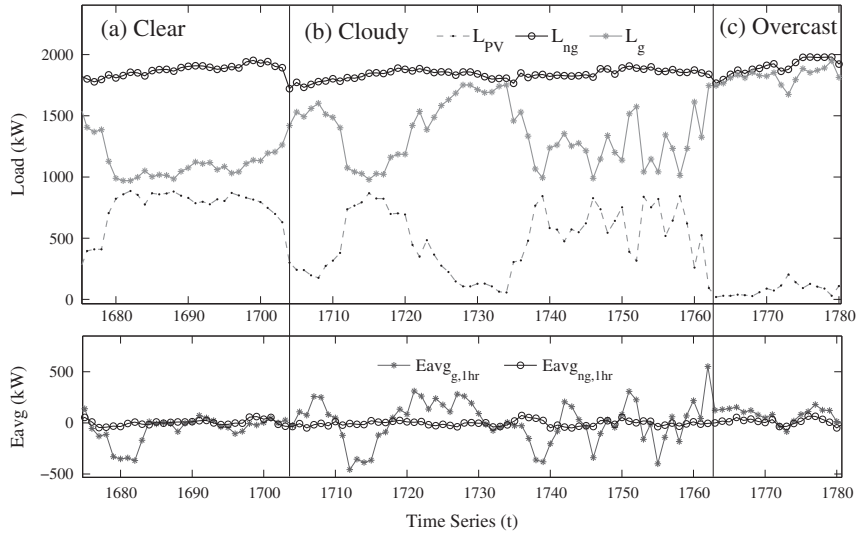

Fig. 7. Time-series of load and error profiles for a clear (a), cloudy (b) and overcast (c) day for 1 h forecast case. For sunny and overcast day, solar variability is the lowest and hence, the error is low too for both high and low solar penetration levels. Whereas on a cloudy day, penetration is medium but error is very high due to variability in solar power output.

variability that can occur for the given daily penetration. The polynomial fit for the dashed line was computed using all the solar power output data for the year 2010. In addition Fig. 6 shows that error magnitude increases with increasing solar variability but not with increasing solar penetration.

Finally, Fig. 7 shows the campus load time-series of load and forecast error profiles for clear, cloudy and overcast days. Solar variability is lowest for sunny and overcast days, resulting in low forecast error regardless of solar penetration levels. On the cloudy day, penetration is medium and the forecast error is very high due to the variability in solar power output. This analysis demonstrates again, that it is the solar variability that affects the forecast error magnitude rather than the penetration level. In [41] it was shown that variability in PV is directly proportional to variability in Global Horizontal Irradiance (GHI). Combining the result from [41] with the present observation, it can be said that the error distribution for a given time horizon can be characterized directly from the variability in solar irradiance for a given time horizon for that particular site.

5. Conclusions

We analyzed the performance of STLF models with and without onsite generation. For the 1-h forecast horizon with onsite solar generation, forecasting skill for the best model (ARMA) is reduced by 9% and R^2 decreases from 0.95 to 0.79 and for 15-min ahead case the forecast skill is reduced by 3% and R^2 decreases from 0.98 to 0.93. As expected, forecast accuracy decreases as forecast horizon and onsite solar generation increase. Similar conclusions have been made for high wind penetration cases [36]. For no onsite generation case, the assumption of persistence holds for the 15-min horizon and the SP models perform as well as AR and ARMA models with $R^2 = 0.98$. Comparing all the models, the AR model performs better in terms of forecast skill. We also showed that the forecast error distribution with onsite solar generation is best characterized by a *t*-distribution, which reflects the heavy tails in the error distribution due to high ramping events in the solar power output.

The impact of solar penetration level on the forecast error was studied for different levels of solar power variability and penetration level. For low variability levels, the error magnitude is mostly uniform and invariant with solar penetration level. As the solar power variability increases the error magnitude increases proportionally as a function of solar penetration, which shows that the solar variability has a more pronounced effect on the error magnitude than the level of penetration. All the results and analyses presented in this work hold for any type of onsite solar generation if it is contributing significantly to the power demand without any storage. Therefore, for long term planning and energy management of sites with expected onsite solar generation both solar penetration level and ground solar irradiance variability of the site should be taken into account.

Acknowledgments

The PI of this project (CFMC) gratefully acknowledges the partial financial support given by the California Energy Commission (CEC) under the PIER RESCO Project PIR-07-036, which is managed by Mr. Hassan Mohammed; and by the National Science Foundation (NSF) EECS (EPAS) award N. 1201986, managed by Dr. Paul Werbos. Partial support by the California Public Utilities Commission through the California Solar Initiative is also gratefully acknowledged.

References

[1] California Public Utilities Commission. Renewables portfolio standard, quarterly report, 1st quarter 2011, 2011.
[2] Lew D, Miller N, Clark K, Jordan G, Gao Z. Impact of high solar penetration in the western interconnection. Technical report NREL/TP-5500-49667, NREL; December 2010.
[3] Denholm P, Margolis RM. Evaluating the limits of solar photovoltaics(pv) in electric power systems utilizing the energy storage and other enabling technologies. Energy Policy 2007;35:4424–33.
[4] WECC Standard BAL-STD-002-0 operating reserves; 2010.
[5] Price PN. Methods for analyzing electric load shape and its variability. Report LBNL-3713E, LBNL, CEC; August 2010.
[6] Lew D, Brinkman G, Ibanez E, Hummon M, Hodge BM, Heaney M. Sub-hourly impacts of high solar penetrations in the western united states. Technical report NREL/CP-5500-56171, NREL; September 2012.
[7] Denholm P, Hand M. Grid flexibility and storage required to achieve very high penetration of variable renewable electricity. Energy Policy 2011;39(3):1817–30.
[8] Gross G, Galiana FD. Short-term load forecasting. Proc IEEE 1987;75(12):1558–73.
[9] Alfares HK, Nazeeruddin M. Electric load forecasting: literature survey and classification of methods. Int J Syst Sci 2002;33(1):23–34.
[10] Hahn H, Meyer-Nieberg S, Pickl S. Electric load forecasting methods: tools for decision making. Eur J Oper Res 2009;199(3):902–7.
[11] Willis H, Northcotegreen J. Spatial electric-load forecasting – a tutorial review. Proc IEEE 1983;71:232–53.
[12] Oamek GE, English BC. A review of load forecasting methodologies. Center for Agricultural and Rural Development, Iowa state University, 84-WP-5; June 1984.
[13] Moghram I, Rahman S. Analysis and evaluation of five short-term load forecasting techniques. IEEE Trans Power Syst 1989;4(4):1484–91.
[14] Abu-El-Magd MA, Sinha NK. Short-term load demand modeling and forecasting: a review. IEEE Trans Syst Man Cybernet 1982;12(3):370–82.
[15] Taylor JW, McSharry PE. Short-term load forecasting methods: an evaluation based on European data. IEEE Trans Power Syst 2008;22:2213–9.
[16] Hippert HS, Pedreira CE, Souza RC. Neural networks for short-term load forecasting: a review and evaluation. IEEE Trans Power Syst 2001;16(1):44–55.
[17] Metaxiotis K, Kagiannas A, Askounis D, Psarras J. Artificial intelligence in short term electric load forecasting: a state-of-the-art survey for the researcher. Energy Convers Manage 2003;44(9):1525–34.
[18] Suganthi L, Samuel AA. Energy models for demand forecasting: a review. Renew Sustain Energy Rev 2012;16(2):1223–40.
[19] Chakhchoukh Y, Panciatici P, Mili L. Electric load forecasting based on statistical robust methods. IEEE Trans Power Syst 2011;26(3):982–91.
[20] Garcia-Ascanio C, Mate C. Electric power demand forecasting using interval time series: a comparison between VAR and iMLP. Energy Policy 2010;38(2):715–25.
[21] Lopez M, Valero S, Senabre C, Aparicio J, Gabaldon A. Application of SOM neural networks to short-term load forecasting: the Spanish electricity market case study. Electric Power Syst Res 2012;91(0):18–27.
[22] Hong WC. Electric load forecasting by seasonal recurrent SVR (support vector regression) with chaotic artificial bee colony algorithm. Energy 2011;36(9):5568–78.
[23] Yang X, Yuan J, Yuan J, Mao H. An improved wm method based on PSO for electric load forecasting. Expert Syst Appl 2010;37(12):8036–41.
[24] Xia C, Wang J, McMenemy K. Short, medium and long term load forecasting model and virtual load forecaster based on radial basis function neural networks. Int J Electr Power Energy Syst 2010;32(7):743–50.
[25] De Felice M, Xin Y. Short-term load forecasting with neural network ensembles: a comparative study [application notes]. IEEE Comput Intell Mag 2011;6(3):47–56.
[26] Elattar EE, Goulermas JY, Wu QH. Generalized locally weighted GMDH for short term load forecasting. IEEE Trans Syst Man Cybernet Part C: Appl Rev 2012;42(3):345–56.
[27] Taylor JW. Short-term load forecasting with exponentially weighted methods. IEEE Trans Power Syst 2012;27(1):458–64.
[28] Mamlook R, Badran O, Abdulhadi E. A fuzzy inference model for short-term load forecasting. Energy Policy 2009;37(4):1239–48.
[29] Rejc M, Pantos M. Short-term transmission-loss forecast for the Slovenian transmission power system based on a fuzzy-logic decision approach. IEEE Trans Power Syst 2011;26(3):1511–21.
[30] Alamaniotis M, Ikonomopoulos A, Tsoukalas LH. Evolutionary multiobjective optimization of kernel-based very-short-term load forecasting. IEEE Trans Power Syst 2012;27(3):1477–84.
[31] Kebriaei HH, Araabi BN, Rahimi-Kian A. Short-term load forecasting with a new nonsymmetric penalty function. IEEE Trans Power Syst 2011;26(4):1817–25.
[32] Pedro HTC, Coimbra CFM. Assessment of forecasting techniques for solar power production with no exogenous inputs. Solar Energy 2012;86(7):2017–28.
[33] Bludszuweit H, Dominguez-Navarro JA, Llombart A. Statistical analysis of wind power forecast error. IEEE Trans Power Syst 2008;23(08858950):983–91.
[34] Hodge BM, Milligan M. Wind power forecasting error distributions over multiple timescales. IEEE 2011(1944–9925):1–8.

[35] Lange M. On the uncertainty of wind power predictions-analysis of the forecast accuracy and statistical distribution of errors. J Solar Energy Eng 2005;127:177–84.
[36] GE Energy. Analysis of wind generation impact on ERCOT ancillary services requirements. Report, Electric Reliability Council of Texas; March 2008.
[37] Ljung L. System identification: theory for the user. Number ISBN 0-13-881640-9 025. Prentice Hall PTR, New Jersey, USA; 1999.
[38] Duda RO, Hart PE, Stork DG. Pattern classification. John Wiley & Sons, Inc.; 2000.
[39] Marquez R, Coimbra CFM. A proposed metric for evaluation of solar forecasting models. J Solar Eng 2013;135.
[40] Hodge BM, Lew D, Milligan M. Short-term load forecasting error distributions and implications for renewable integration studies. Technical report NREL/CP-5500-57340, NREL; January 2013.
[41] Marquez R, Davis B, Kaur A, Coimbra CFM. Characterization and cost analysis of 1 mw PV plant variability integrated with a small campus load. In: ESFuelCell2012-91102. ASME 2012 6th international conference on energy sustainability; July 2012.