Title

Net load forecasts for solar-integrated operational grid feeders

Authors

Chu, Yinghao
Pedro, Hugo TC
Kaur, Amanpreet
et al.

Publication Date

2017-12-01

DOI

10.1016/j.solener.2017.09.052

Peer reviewed

Intra-Hour Net Load Forecasts for Solar-Integrated Operational Grid Feeders

Yinghao Chu, Hugo T. C. Pedro, Kaur Amanpreet, Jan Kleissl, Carlos F. M. Coimbra*

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

Abstract

Net load forecasts for multiple feeders in the San Diego Gas & Electric operating region have been produced for intra-hour forecast horizons. The forecasting methods implemented in this work include Artificial Neural Networks (ANN), Support Vector Regression (SVR), sky-imaging techniques, and other time series methods. Three enhancement methods are implemented to further decrease forecasting errors: (1) decomposing the time series of the net load to remove the daily trend; (2) training two models with daytime and nighttime data and applying them for day period and night period, respectively; and (3) introducing sky image features as exogenous inputs for 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 improvement over reference persistence model. The assessment results show that stochastic-learning models with the enhancement methods significantly outperform the reference persistence model and achieve forecast skills up to 43% depending on the location and forecast horizons.

Keywords: Net Load forecasts; Sky imaging; Support vector machines; Artificial neural networks; Solar integration.

1. Introduction

Uncertainties in electric loads need to be compensated using operating reserves or ancillary generation, which increase the overall costs for utilities, customers, system operators, and other market participants [1]. Short-term load forecasts play a key role in mitigating the uncertainty of loads and are essential to decrease the costs of operation, control, management, balancing and scheduling of the electric grid [2]. The earliest studies of electrical demand/load forecasts date back to 1960 [3]. Since then various load forecasting techniques, based on different methods such as time series analysis and regression [4, 5], stochastic-learning and artificial intelligence [6, 7], hybrid or ensemble models [2, 8–10], have been developed and comprehensively reviewed in the literature. Detailed reviews of recently developed methods can be found in [11].

Despite all these works in previous decades, there has been a renewed interest in forecasting electricity net demand because of the distributed solar generation, which has been growing rapidly due to the net energy metering tariffs and other incentives [12]. The variable rooftop solar generation adds substantial uncertainties in power demand at the substation level [13] and changes the variability of net load time series. The impact of solar penetration on the load profile for the feeder has been discussed previously in literature [12]. In general, for operating regions with higher solar penetration levels, the variability in solar power production propagates into the load profile and increases the relative error of net load forecasts during the daytime.

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

*Corresponding author
Email address: ccoimbra@ucsd.edu (Carlos F. M. Coimbra)

are developed using data from operational grid feeders and are implemented for 10-, 20-, and 30-minute forecast horizons. The ANN/SVR models are further enhanced using three proposed methods: time series detrending, daytime/nighttime forecasting, and using sky image features as exogenous inputs. Performance of the proposed models is assessed in terms of common statistical metrics and compared against a reference persistence model. The major contribution of this work is to develop appropriate forecasting models for net loads of multiple operational grid feeders, which have significant solar penetration levels. Another contribution is to evaluate the effectiveness of three enhancement methods based on data collected from operational feeders. The proposed forecasting models and the recommended enhancement methods are expected to improve the accuracy and robustness of the net load forecasts for public utilities and system operators.

The rest of this paper is organized as followed: the data used for this study is presented in Section 2. The models and enhancement methods are presented in Section 3. Results and discussion are presented in Section 4, where the forecast error distributions are presented and characterized to understand the performance of the proposed models. The conclusions of this work are presented in Section 5.

2. Data

Intra-hour forecasts (up to 30 minutes) are developed and implemented for four feeders: Alpine, Cabrillo, Avocado, and Valley Center located in the operating region of the San Diego Gas & Electric (SDG&E). The operating regions of these four feeders have different solar penetration levels, which are presented in Table 1. Solar penetration is defined as the annual solar power produced divided by the annual total load on the feeder and the total load is calculated as the solar power produced plus the net load.

Table 1: Details about four SDG&E feeders for which the net load forecasting is implemented and tested.

SDGE feeders	Location	Feeder length [km]	No. of customers	No. of PV systems	Solar penetration
Alpine	32.83 N, 116.77 W	34.5	1466	28	2.39%
Cabrillo	32.67 N, 117.24 W	39.6	3761	91	5.79%
Avocado	33.39 N, 117.26 W	177.8	2246	29	13.30%
Valley Center	33.24 N, 117.00 W	51.5	471	19	23.80%

Load data with 10-minute intervals are collected from Oct. 2014 to Mar. 2015 (22550 time instances) from a dedicated Pi System (OSIsoft) and 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. Two UCSD Sky Imagers (USI)[14] are installed near the Alpine and Cabrillo feeders and provide sky-cover information. The sky images are processed to obtain numerical image features, which are used as exogenous inputs to the intra-hour forecasts. Load forecasts for feeders without sky imagers are developed based on endogenous inputs.

3. Methods

3.1. Stochastic-learning models

Artificial Neural Networks (ANN) and Support Vector Regression (SVR) are popular stochastic-learning tools for pattern recognition, data classification and regression, and have proven to be useful for non-linear input/output mapping [15, 16]. Therefore, both ANN and SVR are used in this work to implement the net load forecasts. The weights and parameters of both ANN and SVR are estimated using the training dataset.

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 take in weighted sum of inputs X_j

through various layers and produce an output using an activation function. The ANNs used in this work are feedforward networks, that is, only forward connections between the neurons are allowed. Mathematically, the ANNs can be represented as:

$$Y_i = f\left(\sum_{j=1}^{M}(w_{ij}X_j)\right), \tag{1}$$

where ~~the~~ Y_i is the output from the i-th neuron, f is an activation function (sigmoidal functions in this work), w_{ij} are the weights of the j-th input on the i-th neuron, M is the number of inputs, and X_j is j-th output from previous layer. A supervised learning process using the training data is employed to optimize the weight w_{ij} and bias β_{ij}. In this work, we use the Bayesian regularization process with Levenberg-Marquardt optimization [17]. After the learning process, the ANN model is used to produce~~d~~ the forecast using as inputs the independent data from the training dataset. More details about applications and implementations of ANN can be found in [18–22].

The SVR modeling uses inputs that are known as support vectors and can be mathematically expressed as

$$f(x) = < \boldsymbol{w} \cdot \boldsymbol{x} > + \boldsymbol{b}, \tag{2}$$

where $\boldsymbol{w}$ represents the weights and $\boldsymbol{b}$ is a bias term. The optimization problem to be solved is defined with the following objective:

$$F(w, \varepsilon) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\varepsilon_i, \tag{3}$$

where C determines the penalty assigned to forecast error ε_i. SVR is trained using labeled data to minimize Eq. (3) subjected to the constraints:

$$\begin{cases} y_i - (w \cdot f(x_i) + b) \leq \xi_i \\ (w \cdot f(x_i) + b) - y_i \leq \xi_i \end{cases}. \tag{4}$$

where f is the mapping function (radial basis function in this work) that controls the regression quality, and ξ_i depends on ε_i and a threshold e:

$$\xi_i = \begin{cases} 0 & \text{if} \quad |\varepsilon_i| \leq e \\ |\varepsilon_i| - e & \text{otherwise} \end{cases}. \tag{5}$$

More details about applications and implementations of SVR can be found in [16, 23–25].

3.2. Enhancement methods

To further enhance the accuracy of stochastic-learning models, three enhancement methods are employed and tested: time series detrending, daytime/nighttime forecasting, and augmenting the input variables with sky image features.

3.2.1. Detrending

Detrending process decomposes the time series of the net load to remove any daily trends. Decomposing the load time series and removing the daily trend is expected to enhance the performance of load forecasts [2, 12]. The detrended electric loads can be mathematically expressed as:

$$L(t) = L_{dt}(t) \times \mu(t), \tag{6}$$

where $L(t)$ is the actual load at time t, $L_{dt}(t)$ is the corresponding detrended load, and $\mu(t)$ is the daily trend. The values of the daily trend are calculated as the mean of the daily load profile using the training dataset (shown in Fig. 1). The example time series of original and detrended load $L_{dt}(t)$ is shown in Fig. 1.b and c, respectively. The detrended time series is used to train and optimize the forecasting models, but all the performance analysis in the results section is done with the actual (non-detrended) data.

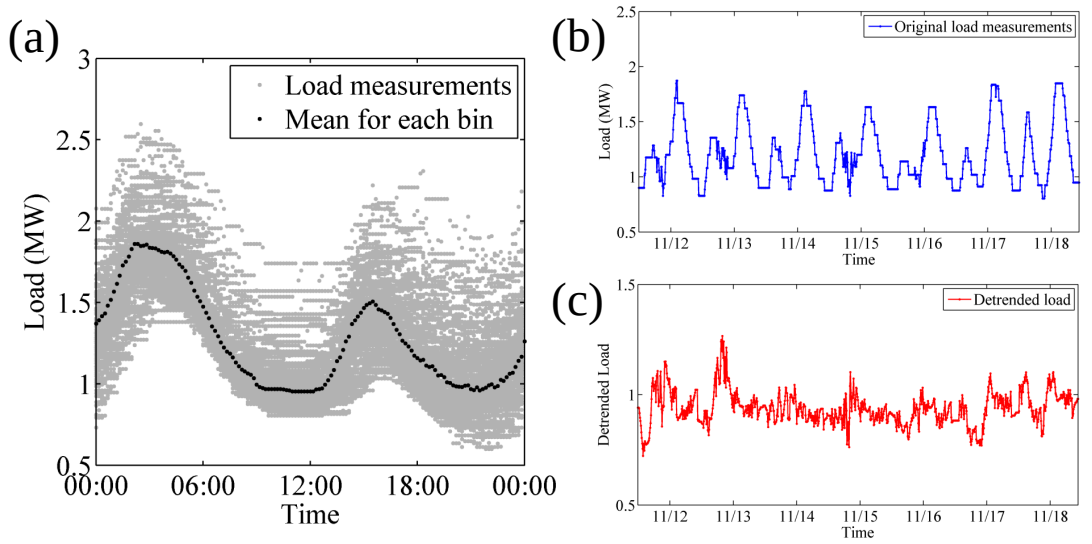

Figure 1: (a) Daily load profile and trend fit for the Alpine feeder. (b) Example time series of 7-day actual load for the Alpine feeder and (c) example time series of 7-day detrended load for the Alpine feeder.

3.2.2. Daytime/nighttime forecasts

The load profile for daytime is different from that of night time because of many reasons such as diurnal human activities and solar power productions. The accuracy of stochastic learning models is highly dependent on training data. For example, an ANN/SVR model trained using the night data may be suboptimal to predict the load for daytime or vice versa [17]. Thus, daytime/nighttime forecasts are achieved by training two models with daytime and nighttime data separately. This strategy can be summarized in 3 steps: (1) separating the training dataset into a day-time set and a night-time set; (2) training two models using the day-time set and the night-time set, respectively; and (3) adaptively applying the two forecasting models in real-time forecasts.

3.2.3. Sky image features as exogenous inputs

Integration of solar generation changes the profile of day-time net load. As the level of solar penetration increases, the variability of day-time net load increases due to the variability of solar energy [12]. The impact of the variability in solar power can be observed in net load (shown in Fig 1.b). The additional variability increases the difficulty in obtaining accurate load forecasts. Sky-cover information provided by local-sensing techniques has been successfully used for intra-hour solar forecast applications [26]. Therefore, sky-cover information is potentially useful exogenous input to improve the forecasting accuracy of solar-integrated net load. In this work, an efficient sky image-processing algorithm, which has been successfully used in intra-hour solar forecast applications [27, 28], is employed to translate the sky images into numerical image features, which in turn are used as exogenous inputs to the stochastic learning models (ANN and SVR).

The image features represent cloud cover information that is important to forecast short-term solar ramps as well as the net load ramps. The algorithm to extract the image features includes three major procedures: (1) image masks are created to exclude the ground obstacles (e.g. trees, buildings); (2) normalized red blue ratios (NRBRs) are calculated for each pixel in the unmasked region of the sky image; (3) image features (mean, standard deviation, entropy) for a sky image are statistically derived based on the NRBRs of all its unmasked pixels. Examples of the image processing are illustrated in figure 2 for a sky camera located in Folsom, California. This algorithm is very fast (less than 2 second for each sky image) and is applicable to images captured by the USI systems deployed in Cabrillo and Alpine. The database of sky image features using sky images collected during the period from Oct 28th 2014 to Mar 31th 2015 is generated. This database is used for both training and

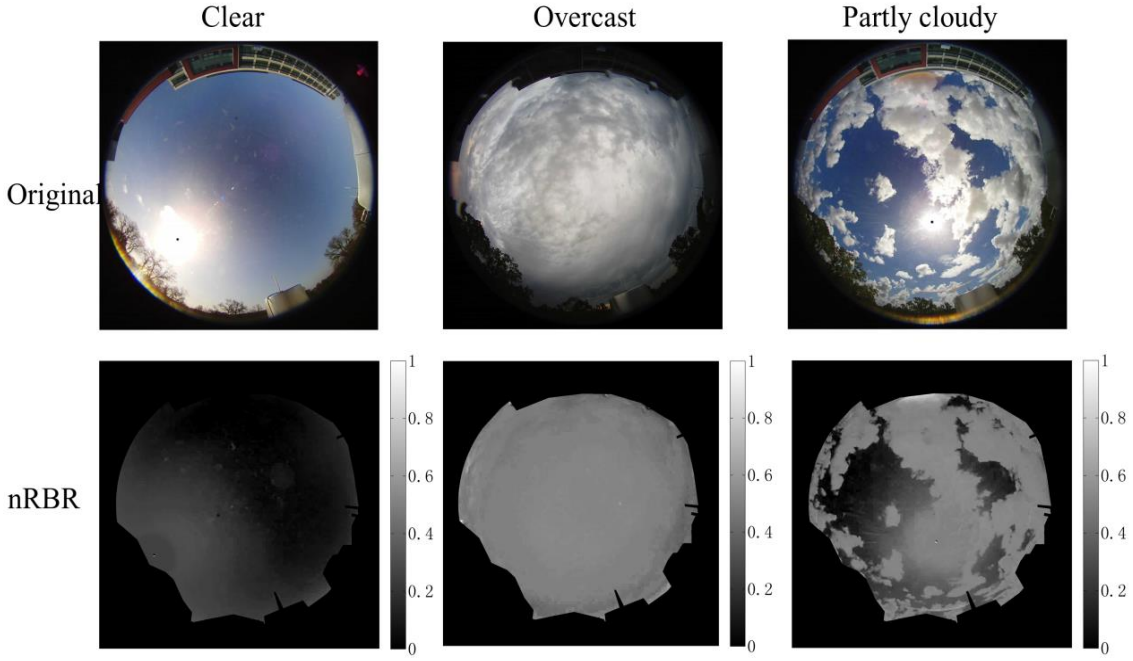

Figure 2: Examples of original images (top row) and normalized RBR (nRBR) images (bottom row). The grey-scale indicates the nRBR magnitudes in each image. These images are captured under different weather conditions by a sky camera located in Folsom, California [28].

validation of the stochastic learning models for the net load forecasting.

3.3. Error metrics

Four statistic error metrics are used in this work to assess the forecast performance: Mean Bias Error (MBE), Mean Absolute Percentage Error (MAPE), relative Root Mean Square Error (rRMSE), and forecast skill over persistence (s). MBE measures the bias of a model and is expressed as:

$$\mathrm{MBE} = \frac{1}{n}\sum_{t=1}^{n}\left(\hat{L}(t) - L(t)\right), \tag{7}$$

where $L(t)$ is the measured net load and $\hat{L}(t)$ is the predicted net load at time t. MAPE measures the accuracy of a model in terms of percentage error and is defined as:

$$\mathrm{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{\hat{L}(t) - L(t)}{L(t)}\right|, \tag{8}$$

rRMSE measures the relative spread in the error:

$$\mathrm{rRMSE} = \frac{\sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(\hat{L}(t) - L(t)\right)^2}}{\frac{1}{n}\sum_{t=1}^{n}|L(t)|}, \tag{9}$$

where the denominator $\frac{1}{n}\sum_{t=1}^{n}|L(t)|$ is the average value of the net load. Forecasting skill (s) measures the improvement of the proposed forecast model (the ANN or the SVR) over the reference model in terms of rRMSE:

$$\mathrm{s} = 1 - \frac{\mathrm{rRMSE}}{\mathrm{rRMSE_p}}, \tag{10}$$

where the subscript p represents the reference persistence model.

The persistence model, which is one of the simplest forecasting models, is based on the assumption that the current state of the system persists between the present time and the time of the forecast [29]. The persistence forecast is expressed as:

$$\hat{L}(t + FH) = L(t), \tag{11}$$

where FH is the forecast horizons that can be 10-, 20-, or 30-minute in this work.

4. Results and discussion

Intra-hour forecasts are implemented for 10-, 20-, and 30-minute forecast horizons with 10-minute resolution. In this section, baseline forecasts are developed using endogenous inputs without employing any of the three enhancement methods discussed in Section 3.2. The endogenous inputs are the lagged values of the net load ranging from zero to one hour in steps of 10 minutes. The baseline forecasts are used to evaluate the effectiveness of the proposed enhancement methods. Then the performance of the stochastic-learning forecasts, with or without enhancement methods, are evaluated and compared against the reference persistence model. Absolute error distributions and sample time series of investigated forecasts are also presented and discussed in this section.

4.1. Enhancement methods

The comparisons of the detrended forecasts and corresponded baseline forecasts in terms of MAPE and rRMSE are illustrated in Fig. 3. Compared to the baseline forecasts, the detrended forecasts mostly have lower MAPEs and rRMSEs. For persistence model, detrending reduces MAPE and rRMSE up to 40% for the Cabrillo feeder. For the ANN and SVR models, detrending decreases the error metrics up to 15% depending on locations and horizons. In general, forecasts that are trained using the detrended data show superior performance over the baseline forecasts for longer horizon forecasts, particularly for 30-minute horizon. This is expected because the variations in actual load time series are usually greater for longer time horizon due to the daily trend.

The comparisons of the daytime/nighttime forecasts and corresponded baseline forecasts in terms of MAPE and rRMSE are illustrated in Fig. 4. The predictions for the daytime/nighttime persistence model are the same as that for the baseline model (persistence model is the same for day and night periods). Therefore, the results of persistence forecasts are not plotted in Fig.4. The figure shows that daytime/nighttime forecasts based on either ANN or SVR do not consistently outperform the baseline forecasts. For the Alpine and Cabrillo feeders, the daytime/nighttime forecasts only achieves marginal improvements ($< 3\%$) over baseline forecasts. For Avocado and Valley center feeders, daytime/nighttime forecasts mostly have higher rRMSEs (up to 5%) than the baseline forecasts. Based on the results, the daytime/nighttime method is not considered as effective enhancement method.

The comparisons of the forecasts that use image features as exogenous inputs (Im-ex) and baseline endogenous forecasts in terms of MAPE and rRMSE are illustrated in Fig. 5. This figure do not consider night time values because image features are only available during the daytime. For the Alpine feeder, the rRMSEs of Im-ex forecasts are 0.03% to 1.5% (average 0.75%) lower than that of the baseline forecasts. For the Cabrillo feeder, the rRMSEs of Im-ex forecasts are -0.15% to 3.5% (average 1.5%) lower than that of the baseline forecasts. When Im-ex inputs are used, overall improvements in term of rRMSE for Cabrillo feeder are significantly higher than that for Alpine feeders. This is expected because the solar penetration for the Cabrillo feeder is nearly twice as large as that of the Alpine feeder. The overall solar penetration levels at both feeders are still very low (2.4% for Alpine feeder, and 5.8% for Cabrillo feeder) when compared to other locations. Therefore, Im-ex forecasts only achieve marginal improvements in terms of rRMSE for these two feeders. The enhancement of Im-ex inputs are expected to be more effective for feeders with higher level of solar penetration.

4.2. Net load forecasts

Based on the results of previous analysis, different enhancement strategies are employed for different locations to obtain 24-hr continuous load forecasts. Forecasts for Alpine and Cabrillo feeders are trained using detrended data and use image features (when available during daytime) as exogenous inputs. Forecasts for Avocado and Valley center, where sky imager is not available, use the detrending technique.

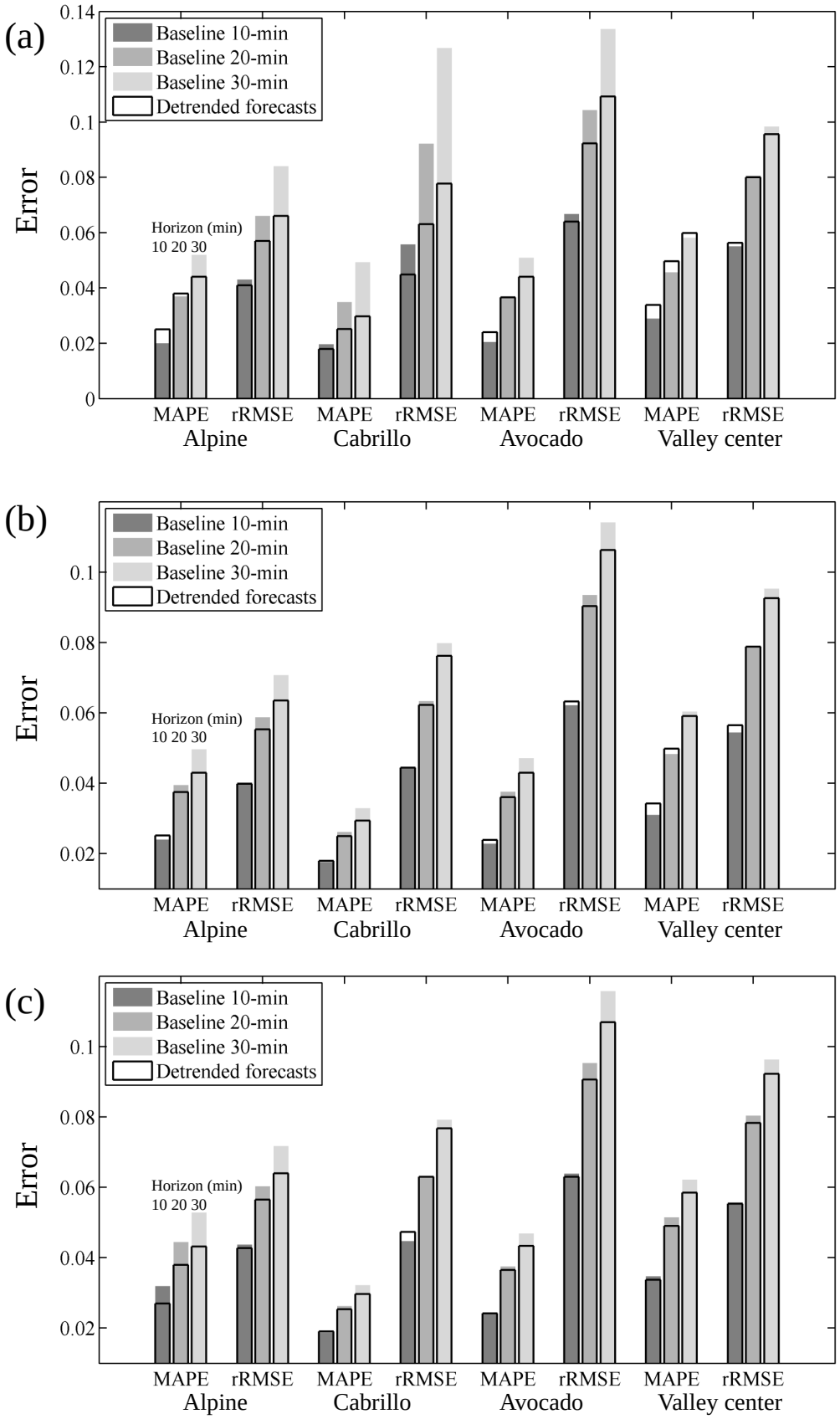

Figure 3: The MAPEs and rRMSEs of the baseline and the detrended forecasts for 10-, 20-, and 30- minute horizons. The evaluated models are (a) persistence model, (b) ANN model, and (c) SVR model. Both MAPE and rRMSE are dimensionless.

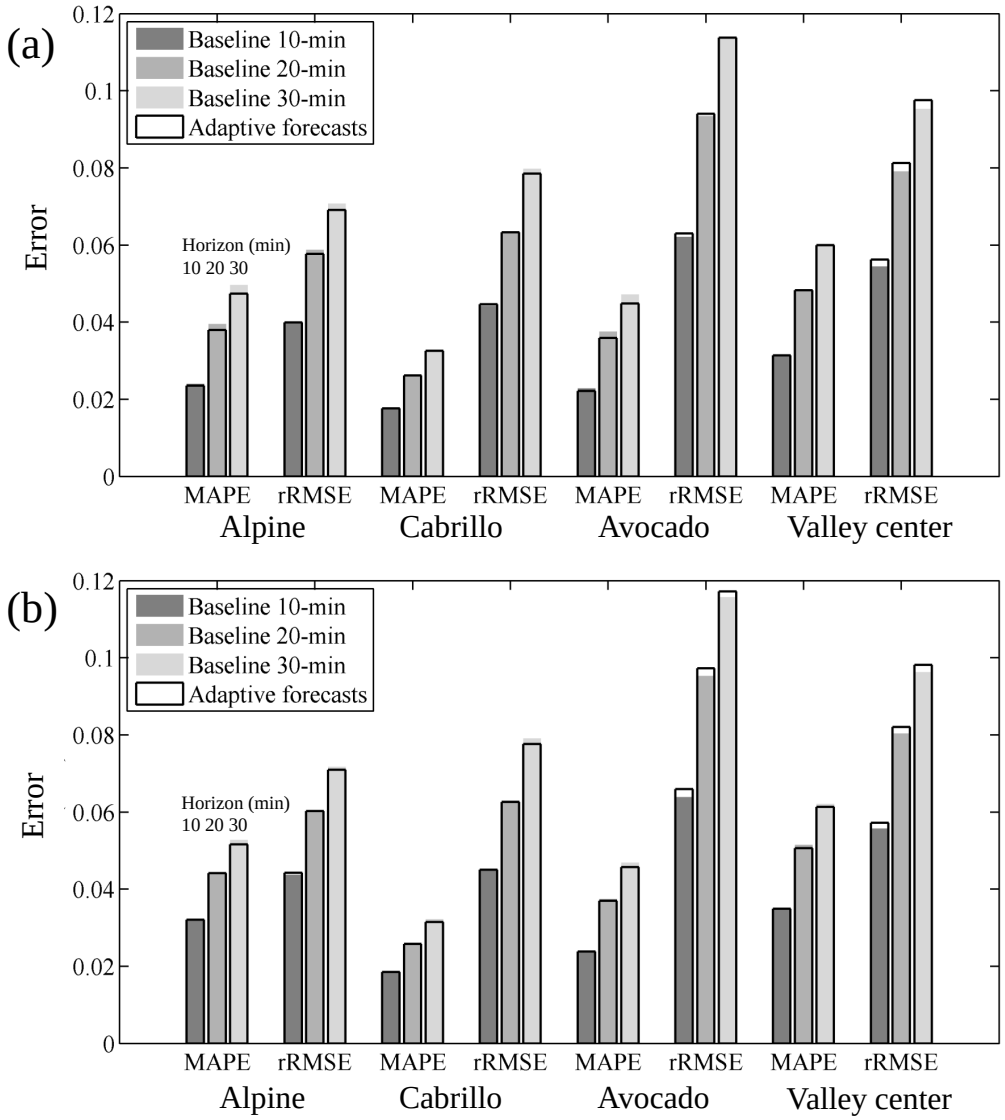

Figure 4: The MAPEs and rRMSEs of the baseline and the daytime/nighttime forecasts for 10-, 20-, and 30- minute horizons. The investigated models are (a) ANN model, and (b) SVR models. Both MAPE and rRMSE are dimensionless.

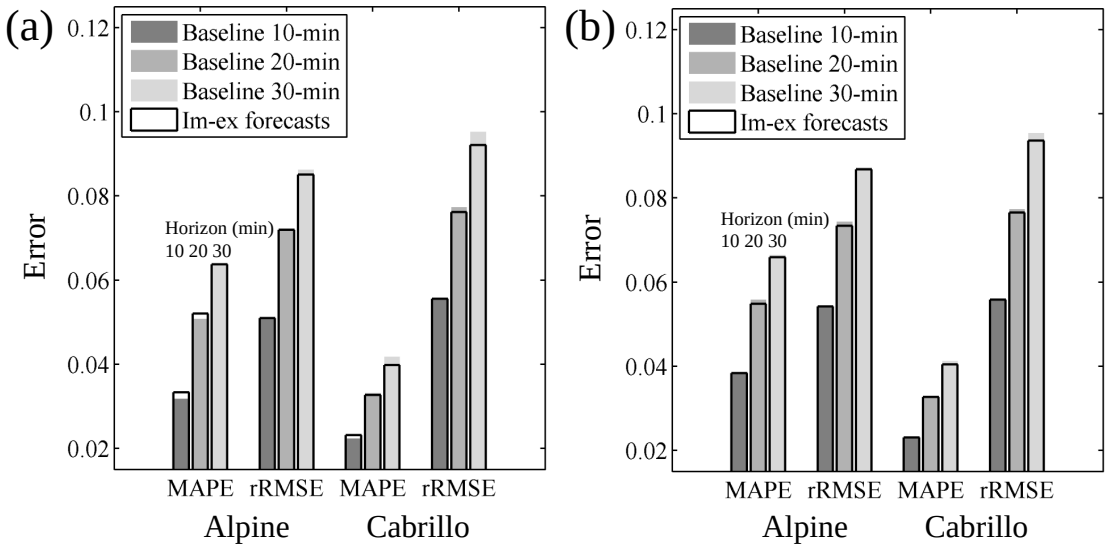

Figure 5: The MAPEs and rRMSEs of the baseline endogenous forecasts and the forecasts with image features as exogenous inputs for 10-, 20-, and 30-minute horizons. The investigated models are (a) ANN model, and (b) SVR models. Both MAPE and rRMSE are dimensionless.

The error metrics for the baseline and the enhanced forecasts models are presented in Table 2. The persistence model shows the smallest MBE in most cases. However, given that all models exhibit small MBE values we compare the performance of models in terms of MAPE and rRMSE. It can be observed that (1) the performance metrics of both stochastic-learning models significantly outperform the reference persistence forecasts; (2) the performance metrics of models with the enhancement methods significantly outperform the baseline forecasts; and (3) both ANN and SVR models achieve higher forecast skills for longer horizon forecasts. ANN model slightly outperforms SVR under most circumstances.

To further understand the forecast performances, both Probability Density Function (PDF) and Cumulative Density Function (CDF) of the distributions of the absolute forecast error are computed and plotted in Fig. 6 and 7. The persistence model is based on the assumption that the net load remains the same with time. Consequently, persistence errors equate to the magnitude of step changes in the net load and the PDF of the persistence errors represents the variability of time series [30] of the net load. For all locations, the persistence models at different error levels tend to have smaller CDF values when forecast horizon increases. This is expected because that the longer horizons usually have larger step changes in the load and are, therefore, more variable and more difficult to forecast [15].

The error distributions of ANN and SVR baseline forecasts share similar behaviors and show comparable performance for high error level ($\varepsilon > 0.2$ MW). The ANN CDFs are slightly greater than the SVR CDFs for the low and moderate error level ($\varepsilon < 0.2$ MW). Therefore, the ANN forecasts slightly outperform the SVR forecasts in terms of the error metrics and achieve the lowest RMSE and highest forecast skills (shown in the Table 2). Comparing to the reference persistence model, the stochastic-learning models (both baseline and enhanced models) significantly reduce the occurrences of moderate and high magnitude errors regardless of location and forecast horizons. Therefore, ANN and SVR PDFs decay to 0 faster than the persistence PDFs. Equivalently, ANN and SVR CDFs approach 1 significantly faster than persistence CDFs. These results indicate that the stochastic-learning forecasts significantly reduce the forecasting uncertainty in the net load for intra-hour forecast horizons. The enhanced forecasts have lower frequency of moderate and high errors than the baseline forecasts and have slightly higher cumulative probability for error level > 0.02 MW. As a result, the PDFs and CDFs of enhanced forecasts exhibit shorter tails than the baseline forecasts, particularly for longer horizon (30-minute).

Example time series of the enhanced forecasts for Cabrillo feeder and corresponding absolute error time series are shown in Fig. 8 for a period of 48 hours. The daytime forecasts consider image features as exogenous inputs while the nighttime forecasts uses only endogenous inputs. Cabrillo feeder has a solar penetration level of 5.79%, and the impact of solar variability on the net load time series can be seen in Fig. 8, particularly during the noon time when the solar power approximates daily maximum. The improvement achieved using the stochastic models with enhancement methods over the reference persistence model is noticeable and as illustrated by the forecast error time series. The highest improvements are observed during the midnight and early morning when the net load time series has a straight and smooth trend. During the mid of days when exogenous inputs are available, both ANN and SVR forecasts achieve significantly lower error than the persistence forecasts, particularly for 30-minute forecasts.

In this work, effects of the spatial differences in the area covered by the feeder are assumed to be negligible and the sky imagery from a single point in near the feeder is assumed to represent the entire area of the feeder. To capture the spatial variability in the net load and further improve the intra-hour forecast accuracy as well as resolution for feeders with higher level of solar penetrations, two kinds of exogenous inputs are recommended: (1) spatial distribution of local solar plants; (2) real-time map of spatial solar irradiance for the feeder area. The irradiance map can be estimated using sensor or image network.

Table 2: Net load forecasting results on the testing dataset for the four SDG&E fedders

			Baseline			Enhanced		
	Forecast Horizon		10	20	30	10	20	30
		MBE	0.000	0.000	0.000	-0.001	-0.001	-0.002
	per	MAPE	0.02	0.037	0.052	0.025	0.038	0.044
		rRMSE	0.043	0.066	0.084	0.041	0.057	0.066
		MBE	0.001	0.002	0.003	0.001	0.001	0.002
Alpine	**ANN**	MAPE	0.024	0.039	0.05	0.025	0.037	0.042
		rRMSE	0.04	0.059	0.071	0.04	0.056	0.064
		s	7.50%	11.60%	16.10%	7.10%	16.30%	24.60%
		MBE	0.003	0.004	0.005	-0.001	-0.002	-0.001
	SVR	MAPE	0.032	0.044	0.053	0.027	0.038	0.043
		rRMSE	0.044	0.06	0.072	0.046	0.061	0.068
		s	-0.60%	9.30%	14.90%	-6.40%	8.90%	19.20%
		MBE	0.000	0.000	0.000	-0.001	-0.003	-0.004
	per	MAPE	0.020	0.035	0.049	0.018	0.025	0.03
		rRMSE	0.056	0.092	0.127	0.046	0.064	0.078
		MBE	0.005	0.008	0.012	0.007	0.011	0.017
Cabrillo	**ANN**	MAPE	0.018	0.026	0.033	0.018	0.024	0.028
		rRMSE	0.044	0.063	0.08	0.044	0.059	0.073
		s	20.60%	31.30%	37.10%	20.10%	35.90%	42.70%
		MBE	0.004	0.007	0.01	0.003	0.017	0.019
	SVR	MAPE	0.019	0.026	0.032	0.02	0.026	0.029
		rRMSE	0.045	0.063	0.079	0.048	0.065	0.074
		s	19.70%	31.60%	37.60%	13.10%	29.50%	41.40%
		MBE	0.000	0.000	0.001	-0.001	-0.002	-0.002
	per	MAPE	0.022	0.037	0.051	0.024	0.037	0.044
		rRMSE	0.067	0.104	0.134	0.064	0.092	0.109
		MBE	-0.004	-0.01	-0.016	-0.004	-0.008	-0.01
Avocado	**ANN**	MAPE	0.021	0.038	0.047	0.024	0.036	0.043
		rRMSE	0.062	0.093	0.114	0.063	0.09	0.106
		s	6.90%	10.40%	14.60%	5.30%	13.40%	20.50%
		MBE	-0.009	-0.007	-0.007	-0.006	-0.017	-0.019
	SVR	MAPE	0.024	0.037	0.047	0.024	0.036	0.043
		rRMSE	0.064	0.095	0.116	0.063	0.091	0.107
		s	4.40%	8.70%	13.40%	5.70%	13.10%	20.00%
		MBE	0.000	0.000	0.000	0.000	0.000	0.001
	per	MAPE	0.029	0.046	0.058	0.034	0.050	0.060
		rRMSE	0.055	0.081	0.098	0.056	0.080	0.096
		MBE	0.000	-0.001	-0.002	-0.001	-0.002	-0.002
Valley Center	**ANN**	MAPE	0.031	0.048	0.06	0.034	0.05	0.059
		rRMSE	0.054	0.079	0.095	0.056	0.079	0.093
		s	1.20%	1.90%	3.10%	-2.50%	2.20%	5.90%
		MBE	-0.005	-0.008	-0.008	-0.001	-0.003	-0.003
	SVR	MAPE	0.035	0.051	0.062	0.034	0.049	0.058
		rRMSE	0.056	0.080	0.096	0.055	0.078	0.092
		s	-1.10%	0.40%	2.10%	-0.40%	2.90%	6.20%

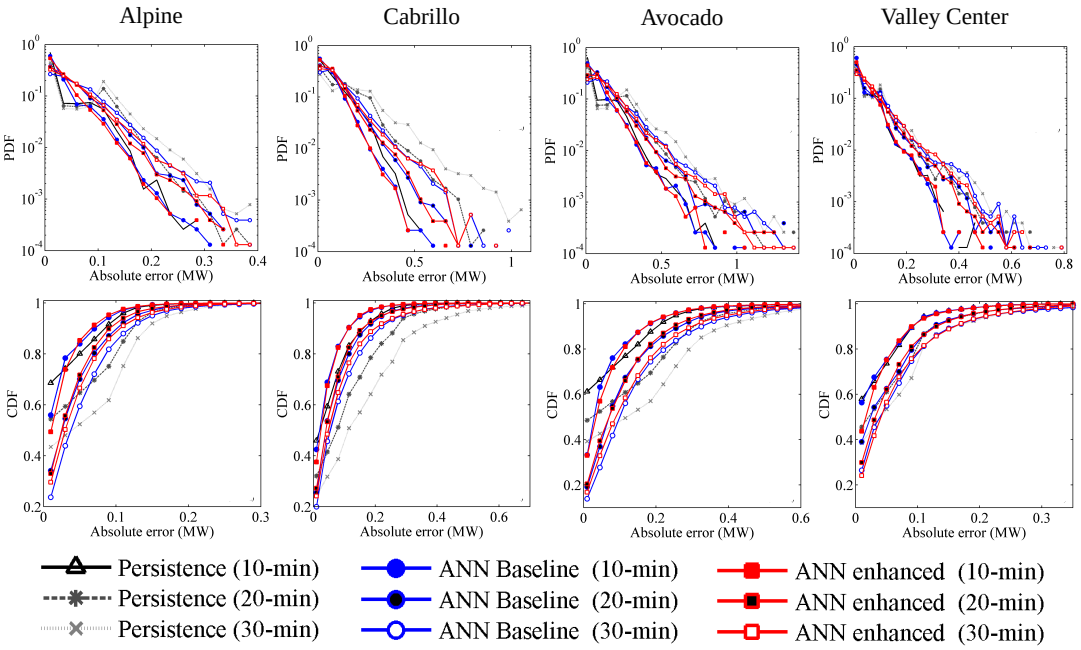

Figure 6: Probability Density Functions (PDFs, top row) and Cumulative Density Functions (CDFs, bottom row) for the persistence, ANN baseline, and ANN enhanced forecasts in the four locations under study. The PDF y-axis is in logarithmic scale such that the PDF details can be observed. The number of bins used to create the PDFs and the CDFs are ranging from 30 to 50. This number is found by increasing the number of bins until the convergence of the CDFs. Persistence forecasts for 10-, 20-, and 30-minute horizon are presented for comparisons.

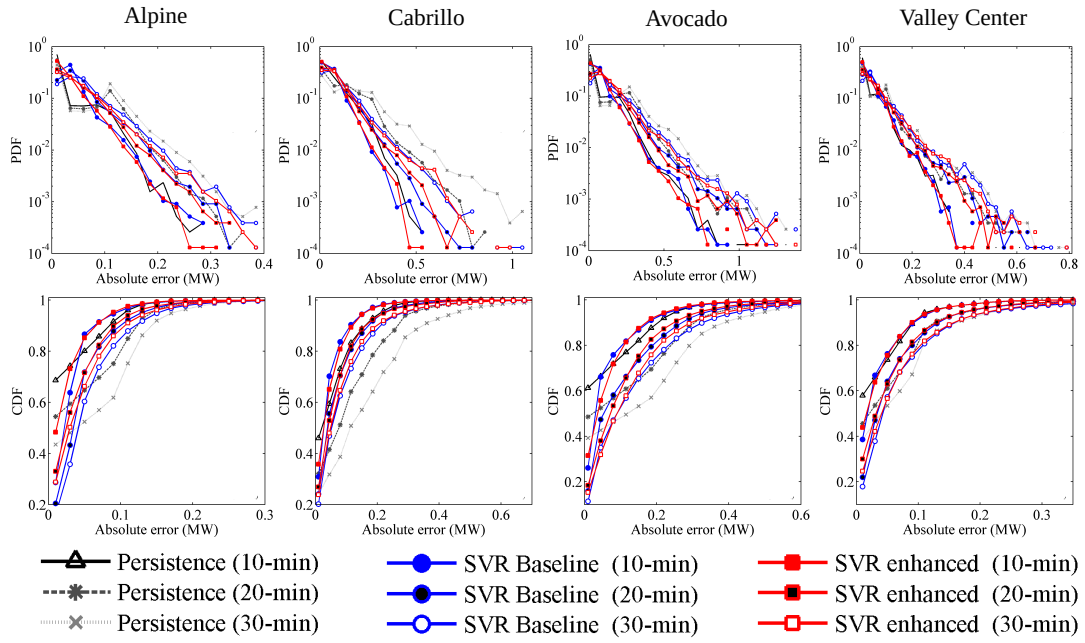

Figure 7: Probability Density Functions (PDFs, top row) and Cumulative Density Functions (CDFs, bottom row) for the persistence, SVR baseline, and SVR enhanced forecasts in the four locations under study. The PDF y-axis is in logarithmic scale such that the PDF details can be observed. The number of bins used to create the PDFs and the CDFs are ranging from 30 to 50. This number is found by increasing the number of bins until the convergence of the CDFs. Persistence forecasts for 10-, 20-, and 30-minute horizon are presented for comparisons.

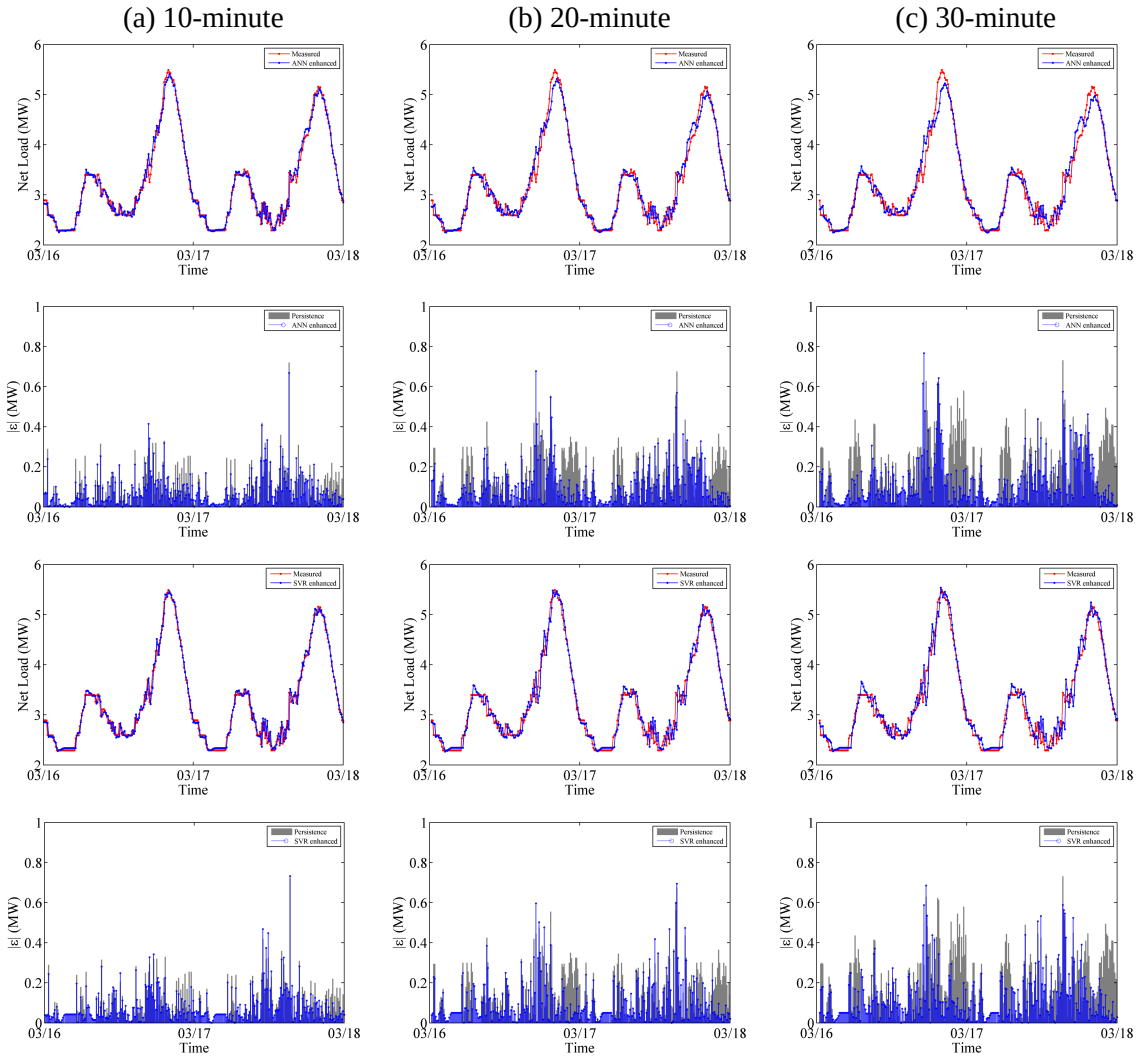

Figure 8: Sample time series of (a) 10-minute, (b) 20-minute, and (c) 30-minute net load forecasts and absolute forecast errors (ε) for the Cabrillo feeder for a period of 48h using ANN and SVR. The timestamps are in PDT.

5. Conclusions

Intra-hour net load forecasts based on stochastic-learning models are successfully implemented for four feeders in the SDG&E operating region. The forecast models analyzed are based on Artificial Neural Network (ANN), Support Vector Machine (SVM), and persistence. Three strategies are introduced to enhance the stochastic-learning models: (1) decomposing the time series of the net load to remove the daily trend; (2) training two models with daytime and nighttime data and applying them for day period and night period, respectively; and (3) introducing sky image features as exogenous inputs for daytime forecasts. The ANN and SVR models are trained and validated using 6-month net load measurements and sky images.

The results show that both ANN and SVR significantly outperform the reference persistence model in terms of MAPE, rRMSE, and forecast skill, particularly for the 30-minute horizon. For example, ANN and SVR forecasts for Cabrillo feeders achieve forecast skills over 40% for the 30-minute horizon. Detrending the net load time series and introducing sky image features as exogenous inputs are identified as useful enhancement methods to reduce the occurrence of moderate and large forecast errors and to improve the overall accuracy of forecasts. Furthermore, the sky-imaging techniques are expected to noticeably enhance the performance of stochastic-learning forecasts for feeders with high solar penetration levels. This work is particularly relevant to grid operation and real-time dispatch for utilities and grid system operators.

Acknowledgments

The authors gratefully acknowledge funding from the California Solar Initiative (CSI) Research, Development, Demonstration, and Deployment (RD&D) Program Grant IV.

References

[1] M. A. Ortega-Vazquez, D. S. Kirschen, Economic impact assessment of load forecast errors considering the cost of interruptions, in: Power Engineering Society General Meeting, 2006. IEEE, IEEE, 8–pp, 2006.

[2] A. Kaur, H. T. C. Pedro, C. F. M. Coimbra, Ensemble re-forecasting methods for enhanced power load prediction, Energy Conversion and Management 80 (2014) 582–590.

[3] G. Gross, F. D. Galiana, Short-term load forecasting, Proceedings of the IEEE 75 (12) (1987) 1558–1573.

[4] M. T. Hagan, S. M. Behr, The time series approach to short term load forecasting, Power Systems, IEEE Transactions on 2 (3) (1987) 785–791.

[5] I. Moghram, S. Rahman, Analysis and evaluation of five short-term load forecasting techniques, Power Systems, IEEE Transactions on 4 (4) (1989) 1484–1491.

[6] H. K. Alfares, M. Nazeeruddin, Electric load forecasting: literature survey and classification of methods, International Journal of Systems Science 33 (1) (2002) 23–34.

[7] K. Metaxiotis, A. Kagiannas, D. Askounis, J. Psarras, Artificial intelligence in short term electric load forecasting: a state-of-the-art survey for the researcher, Energy Conversion and Management 44 (9) (2003) 1525–1534.

[8] H. Hahn, S. Meyer-Nieberg, S. Pickl, Electric load forecasting methods: Tools for decision making, European Journal of Operational Research 199 (3) (2009) 902–907.

[9] M. Alamaniotis, A. Ikonomopoulos, L. H. Tsoukalas, Evolutionary multiobjective optimization of kernel-based very-short-term load forecasting, Power Systems, IEEE Transactions on 27 (3) (2012) 1477–1484.

[10] M. Matijaš, J. A. K. Suykens, S. Krajcar, Load forecasting using a multivariate meta-learning system, Expert systems with applications 40 (11) (2013) 4427–4437.

[11] L. Suganthi, A. A. Samuel, Energy models for demand forecastingA review, Renewable and sustainable energy reviews 16 (2) (2012) 1223–1240.

[12] A. Kaur, H. T. C. Pedro, C. F. M. Coimbra, Impact of onsite solar generation on system load demand forecast, Energy Conversion and Management 75 (2013) 701–709.

[13] P. Denholm, R. M. Margolis, Evaluating the limits of solar photovoltaics (PV) in electric power systems utilizing energy storage and other enabling technologies, Energy Policy 35 (9) (2007) 4424–4433.

[14] H. Yang, B. Kurtz, D. Nguyen, B. Urquhart, C. W. Chow, M. Ghonima, J. Kleissl, Solar irradiance forecasting using a ground-based sky imager developed at UC San Diego, Solar Energy 103 (2014) 502–524.

[15] R. H. Inman, H. T. C. Pedro, C. F. M. Coimbra, Solar forecasting methods for renewable energy integration, Progress in Energy and Combustion Science 39 (6) (2013) 535–576.

[16] C. Chang, C. Lin, LIBSVM: A library for support vector machines, ACM Transactions on Intelligent Systems and Technology 2 (2011) 27:1–27:27.

[17] Y. Chu, H. T. C. Pedro, C. F. M. Coimbra, Hybrid intra-hour DNI forecasts with sky image processing enhanced by stochastic learning, Solar Energy 98 (2013) 592–603.

[18] C. M. Bishop, Neural networks and their applications, Review of scientific instruments 65 (6) (1994) 1803–1832.

[19] S. M. Al-Alawi, H. A. Al-Hinai, An ANN-based approach for predicting global radiation in locations with no direct measurement instrumentation, Renewable Energy 14 (1998) 199–204.

[20] A. Mellit, S. A. Kalogirou, Artificial intelligence techniques for photovoltaic applications: A review, Progress in Energy and Combustion Science 34 (5) (2008) 574–632.

[21] R. Marquez, C. F. M. Coimbra, Forecasting of Global and Direct Solar Irradiance Using Stochastic Learning Methods, Ground Experiments and the NWS Database, Solar Energy 85 (5) (2011) 746–756.

[22] S. X. Chen, H. B. Gooi, M. Q. Wang, Solar radiation forecast based on fuzzy logic and neural networks, Renewable Energy 60 (0) (2013) 195–201.

[23] V. Vapnik, The nature of statistical learning theory, springer, New York, 2000.

[24] C. Huang, L. S. Davis, J. R. G. Townshend, An assessment of support vector machines for land cover classification, International Journal of Remote Sensing 23 (4) (2002) 725–749.

[25] F. Melgani, L. Bruzzone, Classification of hyperspectral remote sensing images with support vector machines, Geoscience and Remote Sensing, IEEE Transactions on 42 (8) (2004) 1778–1790.

[26] Y. Chu, H. T. C. Pedro, L. Nonnenmacher, R. H. Inman, Z. Liao, C. F. M. Coimbra, A Smart Image-Based Cloud Detection System for Intra-hour Solar Irradiance Forecasts, Journal of Atmospheric and Oceanic Technology 31 (2014) 1995–2007.

[27] H. T. C. Pedro, C. F. M. Coimbra, Nearest-Neighbor Methodology for Prediction of Intra-Hour Global Horizontal and Direct Normal Irradiances, Renewable Energy 80 (2015) 770–782.

[28] Y. Chu, M. Li, H. T. C. Pedro, C. F. M. Coimbra, Real-time Prediction Intervals for Intra-hour DNI Forecasts, Renewable Energy 83 (2015) 234–244.

[29] R. Marquez, C. F. M. Coimbra, Intra-Hour DNI Forecasting Methodology Based on Cloud Tracking Image Analysis, Solar Energy 91 (2013) 327–336.

[30] Y. Chu, H. T. C. Pedro, M. Li, C. F. M. Coimbra, Real-time Forecasting of Solar Irradiance Ramps with Smart Image Processing, Solar Energy 114 (2015) 91–104.