

# UC San Diego

## UC San Diego Electronic Theses and Dissertations

### Title

Forecasting for power grids with high solar penetration

### Permalink

<https://escholarship.org/uc/item/0tq176zz>

### Author

Kaur, Amanpreet

### Publication Date

2015

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

**Forecasting for power grids with high solar penetration**

A dissertation submitted in partial satisfaction of the  
requirements for the degree  
Doctor of Philosophy

in

Engineering Sciences (Mechanical Engineering)

by

Amanpreet Kaur

Committee in charge:

Professor Carlos F.M. Coimbra, Chair  
Professor Farhat Beg  
Professor Jan Kleissl  
Professor Tajana S. Rosing  
Professor Daniel L. Rudnick

2015

©

Amanpreet Kaur, 2015

All rights reserved.

The dissertation of Amanpreet Kaur is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

---

---

---

---

---

---

Chair

University of California, San Diego

2015

DEDICATION

*To my parents and brother for their love and support.*

## EPIGRAPH

*If you want to find the secrets of the universe,  
think in terms of energy, frequency and vibration.*

— Nikola Tesla

## TABLE OF CONTENTS

Signature Page . . . . .		iii
Dedication . . . . .		iv
Epigraph . . . . .		v
Table of Contents . . . . .		vi
List of Figures . . . . .		ix
List of Tables . . . . .		xi
Acknowledgements . . . . .		xiii
Vita . . . . .		xv
Abstract of the Dissertation . . . . .		xvi
Chapter 1	Introduction . . . . .	1
	1.1 Notation for forecasting . . . . .	2
	1.2 Organization of the thesis . . . . .	2
Chapter 2	Load forecasting . . . . .	4
	2.1 Related research . . . . .	5
	2.2 Proposed methodology . . . . .	9
	2.2.1 HAM re-forecast . . . . .	9
	2.2.2 DAM re-forecast . . . . .	9
	2.2.3 Preprocessing . . . . .	11
	2.2.4 Detrending . . . . .	11
	2.2.5 Re-forecast models . . . . .	12
	2.2.6 Ensemble re-forecast . . . . .	13
	2.2.7 Recursive least squares model ensemble (RLS-ME) . . . . .	16
	2.3 Data . . . . .	17
	2.3.1 California Independent System Operator (CAISO) . . . . .	17
	2.3.2 Electric Reliability Council of Texas (ERCOT) . . . . .	19
	2.4 Results and discussion . . . . .	19
	2.4.1 Error statistics . . . . .	19
	2.4.2 Residual analysis . . . . .	22
	2.5 Summary . . . . .	24

Chapter 3	Impact of increasing solar penetration on load forecasting . . .	26
3.1	Data . . . . .	29
3.1.1	Preprocessing data . . . . .	29
3.1.2	Training and testing datasets . . . . .	30
3.2	Load forecast methods . . . . .	30
3.2.1	Persistence-Load ( $P_L$ ) . . . . .	31
3.2.2	Smart Persistence-Load ( $SP_L$ ) . . . . .	31
3.2.3	Autoregressive (AR) . . . . .	31
3.2.4	Autoregressive Moving Average (ARMA) . . . . .	32
3.2.5	k-th Nearest Neighbor (kNN) . . . . .	32
3.2.6	Error metrics . . . . .	34
3.3	Results and discussion . . . . .	34
3.3.1	1-hour forecasts . . . . .	35
3.3.2	15-minute forecasts . . . . .	36
3.3.3	Error distribution . . . . .	37
3.3.4	Effect of solar penetration on the forecast error . . . . .	40
3.4	Summary . . . . .	44
Chapter 4	Solar resource forecasting . . . . .	46
4.1	Data sets . . . . .	47
4.1.1	NWP data . . . . .	47
4.1.2	Ground data . . . . .	47
4.1.3	Solar power modeling . . . . .	49
4.2	Solar forecast methods . . . . .	50
4.2.1	Persistence (P) . . . . .	50
4.2.2	Smart Persistence (SP) . . . . .	50
4.2.3	Support Vector Regression (SVR) . . . . .	50
4.2.4	Reforecast . . . . .	51
4.3	Forecast results . . . . .	52
4.3.1	Intra-week forecast (day-ahead) . . . . .	52
4.3.2	Intra-hour forecast . . . . .	53
4.4	Summary . . . . .	58
Chapter 5	Net load forecasting for power grids with high solar penetration	60
5.1	Background . . . . .	61
5.1.1	Solar power generation forecasting . . . . .	62
5.1.2	Load forecasting . . . . .	63
5.1.3	Net load forecasting . . . . .	63
5.2	Microgrids as testbeds . . . . .	66
5.2.1	Testbed data . . . . .	66
5.3	Proposed methodology . . . . .	67
5.3.1	Solar power forecast . . . . .	68
5.3.2	Net load forecast . . . . .	72

	5.4	Model implementation . . . . .	74
	5.5	Results and discussion . . . . .	75
	5.5.1	Clear sky identification . . . . .	75
	5.5.2	Solar power forecast . . . . .	79
	5.5.3	Net load forecast . . . . .	79
	5.5.4	Assessment of forecast uncertainty . . . . .	82
	5.6	Solar and net load forecast errors . . . . .	83
	5.6.1	Cointegration of solar and net load forecast errors	85
	5.7	Summary . . . . .	86
Chapter 6		Benefits of forecasting . . . . .	89
	6.1	Energy imbalance markets . . . . .	90
	6.1.1	Goals . . . . .	90
	6.1.2	Previous work . . . . .	91
	6.1.3	Market design variables . . . . .	92
	6.1.4	EIM in Western Interconnection . . . . .	93
	6.1.5	Forecasting for EIMs . . . . .	95
	6.2	Implications on EIMs . . . . .	96
	6.2.1	Reserve scheduling . . . . .	96
	6.2.2	EIMs with GCT greater than 75 minutes . . . . .	97
	6.2.3	EIMs with GCT lower than 75 minutes . . . . .	97
	6.2.4	Implications for the Western Interconnection EIM	100
	6.3	Summary . . . . .	103
Chapter 7		Conclusions . . . . .	105
Bibliography		. . . . .	109

## LIST OF FIGURES

Figure 2.1:	Schematic of the proposed methodology for re-forecasting. . . .	10
Figure 2.2:	The CAISO ( <i>top</i> ) and ERCOT ( <i>bottom</i> ) time-series. . . . .	18
Figure 2.3:	Time-series of actual CAISO load and HAM load prediction for several days in 2013 by CAISO and NARX re-forecast model ( <i>top</i> ) and the instantaneous absolute error ( <i>bottom</i> ). . . . .	23
Figure 2.4:	Autocorrelation and partial autocorrelation for the residuals from CAISO and NARX re-forecast models for HAM load prediction. . . . .	23
Figure 2.5:	Temporal distribution of the mean error (actual load-forecast HAM load) over the day for CAISO ( <i>left</i> ) and ERCOT ( <i>right</i> ). . . . .	24
Figure 3.1:	UC Merced load profile for 06-Oct-2010 through 09-Oct-2010. . . . .	28
Figure 3.2:	Daily demand load profile for the UC Merced campus with no onsite power generation ( <i>top</i> ) and onsite solar power generation ( <i>bottom</i> ). . . . .	29
Figure 3.3:	Error distribution for 15-minute load forecasting with solar penetration. . . . .	37
Figure 3.4:	Error distribution for 1-hour and 15-minute load forecasting with and without solar penetration. . . . .	39
Figure 3.5:	Cumulative Distribution Function of absolute value of step change in load demand with ( $V_{Lg}$ ) and without ( $V_{Lng}$ ) onsite solar generation. . . . .	40
Figure 3.6:	Error distribution as a function of penetration with different levels of variability. . . . .	41
Figure 3.7:	Time-series of load and error profiles for a clear, cloudy and overcast day for 1 hour forecast case. . . . .	43
Figure 4.1:	Sample day of GHI forecast by NAM and reforecast model. . . . .	53
Figure 4.2:	Mean and standard deviation of day-ahead forecast errors for GHI as a function of hour of the day. . . . .	54
Figure 4.3:	RMSE and standard deviation for short-term GHI forecasts for the 15 minute forecast interval with forecast horizon ranging from 15 to 75 minutes. . . . .	56
Figure 4.4:	RMSE and standard deviation for short-term solar forecasts with horizons of 5 to 75 minutes in 5 minute intervals. . . . .	57
Figure 4.5:	Cumulative frequency distribution (CDF) of the solar forecast errors in five minute intervals for the Persistence (P), Smart persistence (SP) and forecast model (SVR-GA). . . . .	58
Figure 5.1:	Block diagram for net load for UC Merced system. . . . .	67

Figure 5.2:	Time series for the solar power generated for the 10 consecutive days from the year 2011. . . . .	77
Figure 5.3:	Time series for the actual solar power and forecast for 1-h forecast horizon ( <i>top</i> ) and absolute error, AE ( <i>bottom</i> ) with night values removed. . . . .	78
Figure 5.4:	Time series for the net load and solar power forecast for six consecutive days (11/06/2011 to 11/12/2011) from the testset. . . . .	80
Figure 5.5:	Error correlation for net load forecast models. . . . .	82
Figure 5.6:	Comparison of net load forecast errors for both additive and integrated models w.r.t. the solar forecast errors from $SVR_{\alpha,h}$ model during daytime. . . . .	83
Figure 5.7:	Net load and solar forecast uncertainty assessment by applying inverse CDF during daytime. . . . .	84
Figure 5.8:	Cointegrating relation between the hourly integrated net load and solar forecast errors during the daytime for the first consecutive 50 hours. . . . .	86
Figure 6.1:	Timeline for the real-time Energy Imbalance Market in the Western Interconnection for the operating hour beginning at time $T$ (Market 1). . . . .	95
Figure 6.2:	Standard deviation in forecast error for forecast horizons ranging from 1 day to 5 minutes ahead forecast horizons in 1 hour, 15 minute and 5 minute resolution. . . . .	98
Figure 6.3:	Benefits of 5 minute versus 15 minute reserve allocation for real-time markets considering 5 minute data interval. . . . .	99
Figure 6.4:	Flexibility reserve required for 1 MW solar plant for 5 minute intervals. . . . .	100
Figure 6.5:	Energy imbalances (probability for the error to exceed 7.5%) for 1 MW solar plant for 5-75 minutes and 15-75 minutes forecast horizon with 5 minute and 15 minute resolution. . . . .	102
Figure 6.6:	Inverse cumulative frequency distribution of the expected forecast errors that market operator has to optimize for in 5-minute market after scheduling the awards for 15 minute market. . . . .	103

## LIST OF TABLES

Table 2.1:	Ensemble models applied for short term load forecasting. . . . .	8
Table 2.2:	Polynomials used for various model structures . . . . .	13
Table 2.3:	Details for the dataset used for HAM and DAM forecast. . . . .	17
Table 2.4:	Number of model parameters used for CAISO and ERCOT re-forecast time-series models. . . . .	20
Table 2.5:	Statistical error metrics for HAM load forecast using various models for CAISO (3-18-2013 to 06-11-2013) and ERCOT (05-01-2013 to 06-20-2013). . . . .	20
Table 2.6:	Statistical error metrics for the DAM load forecast using various models for CAISO (3-18-2013 to 06-11-2013) and ERCOT (05-01-2013 to 06-20-2013). . . . .	21
Table 3.1:	The coefficients of the polynomial $A(q)$ for AR model for the two cases: no onsite generation (no) and onsite solar generation (yes). . . . .	35
Table 3.2:	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). . . . .	35
Table 3.3:	Statistical error metrics for the 1-hour demand load forecasting with no onsite generation and with onsite solar generation. . . . .	36
Table 3.4:	Statistical error metrics for 15-minute demand load forecasting Without onsite generation and with onsite solar generation. . . . .	36
Table 3.5:	Statistics of average error distribution for no onsite generation and onsite solar generation case for 1-hour and 15-minute lead times. . . . .	38
Table 4.1:	Statistical error metrics for GHI to PV output modeling for 1 MW <sub>p</sub> capacity fixed-array PV plant. . . . .	50
Table 4.2:	Statistical error metrics for day-ahead GHI forecasts in hourly intervals. . . . .	53
Table 4.3:	Features selected by GA for 15 minute and 5 minute forecast intervals for the SVR-GA model. . . . .	54
Table 4.4:	Statistical error metrics for short-term solar irradiance forecast in 15 minute intervals. . . . .	55
Table 4.5:	Statistical error metrics for short-term solar irradiance forecast in 5-minute intervals. . . . .	57
Table 5.1:	Forecast models proposed for solar power forecasting. . . . .	64
Table 5.2:	Thresholds for clear sky solar power identification. . . . .	70
Table 5.3:	Output and input variables for the SVR model. . . . .	74
Table 5.4:	Clear sky identification results . . . . .	76

Table 5.5: Statistical error metrics for hour-ahead solar power forecast (01-01-2011 to 12-31-2011). . . . .	76
Table 5.6: Statistical error metrics for UCM load demand forecast (01-01-2011 to 12-31-2011). . . . .	81
Table 6.1: Summary of benefits of solar forecasting for the real-time energy markets with respect to the Persistence (P) and Smart persistence (SP) forecast model using solar power data for 1MW plant.	101

## ACKNOWLEDGEMENTS

I would like to acknowledge my family, advisors, professors and friends who made this work possible, especially:

Professor Carlos F.M. Coimbra, who convinced me to join University of California, San Diego. I am very grateful for his immense support and patience all the time. He has been a perfect advisor, giving insightful advice and suggestions. Thanks to him for the continuous guidance and mentorship.

Professor Ram Nunna, who encouraged me to pursue a PhD and has been a great support ever since I met him. Professor Jeff Wright for addressing all my concerns when I was planning to join a PhD program and introducing me to Professor Carlos F.M. Coimbra. My committee members Professor Farhat Beg, Professor Jan Kleissl, Professor Tajana S. Rosing and Professor Daniel L. Rudnick for their time, support and feedback.

Hugo T.C. Pedro for providing the initial and continuous support since I joined the research group. Lukas Nonnenmacher for all the brainstorming sessions and teamwork. Other lab members including Ricardo Marquez, Bron Davis, David Larson, Athanassios Zagouras, Rich Inman and Yinghao Chu. Thank you for the instrumentation, database management, and other collaborative work.

My parents for their love, support and encouragement. All the credit for my success goes to them. My brother for his words of wisdom and patient ears. His support kept me going in all the good and bad times. My grandparents and extended family for their love and support.

Special thanks to Andres Cortes for everything he does to support all my endeavors. All my friends especially Daniele Cavalgieri, Minyi Ji, Minu Jacob, Rishi Kapoor, Ashish Tawari, Alaena Alilin, Diana Higuera, Annie Pham, Eduardo Ramirez, David Mateos, Siddharth Joshi, Jyoti Wadhvani, Bharathan Balaji, Arpit Gupta, Navreet Khosa, Neelam Sharma, Nancy Aggarwal, Rini Kalsi, Parneet Jammu, Prabhjot Kaur, Shivani Bisht, Navreet Cheema, Harpreet Wallia, Prabhnoor Kaur, and Gurvinder Kaur..., who contributed in various ways to this journey and made my life in California and Punjab even more fun. Thank you for all the support, celebrations, chats, discussions and adventures.

Co-author acknowledgements: This dissertation contains parts of the following publications with Hugo T.C. Pedro, Lukas Nonnenmacher and Carlos F.M. Coimbra as co-authors:

- A. Kaur, H.T.C. Pedro, and C.F.M. Coimbra, “Impact of onsite solar generation on system load demand forecast”, *Energy Conversion and Management* 75 (0) (2013) 701 - 709.
- A. Kaur, H.T.C. Pedro, and C.F.M. Coimbra, “Ensemble re-forecasting methods for enhanced power load prediction”, *Energy Conversion and Management* 80 (0) (2014) 582 - 590.
- A. Kaur, L. Nonnenmacher and C.F.M. Coimbra, “Net load forecasting for high renewable energy penetration grids”, *Energy*, [Submitted, December 2014].
- A. Kaur, L. Nonnenmacher, H.T.C. Pedro and C.F.M. Coimbra, “Benefits of solar forecasting for energy imbalance market”, *Renewable Energy*, [Submitted, March 2015].

The specific publication from where the content was used can be found at the end of each chapter.

## VITA

- 2008 B. Tech., Electronics and Communication, Punjab Technical University.
- 2010 M. S., Electrical Engineering, California State University Fresno.
- 2015 Ph. D., Engineering Sciences (Mechanical Engineering), University of California, San Diego.

## PUBLICATIONS

A. Kaur, H.T.C. Pedro, and C.F.M. Coimbra, "Impact of onsite solar generation on system load demand forecast", *Energy Conversion and Management* 75 (0) (2013) 701 - 709.

A. Kaur, H.T.C. Pedro, and C.F.M. Coimbra, "Ensemble re-forecasting methods for enhanced power load prediction", *Energy Conversion and Management* 80 (0) (2014) 582 - 590.

A. Kaur, L. Nonnenmacher and C.F.M. Coimbra, "Net load forecasting for high renewable energy penetration grids", *Energy*, [Submitted, December 2014].

A. Kaur, L. Nonnenmacher, H.T.C. Pedro and C.F.M. Coimbra, "Benefits of solar forecasting for energy imbalance market", *Renewable Energy*, [Submitted, March 2015].

L. Nonnenmacher, A. Kaur, C. F. M. Coimbra, "Verification of the SUNY direct normal irradiance model with ground measurements", *Solar Energy* 99 (0) (2014) 246 - 258.

L. Nonnenmacher, A. Kaur and C.F.M. Coimbra, "Benchmarking and valuation of day-ahead forecasts for direct normal irradiance", *Solar Energy*, [Submitted, October 2014].

L. Nonnenmacher, A. Kaur and C.F.M. Coimbra, "Day-ahead resource forecasting for concentrated solar power", *Renewable Energy*, [Submitted, March 2015].

## FIELDS OF STUDY

Major Field: Mechanical Engineering; Electrical Engineering

Areas of Specialization: Energy systems; Power grid; Renewable energy (solar); Electricity markets; Forecasting; Data analytics

ABSTRACT OF THE DISSERTATION

**Forecasting for power grids with high solar penetration**

by

Amanpreet Kaur

Doctor of Philosophy in Engineering Sciences (Mechanical Engineering)

University of California, San Diego, 2015

Professor Carlos F.M. Coimbra, Chair

Global solar photovoltaic capacity increased by 35% from 2013 to 2014, and this upward growth trend is likely to continue. Power grids must adapt to accommodate increasing shares of renewable energy penetration. The impact of increasing solar penetration is quantified in terms of the variability and the predictability of net load behavior. As expected, due to variable nature of solar technologies, the predictability of net load decreases with increasing penetration. The need for novel net load forecasting techniques that allow for improved management of grids with high solar penetration is discussed. Integrated net load forecasting methods (solar power forecasts are used as inputs) are recommended for grid operators and utilities. Analysis of forecast performance reveals that the solar variability plays a dominant role in driving the forecasting errors, even more so than the penetration

levels. Net load and solar forecast errors are found to be co-integrated, sharing a common stochastic drift. Thus, the solar irradiance time series is sufficient to provide necessary information for the future planning of reserve allocation and storage design for power grids. The benefits of proposed techniques are presented for real-time energy imbalance markets. Design variables regulating the electricity markets and grid timelines govern the system dynamics, which in turn highlight the benefits of forecasting. Increased flexibility of operations at shorter time-scales emerges as a key factor for the reliable and efficient management of power grids.

# Chapter 1

## Introduction

Global electricity demand grows by an average of 0.8% every year [1]. To meet the increasing demand in a sustainable manner, it is imperative that we exploit renewable resources. Amongst all the renewable energy sources, it is expected that the larger shares of renewable energy will come from solar and wind energy technologies [1]. Global solar energy capacity has increased from 1.28 GW to 138.86 GW over the past decade (2000-2013) [2], and this upward growth trend is expected to continue. In the future, we will see an increase in grids with high solar and wind energy penetration. However, due to increasing variability and uncertainty in load demand and power generation, the intermittent nature of these technologies presents many challenges in maintaining the grid reliability and stability. Various solutions to compensate for the intermittency have been proposed e.g. hybrid energy systems, storage, demand response, resource sharing, etc. The integration of all these proposed solutions is dependent on the quality of the available information retrieved by forecasting.

Forecasting is a fundamental decision making tool for management and operation of power grids. Intra-year and intra-month forecast is used for long term strategic planning for construction and maintenance of the grid infrastructure, intra-week forecast is used for electricity market participation, resource allocation and unit commitment, intra-hour forecasting is applied for real-time scheduling, load following and automatic generation control, etc.

With the changing power grid infrastructure, forecasting needs are also

evolving to accommodate higher penetration of intermittent renewable energy sources for instance solar and wind energy. There is a need to understand and characterize changing net load profiles with increasing renewable energy penetration where net load represents the load that is traded between the power grids (e.g. between a microgrid and a utility grid). This is important for resource sharing and reserves allocation at the point of common coupling between the interconnected grids and integration of fleet of microgrids with the macrogrid. For such future grid scenarios, the standard forecasting methods need to be replaced with net load forecasting i.e., a combination of load and solar forecasting.

## 1.1 Notation for forecasting

Three parameters that characterize a forecast are forecast horizon ( $k$ ), forecast interval ( $fi$ ) and forecast resolution ( $fr$ ). For a given time  $t$ , the forecast horizon is how many steps/time ahead a forecast needs to be issued. Forecast interval is defined as the time range up to which the forecast is issued i.e  $[t+k, t+k+fi)$  and lastly, the forecast resolution  $fr$  denotes the frequency at which the forecasts are issued. The technical details of a forecast issued will be denoted as  $k/fi^{fr}$ . For instance  $15/10^5$  in minutes implies that a given time  $t$ , the forecast is issued 15 minutes ahead for the forecast interval  $[t+15, t+25)$  minutes and the resolution of forecasts is 5 minutes i.e. two forecast  $\{[t+15, t+20), [t+20, t+25)\}$  in minutes are issued.

## 1.2 Organization of the thesis

A review on state-of-the-art load forecast models and proposed ensemble re-forecast method is presented in Chapter 2. Chapter 3 characterizes the impact of an increasing solar penetration on load forecast capabilities and highlights the need for solar forecasting. Various solar forecasting techniques for forecast horizons ranging from intra day to intra-hour are presented in Chapter 4. Chapter 5 covers the optimal methods to integrate solar and load forecasting. The applications and

benefits of the proposed forecasting methodologies are quantified and discussed in Chapter 6. Lastly, the conclusions are drawn in Chapter 7.

## Acknowledgements

This chapter contains work previously published in:

- A. Kaur, H.T.C. Pedro, and C.F.M. Coimbra, “Impact of onsite solar generation on system load demand forecast”, *Energy Conversion and Management* 75 (0) (2013) 701 - 709.
- A. Kaur, H.T.C. Pedro, and C.F.M. Coimbra, “Ensemble re-forecasting methods for enhanced power load prediction”, *Energy Conversion and Management* 80 (0) (2014) 582 - 590.
- A. Kaur, L. Nonnenmacher and C.F.M. Coimbra, “Net load forecasting for high renewable energy penetration grids”, *Energy*, [Submitted, December 2014].
- A. Kaur, L. Nonnenmacher, H.T.C. Pedro and C.F.M. Coimbra, “Benefits of solar forecasting for energy imbalance market”, *Renewable Energy*, [Submitted, March 2015].

# Chapter 2

## Load forecasting

Load forecast plays a key role in operation, control, and management of the grid. Each forecast depending on its forecast horizon has a specific application for grid balancing and scheduling. Day-ahead market (DAM) load forecasts are required by Independent System Operators (ISOs), utilities, and electricity market participants for operation planning and unit commitment of generating plants. Hour-ahead market (HAM) load forecasts are needed for real-time control and load following. To compensate for the uncertainty in load forecasts ISO maintains an operating reserve [3], which increases cost for ISOs, utilities, and customers [4].

Continuous research is being conducted to lower the uncertainty and increase the accuracy of load forecasts. Various methodologies based on time-series, regression, fuzzy logic, Artificial Neural Networks (ANNs), expert systems, hybrid models, etc., have been proposed [5]. Most of these methods rely on external inputs such as meteorological forecasts, temperature, dew point, etc., [6, 7]. Since no forecast is completely accurate, the errors in meteorological forecasts are propagated to the load forecasts [6]. Thus, the analysis of the forecast error distribution is essential to reveal the error structure which, if systematic and non-random, could be refined for better prediction.

We propose refining techniques that extracts the information from the structured forecast error needed to enhance forecast accuracy. The predictions are ensembled using optimized weights for the models, depending on their global, hourly, and weekly performance. The proposed methodology is applied to forecast load.

DAM load estimates provided by the ISOs are used as an exogenous input to re-forecast HAM and DAM loads.

The major contribution of this study is the demonstration that with the proper re-forecasting and ensemble techniques it is possible to substantially improve publicly available utility and ISOs load forecasts. Moreover, we demonstrate that it can be achieved without requiring any additional information (such as weather forecast). Another contribution is that our algorithms can be used to produce short-term forecasts based on the utility predictions, without requiring exogenous information. The proposed ensemble re-forecast methods for load prediction have several advantages: 1) methods are independent of seasonal cycles; rather they depend on the structure in the forecast residues. However, seasonal effects in the base forecast may influence the re-forecast produced by the proposed algorithm 2) the weights of the time-variant model are updated as the new data becomes available, so, the model is robust for the frequent changes in the time-series, and 3) the models can make predictions for any forecast horizon given that it is equal to or smaller than the forecast horizon of the input estimate.

The load data used in this study is from California Independent System Operator (CAISO) and Electric Reliability Council of Texas (ERCOT). Real-time load measurements, DAM and HAM load forecasts are publicly available on their websites [8, 9].

The organization of the study is as follows: literature review about the load forecast, ensemble methods, and re-forecast is presented in section 2.1; proposed methodology is described in section 2.2; dataset used for validation is explained in section 2.3; results are discussed in section 2.4 and the conclusions are drawn in section 2.5.

## 2.1 Related research

Several comprehensive reviews on load forecasting techniques have been published as research has progressed in this field. Comprehensive reviews on time-series modeling and forecasting can be found in [6, 10, 11]. The authors in [11]

concluded that Box-Jenkins time-series models are well suited for Short Term Load Forecast (STLF), but they do not capture the non-linearities in the load time-series. To circumvent this problem, they proposed a polynomial regression to linearize the relationship between load and temperature before using the time-series models.

A comparative evaluation of linear regression, stochastic time-series, general exponential smoothing, states-space methods, and knowledge-based approaches was presented in [12] by comparing their performance for the same load time-series. They showed that the regression and state space methods do not respond to sudden changes in load. Hence, the authors recommended to update the model parameters automatically and develop models specific to weekends and holidays.

In the 1990s, Artificial Neural Networks (ANNs), fuzzy logic, and Computational Intelligence (CI) techniques became popular and several studies reviewing their application for load forecasting were published [7, 13, 14, 15]. The studies [7, 16] emphasized the need for more rigorous standard tests to compare the ANN model performance and check for over-fitting and over-parameterization. Despite these concerns, [13] stated that ANNs capture the non-linearities in load data better than time-series based regression methods, and the potential of ANNs should not be overlooked.

Results from the Artificial Neural Networks & Computational Intelligence Forecasting competition (NN3 competition) were presented in [17] with a detailed discussion on merits, demerits, and performance of CI and statistical methods. It was observed that ensemble models outperformed the individual models. More recent forecast and optimization methods like support vector regression (SVR), simulated annealing, ant colony optimization, and other hybrid methods for energy forecasting were reviewed in [18, 5, 19]. All these reviews highlight the promising application of optimized hybrid and ensemble methods in load and other areas of energy forecasting [20, 21, 22, 23, 24].

Following the conclusions that combining several methods outperform single methods, we focus on ensemble optimization. Ensemble methods have been rigorously applied and tested for load forecasting (see Table 2.1). For instance in [25], local predictors were trained on selected inputs using k Nearest Neighbors

(kNN). This was beneficial because local predictors were able to retain the forecast accuracy for longer lead times than global predictors. The studies [26, 27] proposed using forecasted weather ensembles for load prediction. In [28], different training sets were used to train a network committee. The study [29] could be seen as a hybrid of [27, 28]. In this work various weather forecasts from different forecast services were combined using adaptive weights and an ensemble comprised of five ANNs.

The study [30] proposed an ensemble of kernels-based gaussian processes (GPs) where the linear model parameters were selected using genetic algorithm (GA), and showed that the ensemble outperforms SVR and ARMA. Various ensemble methods like simple averaging, singular value decomposition (SVD) based weighted averaging (for each hour separately), principal component analysis (PCA), and blind source separation were investigated in [31]. The best results were obtained by blind source separation method, where the data was decomposed into statistically independent components and the time-series was reconstructed by filtering the noise.

The studies [32, 33] proposed using selective ensemble techniques, i.e., a combination of models was optimally selected to form an ensemble. A discrete differential algorithm was used in [32] to select a combination of ANN models whereas a meta-learning technique was applied in [33] to select models based on meta-features like highest ACF, granularity, fickleness, etc. Results for various load profiles with different forecast horizons were presented to show that the meta-learning based ensemble is independent of load profile, forecast horizon, and outperforms various well established load forecasting models.

Motivated by the results from ensemble methods we propose novel re-forecast ensemble methods that filters the structured noise from a given estimate to produce a better prediction. We address the issue of extracting the information from non-random errors to further improve the forecasts. Previously, the re-forecast concept was investigated by [34] for DAM load forecast, where the forecast was updated to account for sudden changes in the weather. In this study no sudden update is made once the forecast has been issued.

**Table 2.1:** Ensemble models applied for short term load forecasting.

Ref.	Exogenous inputs	Input selection	No. of models	Forecast models	Ensemble methods
[25]	Average temperature, maximum and minimum temperature	k Nearest Neighbor	2	ANNs	Mean
[26]	51 Forecasted weather ensembles for temperature, wind speed, and cloud cover		51	Regression model	Mean
[27]	51 Forecasted weather ensembles for temperature, wind speed, and cloud cover		51	ANNs	Mean
[28]	Temperature and day-type variables		3	ANN committee, abductive network committees	Mean
[35]	Temperature, rainfall index, & holiday index			Fuzzy hyperrectangular composite neural network	Chaos-search genetic algorithm and simulated annealing
[29]	Forecasted weather from different services combined using adaptive weights	Bootstrap sampling	5	ANNs	Median
[32]	Temperature		20	ANNs	Mean and entropy weighted method for models selected using discrete differential algorithm
[31]	Type of the day, season of the year		3	Multilayer perceptron layer (MLP), self organizing map (SOM), and fuzzy SOM	Mean, singular value decomposition based weighted averaging (for each hour separately), principal component analysis decomposition, and blind source separation
[30]			3	Kernel-based gaussian processes regression	Evolutionary optimization
[33]	Temperature (if available)		7	Random walk algorithm, autoregressive moving average, similar days algorithm, layer recurrent neural network, MLP, $\nu$ -support vector regression, and robust LS-SVM	Meta learning

## 2.2 Proposed methodology

The schematic for the proposed methodology to improve the accuracy of publicly available utility and ISOs load forecasts and generate forecasts for shorter forecast horizon with no exogenous input is shown in Fig. 2.1. A base forecast with given forecast horizon  $k'$  is required to apply this methodology to produce  $k$  steps ahead forecast s.t.  $k \leq k'$ . For  $k = k'$ , if the forecast residuals are white then it is considered as the final forecast otherwise re-forecast and ensemble techniques are applied unless the residuals are white. Similarly, for  $k < k'$ , re-forecast and ensemble technique are applied directly to produce a short-term forecast and then again the residues are checked for whiteness. In this study, the proposed methodology is applied for  $k = 1$  and  $k \in \{1, 2, \dots, 24\}$  forecast horizon (in hours), the forecast interval and forecast resolution is 1 hour for both cases. The details are explained below.

### 2.2.1 HAM re-forecast

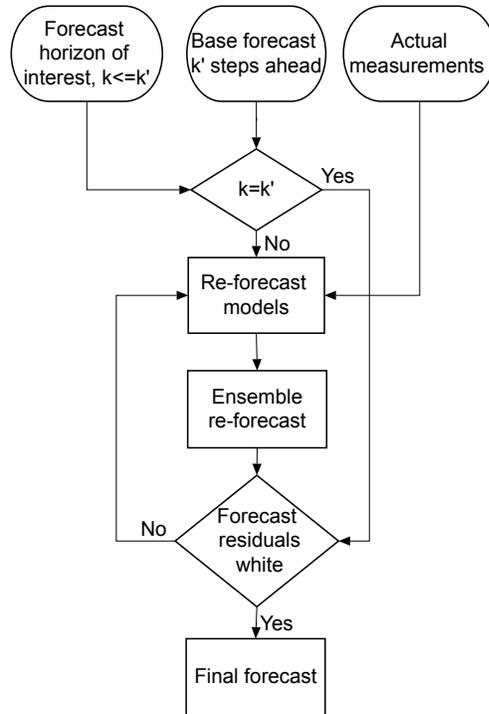
Using real-time system measurements, a given input  $u(t)$  is processed every hour by ensemble re-forecast models ( $m$ ) to produce HAM forecast  $\hat{y}(t+k)$ ,

$$\hat{y}(t+k) = m(y(t), y(t-1), \dots, u(t), u(t-1), u(t-2), \dots), \quad (2.1)$$

where  $t$  is time and  $k = 1$  is the forecast horizon. We use 24-hourly DAM load estimates provided by the ISO as an exogenous input, filter it every hour and produce HAM load forecasts for the system. For example, ISO DAM forecast produced on June 11, 2013 for the following 24 hours is corrected hourly by our re-forecast in order to produce HAM point forecast.

### 2.2.2 DAM re-forecast

The above idea is extended for DAM load prediction with forecast horizons  $k \in \{1, 2, \dots, 24\}$ . In this case, hourly forecasts are made at midnight for each hour  $h \in \{1, 2, \dots, 24\}$ , i.e.,  $\hat{Y} = [\hat{y}(t+1), \hat{y}(t+2), \dots, \hat{y}(t+24)]^T$ . The time-series models  $m(h)$  are built specific for each hour of the day with an exogenous



**Figure 2.1:** Schematic of the proposed methodology for ensemble re-forecasting. A base forecast upto  $k'$  steps ahead forecast horizon is required to produce  $k$  steps ahead forecast s.t.  $k \leq k'$ . If the forecast residuals are white then it is consider as the final forecast else re-forecast and ensemble techniques are applied unless the residuals are white.

input consisting of estimates for that hour only  $u(t-24n)|_{n=0,1,\dots}$  and corresponding actual measurements  $y(t-24n)|_{n=1,2,\dots}$ . In summary, we use the twenty-four hourly DAM estimated load provided by the ISO as an input, refine it by applying the models developed for each hour, combine the forecast from all the models and produce a vector of 24 load forecasts corresponding to all the hours of the next day.

### 2.2.3 Preprocessing

The time-series is preprocessed to remove the outliers and days with missing data. After preprocessing, the hourly time-series is generated and separated into three disjoint datasets: training set ( $\mathbf{T}$ ), simulation set ( $\mathbf{S}$ ) and validation set ( $\mathbf{V}$ ). The data points in  $\mathbf{T}$  were used to derive a polynomial fit to detrend the time-series and develop the models which were then tested using the  $\mathbf{S}$  dataset. The forecasts from  $\mathbf{T}$  and  $\mathbf{S}$  were combined to optimize the weights for the ensemble.

### 2.2.4 Detrending

In this work we applied a simple polynomial fit that captures the daily trend for the electric load. Using a simple least squares method (from the Curve Fitting toolbox available in Matlab) we determine that daily trends are best fitted by a 6th order polynomial as a function of time of the day  $t$ , specific for each weekday  $w \in \{1, 2, \dots, 7\}$  where Sunday is represented as 1, Monday as 2, and so on. The actual load data can be expressed as a sum of the polynomial fit and the detrended load, i.e.,

$$y(t, w) = \psi(t, w) + \delta(t, w), \quad (2.2)$$

where  $y$  represents the actual load and  $\psi$  is the corresponding detrended load using a polynomial fit for each day of the week  $\delta(t, w)$ . Similarly, the input for the models, i.e., a given estimate  $u$  is detrended,

$$u(t, w) = \mu(t, w) + \delta(t, w), \quad (2.3)$$

where  $\mu$  is the detrended input using the polynomial fit  $\delta(t, w)$ .

## 2.2.5 Re-forecast models

This study focuses on extracting the information from non-random noise to produce better forecasts using time-series models with the assumption that the model's residue is a white Gaussian noise. The filtered output produced by any of the models is called a re-forecast.

Various time-series models are applied at re-forecast stage. The linear and non-linear time-series models used in this work are well defined in the literature [6, 10, 11, 36]. Only a brief description is provided here.

### Linear models

A generalized model (GM) that linearly combines current and past values of the input  $\mu(t)$  and past values of the output  $\psi(t)$  to model current output can be defined as [36],

$$A(q)\psi(t) = \frac{B(q)}{F(q)}(q)\mu(t - n_k) + \frac{C(q)}{D(q)}(q)\epsilon(t), \quad (2.4)$$

where  $t$  represents time,  $n_k$  is the input-output delay parameter,  $\epsilon$  is assumed to be white noise,  $q$  is a shift operator, and  $A(q), B(q), C(q), D(q)$ , and  $F(q)$  are the polynomials of order  $n_a, n_b - 1, n_c, n_d$ , and  $n_k$ , i.e.,

$$q^{\pm N}\psi(t) = \psi(t \pm N), \quad (2.5)$$

$$A(q) = 1 + a_1q^{-1} + \cdots + a_{n_a}q^{-n_a}, \quad (2.6)$$

$$B(q) = b_1 + \cdots + b_{n_b}q^{-n_b+1}, \quad (2.7)$$

$$C(q) = 1 + c_1q^{-1} + \cdots + c_{n_c}q^{-n_c}, \quad (2.8)$$

$$F(q) = 1 + f_1q^{-1} + \cdots + f_{n_f}q^{-n_f}, \quad (2.9)$$

$$D(q) = 1 + d_1q^{-1} + \cdots + d_{n_d}q^{-n_d}. \quad (2.10)$$

Based on the application of GM and information about the physical system one or many polynomials can be fixed to unity. In this study we also apply the other forms of GM such as Autoregressive model with exogenous input (ARX),

Autoregressive moving average model with exogenous input (ARMAX), and Box-Jenkins model (BJ). Table 2.2 lists the polynomials used in each one of these models.

**Table 2.2:** Polynomials used for various model structures

Model Parameters	$A(q)$	$B(q)$	$C(q)$	$D(q)$	$F(q)$
ARX	✓	✓			
ARMAX	✓	✓	✓		
BJ		✓	✓	✓	✓
GM	✓	✓	✓	✓	✓

### Non-linear model

To model the non-linearities in the load time-series we use the Non-linear Autoregressive model (NARX). In this model the current output is expressed as a non-linear combination of the input and the past values of the output,

$$\psi(t) = f(\psi(t-1), \psi(t-2), \dots, \mu(t), \mu(t-1), \mu(t-2), \dots), \quad (2.11)$$

where  $f(\cdot)$  can be computed using any non-linear estimator. In this study, we use a wavelet network non-linear estimator. The prediction error method (PEM) is applied to derive the model parameters that minimize the weighted norm of the prediction error for the dataset  $\mathbf{T}$ .

### 2.2.6 Ensemble re-forecast

Ensemble re-forecast is produced by combining the above re-forecasts. The trend  $\delta(t, w)$  is added to the re-forecast  $\hat{\psi}(t)$  to produce a final re-forecast  $\hat{y}(t)$ ,

$$\hat{y}(t, w) = \hat{\psi}(t, w) + \delta(t, w). \quad (2.12)$$

We investigate three ways of forming an ensemble by optimizing the weights based on global, hourly, and weekly performance of the models. The mathematical details about the proposed ensembles are provided in the next subsections.

### Least squares model ensemble (LS-ME)

An ensemble forecast  $\hat{\phi}$  for any given time  $t$  is produced by linearly combining forecasts from  $n$  forecasting models for  $k$  forecast horizons,

$$\hat{Y} = \begin{pmatrix} \hat{y}_{1,1} & \hat{y}_{1,2} & \cdots & \hat{y}_{1,n} \\ \hat{y}_{2,1} & \hat{y}_{2,2} & \cdots & \hat{y}_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{y}_{k,1} & \hat{y}_{k,2} & \cdots & \hat{y}_{k,n} \end{pmatrix}, \quad (2.13)$$

using the weights  $W = [\omega_1, \omega_2, \dots, \omega_n]^T$ , i.e.,

$$\hat{\phi} = \hat{Y}W \quad (2.14)$$

For HAM load prediction  $k = 1$  and for DAM  $k = \{1, 2, \dots, 24\}$ . The weights  $\omega_i$  are optimized based on the global performance of the models applied to the T and S datasets, which consist of  $p$  actual load measurements  $Y = [y_1, y_2, \dots, y_p]^T$ . Each measurement can be uniquely identified by a timestamp  $\mathcal{T} = [t_1, t_2, \dots, t_p]^T$ . Correspondingly, there are  $p$  hourly forecasted data points from  $n$  forecasting models  $\hat{y}_{i,j}$  for  $i = 1, 2, \dots, p$  and  $j = 1, 2, \dots, n$ ,

$$\hat{Y} = \begin{pmatrix} \hat{y}_{1,1} & \hat{y}_{1,2} & \cdots & \hat{y}_{1,n} \\ \hat{y}_{2,1} & \hat{y}_{2,2} & \cdots & \hat{y}_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{y}_{p,1} & \hat{y}_{p,2} & \cdots & \hat{y}_{p,n} \end{pmatrix}. \quad (2.15)$$

The forecasted data points  $\hat{Y}$  are combined using  $W = [\omega_1, \omega_2, \dots, \omega_n]^T$  to produce a final forecast  $\hat{\phi} = [\hat{\phi}_1, \hat{\phi}_2, \dots, \hat{\phi}_p]^T$  with an error  $E = [e_1, e_2, \dots, e_p]^T$ ,

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{pmatrix} = \begin{pmatrix} \hat{y}_{1,1} & \hat{y}_{1,2} & \cdots & \hat{y}_{1,n} \\ \hat{y}_{2,1} & \hat{y}_{2,2} & \cdots & \hat{y}_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{y}_{p,1} & \hat{y}_{p,2} & \cdots & \hat{y}_{p,n} \end{pmatrix} \begin{pmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_n \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_p \end{pmatrix} = \begin{pmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \\ \vdots \\ \hat{\phi}_p \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_p \end{pmatrix}, \quad (2.16)$$

such that  $E = Y - \hat{Y}W$ , where  $W$  is approximated using linear least squares with an objective function of minimizing the sum of squares of errors, i.e.,  $G(T, W) = \frac{1}{2}E^T E$ , whose global solution is  $W = (\hat{Y}^T \hat{Y})^{-1} \hat{Y}^T Y \in \mathbb{R}^n$ .

### Least squares hourly ensemble (LS-HE)

In this model an ensemble forecast  $\hat{\phi}(h)$  depends on the hour of the day  $h$ , which can be easily retrieved from the timestamp  $t$ , i.e.,  $h \subset t$ .  $\hat{\phi}(h)$  is produced by linearly combining forecasts for the  $h^{th}$  hour from  $n$  forecasting models with  $k$  forecast horizons  $\hat{y}_{ij}(h)$ ,  $i \in \{1, 2, \dots, k\}$  and  $j \in \{1, 2, \dots, n\}$  using the weights  $\omega_j(h)$ ,

$$\hat{\phi}(h) = \sum_{i=1}^k \sum_{j=1}^n \hat{y}_{i,j}(h) \omega_j(h). \quad (2.17)$$

Using the timestamps  $\mathcal{T}$ , only  $q$  data points with  $q \leq p$  corresponding to  $h^{th}$  hour are considered to compute the weights  $w_i(h)$  and the same procedure as defined above is followed, i.e.,

$$\begin{pmatrix} y_1(h) \\ y_2(h) \\ \vdots \\ \hat{y}_q(h) \end{pmatrix} = \begin{pmatrix} \hat{y}_{1,1}(h) & \hat{y}_{1,2}(h) & \cdots & \hat{y}_{1,n}(h) \\ \hat{y}_{2,1}(h) & \hat{y}_{2,2}(h) & \cdots & \hat{y}_{2,n}(h) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{y}_{q,1}(h) & \hat{y}_{q,2}(h) & \cdots & \hat{y}_{q,n}(h) \end{pmatrix} \begin{pmatrix} \omega_1(h) \\ \omega_2(h) \\ \vdots \\ \omega_n(h) \end{pmatrix} + \begin{pmatrix} e_1(h) \\ e_2(h) \\ \vdots \\ e_q(h) \end{pmatrix} = \begin{pmatrix} \hat{\phi}_1(h) \\ \hat{\phi}_2(h) \\ \vdots \\ \hat{\phi}_q(h) \end{pmatrix} + \begin{pmatrix} e_1(h) \\ e_2(h) \\ \vdots \\ e_q(h) \end{pmatrix}. \quad (2.18)$$

The above algorithm is repeated twenty-four times  $\forall h \in \{1, 2, \dots, 24\}$  to compute individual weights for each hour of the day for all  $n$  models. In this case,  $W \in \mathbb{R}^{n,h}$ .

### Least squares weekday ensemble (LS-WE)

This model produces ensemble forecast  $\hat{\phi}(d)$  depending on the day of the week  $d \in \{1, 2, \dots, 7\}$ , which can be computed from the timestamp  $t$ , i.e.,  $d \subset t$ .

$\hat{\phi}(d)$  is produced by linearly combining forecasts from  $n$  forecasting models with  $k$  forecast horizons  $\hat{y}_{ij}(h)$ ,  $i \in \{1, 2, \dots, k\}$  and  $j \in \{1, 2, \dots, n\}$  using the weights  $\omega_i(d)$  specific for each weekday,

$$\hat{\phi}(d) = \sum_{i=1}^k \sum_{j=1}^n \hat{y}_{i,j}(d) \omega_j(d). \quad (2.19)$$

In this case, from the dataset using  $\mathcal{T}$ , only  $r$  data points with  $r \leq p$  corresponding to  $d$  are considered to compute the weights  $\omega_i(d)$ ,

$$\begin{pmatrix} y_1(d) \\ y_2(d) \\ \vdots \\ \hat{y}_r(d) \end{pmatrix} = \begin{pmatrix} \hat{y}_{1,1}(d) & \hat{y}_{1,2}(d) & \cdots & \hat{y}_{1,n}(d) \\ \hat{y}_{2,1}(d) & \hat{y}_{2,2}(d) & \cdots & \hat{y}_{2,n}(d) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{y}_{r,1}(d) & \hat{y}_{r,2}(d) & \cdots & \hat{y}_{r,n}(d) \end{pmatrix} \begin{pmatrix} \omega_1(d) \\ \omega_2(d) \\ \vdots \\ \omega_n(d) \end{pmatrix} + \begin{pmatrix} e_1(d) \\ e_2(d) \\ \vdots \\ e_r(d) \end{pmatrix} = \begin{pmatrix} \hat{\phi}_1(d) \\ \hat{\phi}_2(d) \\ \vdots \\ \hat{\phi}_r(d) \end{pmatrix} + \begin{pmatrix} e_1(d) \\ e_2(d) \\ \vdots \\ e_r(d) \end{pmatrix}. \quad (2.20)$$

The above algorithm is repeated seven times  $\forall d \in \{1, 2, \dots, 7\}$  to compute individual weights for each weekday and forecasting model. For this ensemble  $W \in \mathbb{R}^{n,d}$ .

### 2.2.7 Recursive least squares model ensemble (RLS-ME)

This method only applies to point forecast with one forecast horizon. For real-time application, this model produces an hourly ensemble forecast  $\hat{\phi}(t)$  depending on the time of the forecast  $t$ .  $\hat{\phi}(t)$  is produced by linearly combining forecasts from  $n$  forecasting models,  $\hat{y}_i(t)$ ,  $i \in \{1, 2, \dots, n\}$  using the weights  $w_i(t)$  updated based on the last measurement,

$$\hat{\phi}(t) = \sum_{i=1}^n \hat{y}_i(t) \omega_i(t). \quad (2.21)$$

The model is initialized using the weights computed by the LS-ME method as they represent the best estimate of weights for the individual model. The weights  $w_i(t)$  are updated using the following RLS algorithm [37],

$$\omega(t) = \omega(t-1) + K(t)(y(t) - \hat{y}(t)')\omega(t-1), \quad (2.22)$$

$$K(t) = \frac{P(t-1)\hat{y}(t)}{1 + \hat{y}(t)'P(t-1)\hat{y}(t)}, \quad (2.23)$$

and

$$P(t) = (I - K(t)\hat{y}(t)')P(t-1). \quad (2.24)$$

## 2.3 Data

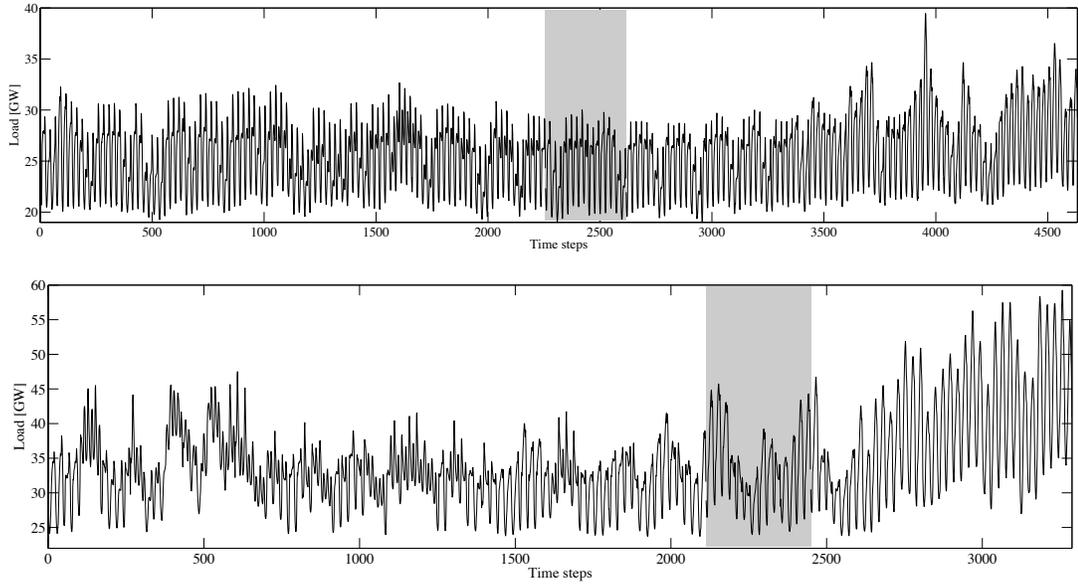
As explained in section 3.3, the collected data was preprocessed and separated into three disjoint datasets: training set ( $\mathbf{T}$ ), validation set ( $\mathbf{V}$ ), and testing set ( $\mathbb{T}$ ). For details see Table 2.3 and Fig. 2.2. The results presented in this study are for the  $\mathbb{T}$  dataset. As mentioned above, we implemented load re-forecast models for two ISOs: CAISO and ERCOT.

**Table 2.3:** Details for the dataset used for HAM and DAM forecast.

Data source	Dataset label	Time-period	Data points
CAISO	$\mathbf{T}$	11/2/2012 to 2/28/2013	2256
	$\mathbf{V}$	3/1/2013 to 3/17/2013	360
	$\mathbb{T}$	3/18/2013 to 6/11/2013	2016
ERCOT	$\mathbf{T}$	12/01/2012 to 04/14/2013	1848
	$\mathbf{V}$	04/15/2013 to 04/30/2013	336
	$\mathbb{T}$	05/01/2013 to 06/20/2013	528

### 2.3.1 California Independent System Operator (CAISO)

CAISO operates the power grid and electricity market for California (CA). It covers almost 80% of the California-Mexico (CAMX) power area [38, 39]. The



**Figure 2.2:** The CAISO time-series (*top*) shows load measurements from November 2, 2012 to June 11, 2013. The ERCOT time-series (*bottom*) shows load measurements for Texas, ranging from December 1, 2012 to June 20, 2013. The measurements on the left of the shaded region represent training set  $\mathbf{T}$ , the shaded portion highlights validation set  $\mathbf{V}$ , and the time-series on the right of the shading region is testing set  $\mathbf{T}$ . The influence of heat and summer weather can be seen in  $\mathbf{T}$  with sudden spikes and increase in load demand. It can be observed that the ERCOT load profile is highly variable and even daily trends are change as the seasons change from winter to summer season.

remaining 20% of CA’s state load is covered by the Los Angeles Department of Water and Power, the Sacramento Municipal Utility District, and the Imperial Irrigation District. Their load forecasts are based on the weather data provided by the NOAA National Weather Service for various stations in CA.

### 2.3.2 Electric Reliability Council of Texas (ERCOT)

ERCOT covers almost 85% of Texas’ load [40]. With centralized control ERCOT is responsible for transmission reliability, wholesale open access, and management of wholesale market for balancing energy and ancillary services for Texas.

## 2.4 Results and discussion

The above methods were implemented in Matlab using the system identification toolbox. The results and discussion presented here are for the re-forecasts and ensemble re-forecasts produced for the validation set, i.e.,  $\hat{y}, \hat{\phi} \in \mathbb{T}$ . For the RLS algorithm, the HAM re-forecast started two days before the validation set to discard the initial effect of sudden unstable changes in weights. However, only the results from  $\mathbf{V}$  were used in the error analysis.

For HAM, the parameters of the CAISO time-series model were selected using minimum description length (MDL) criterion. Using those parameters as reference, similar parameters were selected for ERCOT with small changes (Table 2.4). For DAM, global search was performed using data points in  $\mathbf{T}$  and  $\mathbf{S}$ . Before applying least squares ensemble methods for DAM load, the forecast vector was produced by combining the re-forecast from various models. They are denoted as ARX-E, ARMAX-E, BJ-E, GM-E, and NARM-E to represent their ensemble for DAM load forecasts.

### 2.4.1 Error statistics

Standard error metrics commonly used to compare the performance of the forecast models are: Mean Absolute Percentage Error (MAPE), Mean Absolute

**Table 2.4:** Number of model parameters used for CAISO and ERCOT re-forecast time-series models.

Model Parameters	$n_a$	$n_b$	$n_c$	$n_d$	$n_f$	$n_k$
CAISO-HAM	2	3	5	5	3	24
ERCOT-HAM	2	3	5	5	3	23
CAISO-DAM	2	2	2	2	2	0
ERCOT-DAM	2	2	2	2	2	0

**Table 2.5:** Statistical error metrics for HAM load forecast using various models for CAISO (3-18-2013 to 06-11-2013) and ERCOT (05-01-2013 to 06-20-2013).

Model	CAISO				ERCOT			
	MAPE (%)	MBE (MW)	MAE (MW)	RMSE (MW)	MAPE (%)	MBE (MW)	MAE (MW)	RMSE (MW)
ISOs	1.57	255.64	379.99	519.05	2.20	26.32	798.75	1009.12
ARX	0.88	45.99	224.13	303.66	1.91	286.80	695.64	897.12
ARMAX	0.88	39.06	224.37	303.48	1.74	225.46	630.53	818.15
BJ	0.90	34.63	228.67	306.75	1.77	211.65	644.91	836.70
GM	0.90	-0.03	229.52	308.60	1.68	203.95	614.67	792.07
NARX	0.84	34.60	215.42	292.90	1.79	313.26	650.85	853.72
LS-ME	0.85	31.70	218.09	294.37	1.63	217.28	593.16	769.44
LS-DE	0.89	28.76	225.52	304.72	1.64	214.80	598.57	771.69
LS-HE	0.88	52.43	227.42	300.11	1.39	201.37	518.64	707.58
RLS-ME	0.88	58.65	224.77	302.63	1.70	231.16	650.03	815.37

Error (MAE), Mean Bias Error (MBE), and Root Mean Square Error (RMSE). MAPE measures the accuracy of a method in terms of percentage error. A MAPE of zero implies a perfect fit, but there is no upper bound on its value. The mean of absolute errors is represented by MAE and bias in the forecast is represented by MBE. RMSE measures the average squared distance between the measured and forecasted load. The results after applying these metrics for the proposed models are presented in Table 2.5 and 2.6. The individual models are well-established base models for load forecasting. They were evaluated before combining to compare the re-forecasts using a single model and ensemble model. This gives an insight to the reader how much improvement is achieved by re-forecast using a single model and then further by combining them.

### HAM re-forecast

For CAISO, NARX performed the best with a MAPE of 0.85% which presents a 47% improvement over the internal HAM forecast produced by CAISO.

**Table 2.6:** Statistical error metrics for the DAM load forecast using various models for CAISO (3-18-2013 to 06-11-2013) and ERCOT (05-01-2013 to 06-20-2013).

Model	CAISO				ERCOT			
	MAPE (%)	MBE (MW)	MAE (MW)	RMSE (MW)	MAPE (%)	MBE (MW)	MAE (MW)	RMSE (MW)
ISOs	2.26	370.37	570.55	718.86	3.76	-108.48	1359.71	1663.55
ARX-E	1.64	155.61	437.30	604.51	2.81	430.22	1098.04	1669.98
ARMAX-E	1.99	-52.29	518.97	683.69	2.98	522.77	1173.04	1667.83
BJ-E	1.48	55.71	393.91	558.10	2.76	129.95	1026.14	1357.29
GM-E	1.47	55.25	391.04	545.81	2.00	64.09	776.37	1140.42
NARX-E	1.61	72.91	429.30	597.77	3.00	703.77	1183.02	1778.51
LS-ME	1.47	8.87	391.11	544.35	2.15	49.94	819.27	1142.21
LS-DE	1.54	13.33	407.95	562.36	2.31	143.02	889.64	1283.66
LS-HE	1.55	2.27	411.30	563.74	2.38	293.66	920.19	1337.18

The GM forecast had the lowest bias in the forecast. NARX and LS-ME performance is almost the same in terms of RMSE. Although BJ showed improvement in forecast when compared to CAISO, it performed the worst out of the other re-forecast models.

For ERCOT, the re-forecast showed an improvement of 36% over the existing HAM forecast produced by the ISO. Compared to all the models, LS-HE performed the best with the lowest MAPE of 1.39% and RMSE of 707.58 MW. As in the previous case, all the models showed improvement over the internal forecast provided by the ISO. Amongst all the forecast methods implemented in this study, ARX performed the worst. In terms of MBE, no model was able to perform better than the ERCOT forecast. In both cases, the RLS-ME forecast model performed similar to any other LS-based ensemble model.

### DAM re-forecast

The re-forecasts combined from various GM models (GM-E) showed the best results with a MAPE of 1.47% and 2.00% in comparison to all other methods for both CAISO and ERCOT datasets. In terms of MBE, LS-based ensemble methods performed the best for both cases.

Comparing the results from both HAM and DAM re-forecasts, it can be inferred that ensemble methods are robust for any forecast horizon or error statistics. This highlights the promising application of ensemble re-forecasting for better prediction. Moreover, all the time-series models used are based on the assumption

that  $\epsilon$  is white-noise which is validated using a correlation analysis presented in the next section.

## 2.4.2 Residual analysis

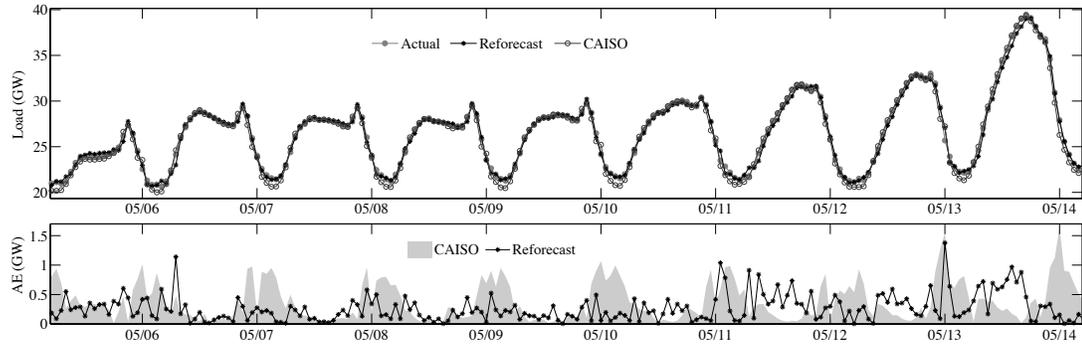
### Correlation analysis

The assumption in time-series forecasting is that the residue of the forecast should be a white noise. If the residues are not white, then the information in the residues can be used to enhance model performance. In our experience the residuals from CAISO and ERCOT forecasts showed a periodic correlation in ACF that served as a motivation to apply re-forecast. The autocorrelation function (ACF) and partial correlation function (PCF) of the forecast residuals produced by the ISO and the re-forecast model for HAM are shown in Fig. 2.4. The residuals from the re-forecast are white and performance of the models shows improvement, which validates the assumption. One drawback of this analysis is that it does not provide any information about the time of the errors which is very important in load forecasts as the errors during peak times have higher implications on the stability of the grid than off-peak times. Thus, a temporal analysis of errors is discussed below.

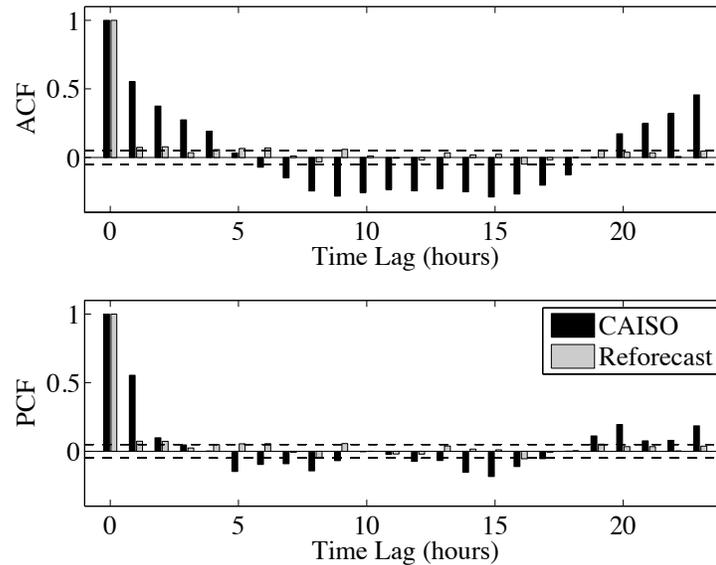
### Temporal analysis

For CAISO load demand, midnight to 5:00 am PDT in the morning is considered super off-peak time, 6:00 am to 12:00 noon and 8:00 pm to midnight PDT is considered off-peak time, and noon to 8:00 pm PDT is considered an on-peak time. The grid is very sensitive during on-peak times because of high demand. Figure 2.5 shows the mean and standard deviation in HAM load forecast errors from CAISO and re-forecast model (NARX). This analysis shows that the CAISO forecast is the worst during super off-peak times. As the time-period changes to off-peak and on-peak times, the forecast improves significantly. Hence, the forecast error is smallest during on-peak times with the lowest standard deviation.

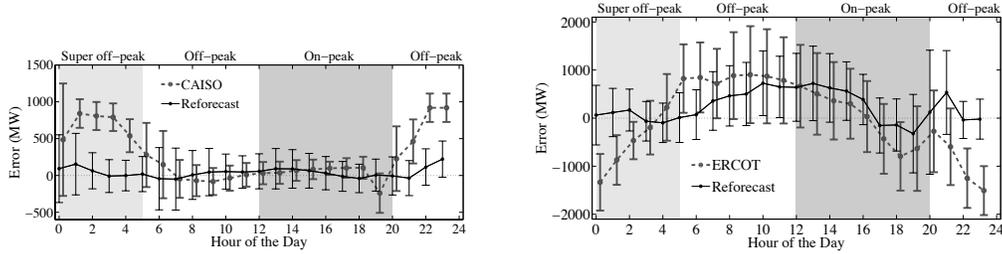
The performance of the re-forecast model is consistent throughout the day.



**Figure 2.3:** Time-series of actual CAISO load and HAM load prediction for several days in 2013 by CAISO and NARX re-forecast model (*top*) and the instantaneous absolute error (*bottom*). The time-series shows change in load profiles for weekdays and weekends. In addition, it can be seen that a day like 05-13-2013 has a different load profile than the rest of the days, which can be attributed to the weather conditions. The CAISO model does not perform well during off-peak hours (22:00-8:00 PDT) and the forecast for this period is corrected by the re-forecast model.



**Figure 2.4:** Autocorrelation and partial autocorrelation for the residuals from CAISO and NARX re-forecast models for HAM load prediction. The dashed line represents the 95% confidence interval. It can be noticed that the residuals from re-forecast model are white whereas the ACF for CAISO residuals is greater than 0.05 which implies the presence of structured systematic errors. Similar pattern was observed for the ERCOT dataset.



**Figure 2.5:** Temporal distribution of the mean error (actual load-forecast HAM load) over the day for CAISO (*left*) and ERCOT (*right*). The error-bar shows the standard deviation in error for that hour of the day where 0 represents a perfect forecast with 0 MW error. It can be noticed that the accuracy of the CAISO model significantly improves during on-peak times than super off-peak and off-peak times whereas the performance of the re-forecast model (NARX) is consistent over the day. A similar pattern can be observed for ERCOT where re-forecast (LS-HE) shows improvement only for partial hours during on-peak times (5:00 pm to 9:00 pm) and significant improvement during super off-peak and off-peak times.

The re-forecast model significantly improves the forecast in super-off peak and off-peak times. During on-peak times, the re-forecast shows improvement only for certain hours (6:00 pm to 8:00 pm PDT). The variance in the CAISO forecast is smaller than the re-forecast with almost the same mean value. It can be inferred that for on-peak times CAISO forecast is the best (12:00 noon- 6:00 pm PDT) compared to any re-forecasts whereas for all other times the re-forecast has potential to improve the forecast produced by the ISO. Similar results were found for ERCOT (see Fig. 2.5) and DAM load forecasts.

## 2.5 Summary

We proposed a re-forecast methodology to enhance performance of STLF load prediction by extracting the information from the structured non-random errors in the given estimate. The re-forecast ensembles consist of various time-series models combined using least squares optimization. Different ensemble combinations, with the weights optimized based on global, hourly, and weekday performance of the models, are proposed and investigated in this study. For real-time load prediction, the weights are updated at every time-step new data becomes

available using recursive least squares. The results are presented for two types of load forecasts: HAM and DAM load for both CAISO and ERCOT. For HAM, point load forecast is produced at every hour. For DAM point forecasts are produced at midnight for all twenty-four hours of the next day. The re-forecast results showed an improvement over the internal forecast provided by the ISOs.

For HAM, NARX and LS-ME performed the best for CAISO. For ERCOT, LS-HE performed the best with a MAPE of 1.49% as compared to ERCOT forecast MAPE of 2.24%. RLS-ME improved forecast accuracy with a MAPE of 0.88% for CAISO and 1.70% for ERCOT. Similarly, for DAM load prediction, GM-E and LS-ME performed the best for both CAISO and ERCOT dataset. Therefore, we can conclude that the methods proposed in this study can significantly enhance the accuracy of load prediction. This facilitates prospective improvements in all aspects of grid management and operations.

Correlation analysis of the forecast residuals from ISOs revealed structured non-random errors in their estimates. Application of re-forecast refined the forecast with white noise errors. Temporal analysis of the residuals showed that the internal forecasts by the ISOs have the highest error during off-peak times and their forecast accuracy significantly improves for on-peak times. The re-forecast refines the forecast with significant improvements during off-peak times and small improvements during on-peak times. Similar results from two ISOs with different load profiles and forecast horizons validates the robustness of our models. Hence, the proposed methodology is applicable to ISOs and power utilities, and has the ability to substantially enhance the accuracy of current load forecasts.

## Acknowledgements

This chapter contains work previously published in:

- A. Kaur, H.T.C. Pedro, and C.F.M. Coimbra, “Ensemble re-forecasting methods for enhanced power load prediction”, *Energy Conversion and Management* 80 (0) (2014) 582 - 590.

# Chapter 3

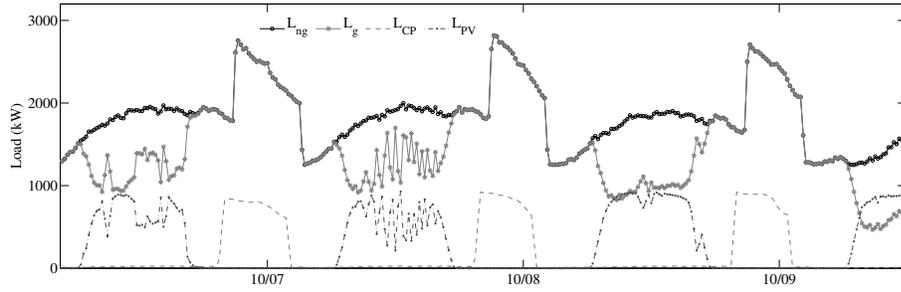
## Impact of increasing solar penetration on load forecasting

California generated approximately 20% of its in-state power using renewable energy resources in 2011 [41]. 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 [42, 43]. 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)[3], 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 [44] – see Fig. 3.1. 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 [42, 45, 46]. Accurate forecasts of both demand and supply profiles are being pursued 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 [47]. Several reviews on load forecasting methods have been published [15, 18, 10, 48, 12, 49, 50, 7, 13, 5] and there is continuous research to develop better methods [51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63]. Because load profiles are expected to change with onsite renewable energy generation especially with intermittent solar generation [45]. It is very important to analyze the forecast model performance for such scenarios. Therefore, two cases were studied for 15-minute and 1-hour 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. 3.1 the former is greatly affected



**Figure 3.1:** 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.

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 [23]. Like previous works that have studied the error distribution for wind forecasts [64, 65, 66, 67], we characterized the error distribution of our predictions in order to understand the impact of additional variability in forecast accuracy.

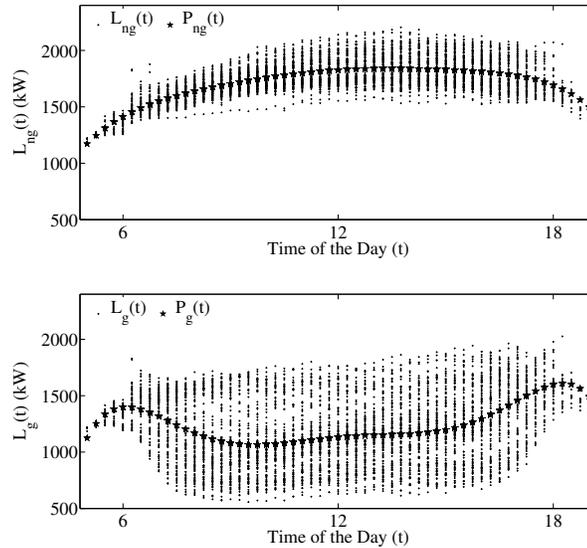
The data used for this study is presented in section 3.1. The models are described in section 3.2. Results and discussion are presented in section 3.3, where the accuracy of the forecasting models is evaluated and compared using standard statistical error metrics. The forecast error distribution is presented 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 3.4.

### 3.1 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 minute backward averages.

#### 3.1.1 Preprocessing data

The time-series was decomposed by removing the daily trend calculated over the whole year. A 6<sup>th</sup> order polynomial ( $P$ ) was fit to both cases as shown in Fig. 3.2. The detrended demand loads can be represented as no onsite generation  $L_{ng}^{dt}(t) = L_{ng}(t) - P_{ng}(t)$  and onsite generation  $L_g^{dt}(t) = L_g(t) - P_g(t)$ .



**Figure 3.2:** 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 6<sup>th</sup> order polynomial ( $P$ ) to daily load profile over the whole year.

### 3.1.2 Training and testing 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 Jan 2010 to Dec 2010 minus a week from each month.
2. Testing 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).

## 3.2 Load forecast methods

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 ( $q^{-N}$ )[36]

$$q^{\pm N} L(t) = L(t \pm N) \tag{3.1}$$

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

### 3.2.1 Persistence-Load ( $P_L$ )

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

$$\hat{L}(t) = q^{-N} L(t) \quad (3.2)$$

where  $N$  is 1 for the 15-minute and 4 for the 1-hour forecast horizons, respectively.

### 3.2.2 Smart Persistence-Load ( $SP_L$ )

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

1.  $SP1_L$ : this model assumes that the difference, with respect to the trend, persists to the future. In mathematical terms it can be represented as

$$\hat{L}(t) = q^{-N} L(t) - q^{-N} P(t) + q^N P(t). \quad (3.3)$$

2.  $SP2_L$ : 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

$$\hat{L}(t) = \frac{q^{-N} L(t)}{q^{-N} P(t)} q^N P(t). \quad (3.4)$$

### 3.2.3 Autoregressive (AR)

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 [36]

$$L(t) + a_1L(t-1) + \dots + a_{n_a}L(t-n_a) = e(t). \quad (3.5)$$

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

$$A(q) = 1 + a_1q^{-1} + \dots + a_{n_a}q^{-n_a}, \quad (3.6)$$

the AR-model becomes

$$A(q)L(t) = e(t) \quad (3.7)$$

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

### 3.2.4 Autoregressive Moving Average (ARMA)

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

$$\begin{aligned} L(t) + a_1L(t-1) + \dots + a_{n_a}L(t-n_a) = e(t) + c_1e(t-1) + \\ \dots + e_{n_c}L(t-n_c), \end{aligned} \quad (3.8)$$

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

$$A(q) = 1 + a_1q^{-1} + \dots + a_{n_a}q^{-n_a}, \quad (3.9)$$

$$C(q) = 1 + c_1q^{-1} + \dots + c_{n_c}q^{-n_c}, \quad (3.10)$$

can be rewritten as

$$A(q)L(t) = C(q)e(t). \quad (3.11)$$

### 3.2.5 k-th Nearest Neighbor (kNN)

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 [68]. 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  $A_{ij}$

$$e_i = \sqrt{\sum_j (A_{ij} - p_j)^2}. \quad (3.12)$$

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

$$\hat{L}(t) = \frac{\sum_{i=1}^N w_i L_{train,i}}{\sum_{i=1}^N w_i}, \quad (3.13)$$

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  $w_i$ .

### **kNN1**

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

### **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,  $\vec{w} = \{10, 9, \dots, 2, 1\}$ .

### 3.2.6 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 ( $s$ ) [69] is used to compare the accuracy of the forecasting models with respect to the persistence model

$$s_{\text{Model}} \approx 1 - \frac{\text{RMSE}_{\text{Model}}}{\text{RMSE}_{\text{Persistence}}}. \quad (3.14)$$

## 3.3 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 3.1 and for ARMA, the coefficients for  $A(q)$  and  $C(q)$  are listed in Table 3.2.

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.

**Table 3.1:** 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-hour	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-minute	no	$1 - 0.79q^{-1} - 0.11q^{-3} - 0.06q^{-4}$
	yes	$1 - 0.92q^{-1} + 0.05q^{-2} - 0.08q^{-3}$

**Table 3.2:** 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-hour	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-minute	no	$1 - 1.14q^{-1} + 0.15q^{-2}$	$1 - 0.35q^{-1} - 0.13q^{-2}$
	yes	$1 - 0.95q^{-1}$	$1 - 0.03q^{-1} - 0.08q^{-2}$

### 3.3.1 1-hour forecasts

The results for various statistical error metrics for this forecasting horizon are tabulated in Table 3.3. 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-hour 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 3.3). 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

**Table 3.3:** Statistical error metrics for the 1-hour 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 <sup>2</sup>	s	MAE(kW)	MBE(kW)	RMSE(kW)	R <sup>2</sup>	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	0.00
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 3.4:** Statistical error metrics for 15-minute demand load forecasting Without onsite generation and with onsite solar generation.

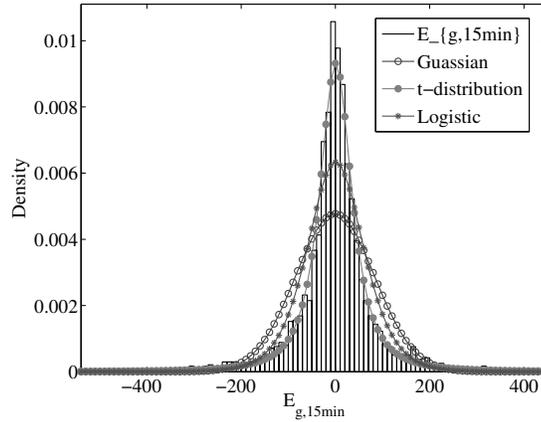
Model	Without onsite generation					With onsite solar generation				
	MAE (kW)	MBE(kW)	RMSE (kW)	R <sup>2</sup>	s	MAE(kW)	MBE(kW)	RMSE(kW)	R <sup>2</sup>	s
P	20.08	1.93	26.54	0.98	0	62.81	8.82	101.12	0.92	0
SP1	18.86	-0.11	24.71	0.98	0.06	60.58	2.28	98.24	0.92	0.03
SP2	18.96	-0.13	24.79	0.98	0.06	59.91	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

=-0.003. The reason for such poor performance could be related to the absence of similar patterns in the training set.

### 3.3.2 15-minute forecasts

The results for various statistical error metrics for this forecast horizon are tabulated in Table 3.4. For this horizon and no onsite 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-minute 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 penetra-



**Figure 3.3:** Error distribution for 15-minute 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.

tion, the basic assumption behind the persistence model does not hold because of variability in solar power output for the 15-minute 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-hour 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.

### 3.3.3 Error distribution

In order to analyze the forecast error distribution for the two cases, we started by computing the mean forecasting error  $E_{avg}$ , that is the average of the error for all the models. With this approach the resulting differences in  $E_{avg}$  depend solely on the variability of the time-series and the forecast horizon. Given

**Table 3.5:** Statistics of average error distribution for no onsite generation and onsite solar generation case for 1-hour and 15-minute lead times.

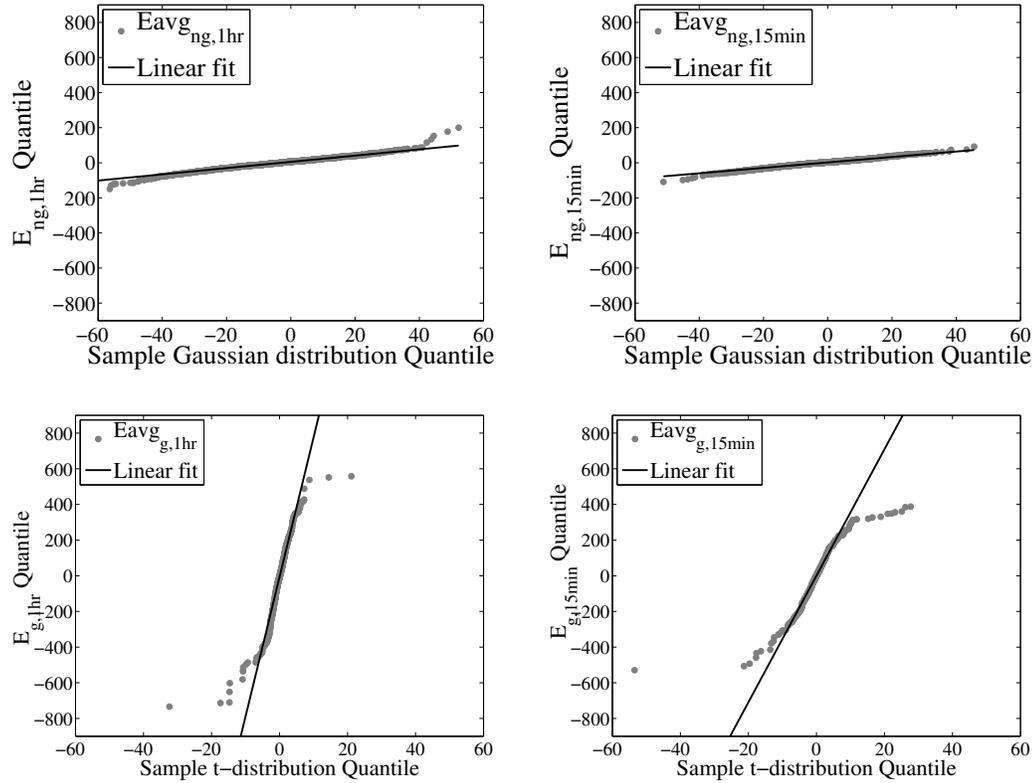
Moment	$Eavg_{ng,1hr}$	$Eavg_{g,1hr}$	$Eavg_{ng,15min}$	$Eavg_{g,15min}$
$\mu$	-7.05	-13.27	-2.94	-0.79
$\sigma$	33.88	137.12	23.04	83.58
$s$	-0.06	-0.46	-0.25	-0.38
$k$	4.92	6.01	3.66	9.15

that we have two cases and two forecast horizons, this operation results in 4 different error time-series. The mean ( $\mu$ ), standard deviation ( $\sigma$ ), skewness ( $s$ ) and kurtosis ( $k$ ) for each case are tabulated in Table 3.5. As expected the  $\sigma$ ,  $s$  and  $k$  values are much higher for the high solar penetration scenario.

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.

In order to identify the best matching distribution, various distributions were compared with the error distribution as shown in Fig. 3.3. We used a Q-Q plot (where Q stands for “Quantile”) [65, 70] 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. 3.4 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  $\mu$  and  $\sigma$  of the  $Eavg$  for each case, and Degrees of Freedom ( $v$ ) for the sample t-distribution quantile were computed based on the Maximum Likelihood principle.

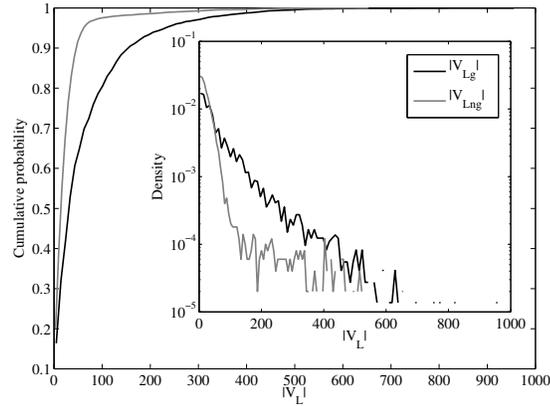


**Figure 3.4:** Error distribution for 1-hour and 15-minute load forecasting with and without solar penetration. With no solar penetration the error distribution can be perfectly characterized by the Gaussian distribution as shown in the first and second plot. Whereas with solar penetration the error distribution can be better characterized by t-distribution. Third and fourth plot shows the sample t-distribution Q-Q plots for 1-hour and 15-minute cases with  $v = 2.6$  and  $v = 1.8$ .

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. 3.3, 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. 3.4. These findings are in agreement with the results published for wind error forecasting distribution [64, 65, 66, 70].

From this analysis we can observe the effect of the solar power on the

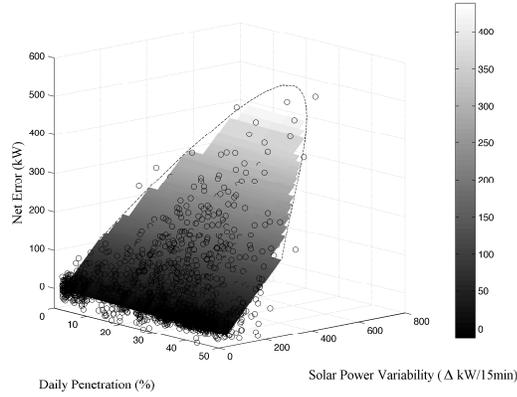
distribution of the forecasting error. This effect can also be explained by analyzing the Cumulative Distribution Function (CDF) of normalized 15-minute step change in load with ( $V_{Lg}$ ) and without ( $V_{Lng}$ ) onsite solar generation as shown in Fig.3.5. It can be observed that the probability for  $V_{Lng} > 100$  kW/15min 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.



**Figure 3.5:** 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/15min 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.

### 3.3.4 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 the solar variability is always mentioned as one of the major challenges in promoting higher levels of solar penetration



**Figure 3.6:** 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 minute forecast and the surface plot shows the computed surface fit between the error, daily solar penetration and solar power variability for 15 minute 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.

[42]. For this purpose we defined daily solar penetration ( $S_{PV}(d)$ ) and solar power variability ( $V_{PV,N}$ ) as

$$S_{PV}(d)(\%) = \frac{\sum_{t \in DT} L_d(t)}{\sum_{t \in DT} L_{PV}(t)} (\%) \quad (3.15)$$

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 ( $d$ ).

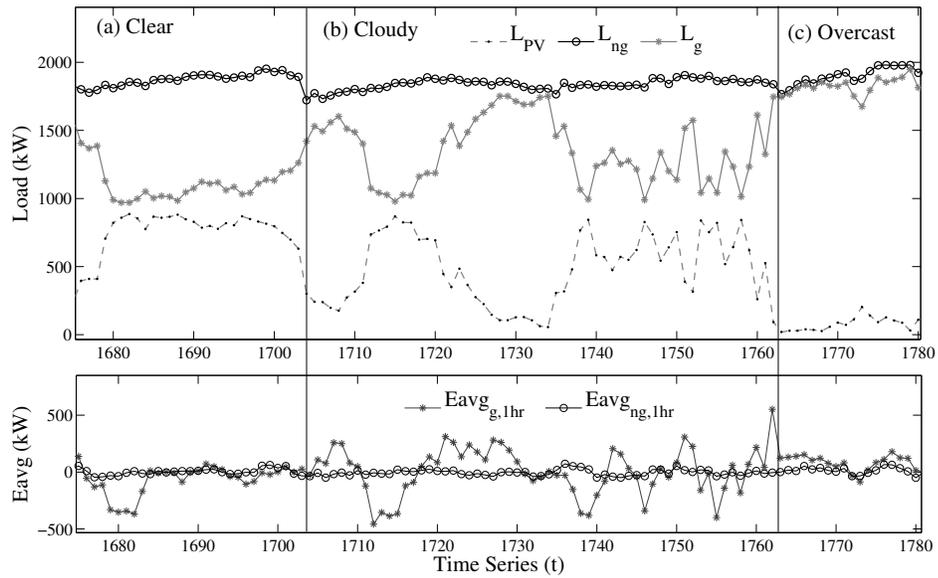
$$V_{PV,N} = \sqrt{\frac{1}{N} \sum (\Delta L_{PV}(t) - \overline{\Delta L_{PV}(t)})^2} \quad (3.16)$$

where  $N = 4$  for the 1 hour forecast and  $N = 1$  for the 15-minute 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  $\overline{\Delta L_{PV}(t)}$  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. Figure 3.6 shows the distribution of net error as a function of increasing solar penetration with different levels of solar variability for the 15 minute horizon. The solar power variability for 1 hour horizon varies between 0.01 kW/hr

to 399.58 kW/hr and for 15 minute horizon it varies from 0 kW/15 min to 800 kW/15 min. The smaller range for the 1 hour horizon is due to the fact that as the time horizon increases solar variability reduces due to the averaging. For the 1-hour onsite solar generation case the average error values vary between -153.65 kW and 200.11 kW and for 15 minute they vary between -109.35 kW and 91.92 kW. The daily solar penetration varies between 3% to 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. 3.6 shows the maximum magnitude of solar 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. 3.6 shows that error magnitude increases with increasing solar variability but not with increasing solar penetration.

Finally, Fig. 3.7, 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 [71] it was shown that variability in PV is directly proportional to variability in Global Horizontal Irradiance (GHI). Combining the result from [71] 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.



**Figure 3.7:** Time-series of load and error profiles for a clear, cloudy and overcast day for 1 hour 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.

### 3.4 Summary

The performance of STLF models with and without onsite generation is compared. For the 1-hour 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-minute 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 [67]. For no onsite generation case, the assumption of persistence holds for the 15-minute 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. This highlights the need for solar forecasting. 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

This chapter contains work previously published in:

- A. Kaur, H.T.C. Pedro, and C.F.M. Coimbra, “Impact of onsite solar gener-

ation on system load demand forecast", Energy Conversion and Management  
75 (0) (2013) 701 - 709.

# Chapter 4

## Solar resource forecasting

Increasing awareness about the adverse effects of conventional energy resources have accelerated the penetration of renewable energy resources. Although solar energy has a high potential to meet human energy needs, the variable and uncertain nature of solar energy has impeded the growth of large-scale solar farms. Solar irradiance forecast is a viable solution to cope with this stochastic nature of solar energy and a necessary tool to manage and operate large solar farms in deregulated electricity market. Utilizing the forecasts, the large-scale solar farm operators can bid predicted solar energy in the electricity market and make smart decisions about curtailment and storage systems to avoid sudden ramps in the solar power output. Depending on the forecast horizon, different methods are applied for solar forecasting. For intra-week forecasts, Numerical Weather Prediction (NWP) models are known to generate good results. This usually holds true for horizons greater than 4 hours. For intra-day forecasts below 4 hours, mostly satellite image based prediction methods are used [72]. For intra-hour forecasts, local sky conditions and sky images are utilized along with other available meteorological data for the site [73, 21, 22, 24, 74, 75, 76, 77, 78]. Furthermore, the forecast models with no external inputs are also suggested [23] for solar power for intra-hour forecasts where information is derived from the various characteristics in the time-series. Also, forecast models are suggested that selects inputs based on spatial and temporal distributions [79]. A detailed review on solar forecasting methodologies can be found in [19].

Datasets used in this study are described in section 4.1, forecasts models are defined in section 4.2, results are provided in section 4.3 and and conclusions are drawn in section 4.4.

## 4.1 Data sets

### 4.1.1 NWP data

Day-ahead Global Horizontal Irradiance (GHI) forecasts generated at the 00:00 coordinated universal time (UTC) are downloaded from National Oceanic and Atmospheric Administration (NOAA) servers for December 2012 to December 2013 and degraded for Folsom, CA. This forecast is generated with the Numerical Weather Prediction (NWP) based on the North American Model (NAM). Forecast horizons ranging from 9 - 35 hours with hourly time-resolution are used. The performance of the NAM model is extensively evaluated. A general over-prediction of GHI is well known [80, 81, 82]. A reforecasting technique is applied on the GHI time-series forecast to remove the bias and structured errors. To evaluate the model performance, the data set is divided into three disjoint data sets: training set (12-20-2012 to 1-15-2013), validation set (1-16-2013 to 1-31-2013) and the test set (2-1-2013 to 12-31-2014). The training set is used to train the models, the validation set is used for optimization as well as for feature selection for the forecast model. The testing set is always kept as an independent set to access the model performance and report results.

### 4.1.2 Ground data

GHI and Direct Normal Irradiance (DNI) data were collected at Folsom, CA, (located at  $38.63^{\circ}$  N and  $121.14^{\circ}$  W) using a rotating shadow-band radiometer (RSR). A CR1000 data logger from Campbell Scientific running on a one minute sampling rate was used to store the data. The sensor of the RSR is a Licor-200SZ photodiode, periodically shaded to provide diffuse irradiance values. DNI values are calculated from GHI and diffuse irradiance values, with a program embedded

in the data logger. Sky imagery was acquired with an off-the-shelf fish-eye lens security camera (Vivotek, model FE8171V) fixed on a horizontal surface and pointed to the zenith, providing one picture every minute. The total costs for the deployed instruments are below \$10,000 USD. The data sets of irradiance and sky imagery cover December 2012 to December 2013.

### Data preprocessing

The collected ground data and sky images are pre-processed to represent 5 and 15 minute averages. Irradiance time-series and images are time-matched and divided into three disjoint data sets: training, validation and testing set as discussed above. In this case, each data set contains subsets of all available months to capture the full seasonal variation and various sky conditions.

### Feature definitions

Using the time-series and the sky images, features are calculated to be used as an input for the forecast models. To calculate the features, the clearness index for GHI and DNI is defined as  $k_t = \frac{GHI}{GHI_{cs}}$  and  $k_b = \frac{DNI}{DNI_{cs}}$ . The computed features include entropy, backward averages, and variability for both  $k_t$  and  $k_b$ .

Entropy is defined as,

$$E_i = - \sum_{j=1; p_{ij} \neq 0}^n p_{ij} \log_2(p_{ij}), \quad i = (1, 2, \dots, 24) \quad (4.1)$$

where  $p_{ij}$  is the relative frequency for the  $j$ th bin out of the 200 bins in the range  $[0,2]$  for data in the interval  $[t - i\delta, t]$  and  $\delta = 5$  minutes is the minimum window size. The index  $i$  ranges from 1 to 24, indicating the smallest window is 5 minutes and the largest is 120 minutes.

Backward averages are defined as,

$$B_i(t) = \frac{1}{N} \sum_{t \in [t-i\delta, t]} k_t(t), \quad i = (1, 2, \dots, 24) \quad (4.2)$$

where  $N$  is the number of data points in the interval  $[t - i\delta, t]$ .

Variability is defined as:

$$V_i(t) = \sqrt{\frac{1}{N} \sum_{t \in [t-i\delta, t]} \Delta k_t(t)^2}, \quad i = (1, 2, \dots, 24) \quad (4.3)$$

where  $\Delta k_t(t) = k_t(t) - k_t(t - \delta t)$ .

The imagery features include entropy, mean, and standard deviation for blue channel, green channel, red channel, red/blue ratio, and normalized red/blue ratio. All imagery features have minimum window sizes of 1 minute, maximum window sizes of 10 minutes, window increments of 1 minute, and feature length of 24. More details can be found in [83].

### 4.1.3 Solar power modeling

The PVwatts tool, available from the National Renewable Energy Laboratory (NREL), is applied to get solar data for 1 MW fixed, (open rack) array, commercial solar plant for the same location as the irradiance data sets. The characteristics chosen for the power plant are: array tilt = 20 degrees, array azimuth = 180 degrees, and system losses = 14%. The data provided by this software is hourly data. A model is derived between the solar power produced and GHI to be able to work with a higher temporal resolution.

The data sets generated by PVwatts are divided into two disjoint data sets. One data set is used to derive the curve fit for the model, whereas the second data set is used to test the goodness of fit. The model fit  $m$  is computed as a function of ground GHI and day of the year (DOY) i.e.  $m = f(\text{GHI}, \text{DOY})$ . DOY is used to account for the seasonal variations. The fit was calculated using polynomial fitting with degree 6. The results obtained using the fit function are given in table 4.1. Results show that the coefficient of determination is 1, with very low MBE and RMSE. This means that all the ramps in solar power are captured by the model fit. The derived fit is used to model solar power  $P$  for the given location, which is then used in the analysis to draw conclusions from the perspective of a solar power producers. The solar power forecast  $\hat{P}$  is computed as  $\hat{P} = f(\widehat{\text{GHI}}, \text{DOY})$ .

**Table 4.1:** Statistical error metrics for GHI to PV output modeling for 1 MW<sub>p</sub> capacity fixed-array PV plant.

Model	MAE (kW)	MBE (kW)	RMSE (kW)	rRMSE -	R <sup>2</sup> -
GHI-PV model	9.27	-0.71	16.81	9.21	1.00

## 4.2 Solar forecast methods

### 4.2.1 Persistence (P)

The persistence model is used as the basic reference model. It is based on the assumption that the current conditions will persist so that,

$$\widehat{I}_p(t+k) = I(t), \quad (4.4)$$

where  $\widehat{I}_p(t)$  represents a GHI prediction from the persistence model,  $k$  represents the forecast horizon, and  $I(t)$  is the measured GHI value at time  $t$ .

### 4.2.2 Smart Persistence (SP)

The smart persistence model is based on the same assumption as persistence model but it corrects for the deterministic diurnal variation in solar irradiance. It is defined as,

$$\widehat{I}_{sp}(t+k) = k_t(t) * I_{CS}(t+k), \quad (4.5)$$

where  $\widehat{I}_{sp}(t)$  represents a GHI prediction, and  $I_{CS}$  represents the estimated clear sky solar irradiance [84, 85, 86].

### 4.2.3 Support Vector Regression (SVR)

Support vector regression is a machine learning technique [87], [88]. Using a set of training inputs  $U$ , the objective is to find a function  $f(u)$  using weights  $w$  that has  $\epsilon$  deviation from the actually obtained targets  $k_t$

$$f(u) = \langle w, u \rangle + b \quad \text{with} \quad w \in U, b \in \mathbb{R}. \quad (4.6)$$

In this study, time-series formulation is applied  $k_t(t) = f(u(t))$ . To select the inputs and parameters for SVR model, Genetic Algorithms are applied. Hence, this method is referred to as SVR-GA.

Genetic algorithm is solution space search technique inspired in natural selection and the survival of the fittest [89, 90]. The algorithm starts with a population of individuals that encodes the parameters that determine an individual layout in the population. In this work the parameters consist of binary values that control the inclusion/exclusion of the various input variables, and real values that determined the SVR parameters. The MSE between the measured values and the forecasted values is used as the fitness of the GA individual. The GA optimizes these parameters by evolving an initial population based on the selection, crossover and mutation operators with the objective to minimize the MSE forecast error.

#### 4.2.4 Reforecast

The reforecast model proposed for refining day-ahead load prediction by correcting for structured errors and bias in the forecast is applied here [91]. We apply the auto-regressive (AR) model which is a linear, time-series regression model. Using this model, the output can be expressed as a linear combination of past outputs and measured values,

$$A(q)I_r(t) = B(q)I_{NAM}(t - n_k)(t) + e(t) \quad (4.7)$$

where  $q^N$  is the shift operator,  $q^{\pm N} = I(t \pm N)$ ,  $A(q)$  that represents the combination of past values of output using coefficients,  $a_1, a_2, \dots, a_{na}$ ,  $A(q) = 1 + a_1q^{-1} + \dots + a_{na}q^{-na}$ ,  $B(q) = b_1 + \dots + b_{nb}q^{-nb+1}$  and  $n_k$  is time delay parameter. The part of the time-series not modeled is represented as  $e(t)$ . For more details refer to [36]. In this study, the external input  $I_{NAM}(t)$  consists of GHI forecast values predicted by NAM.

### 4.3 Forecast results

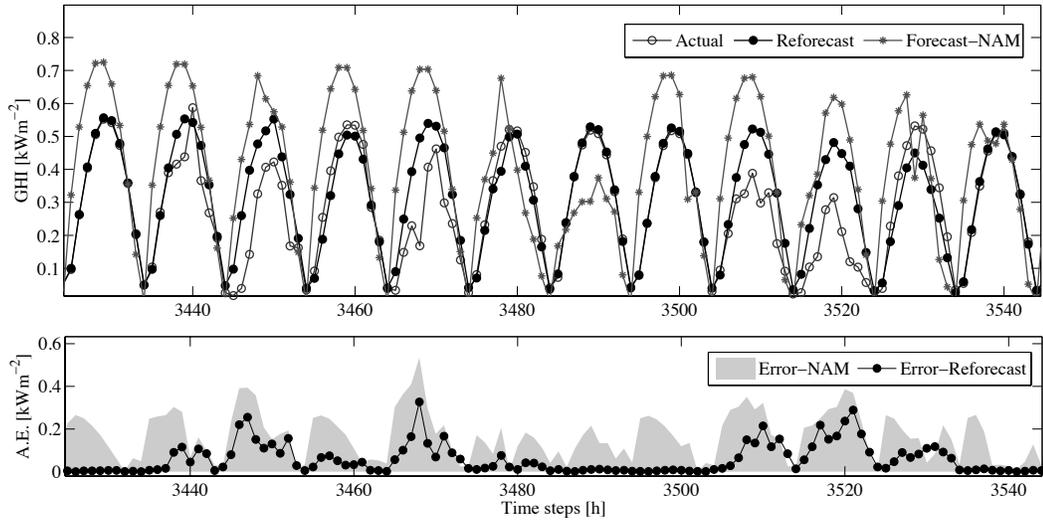
The results obtained using the test set are reported and discussed. The statistical metrics used to compare the forecast results are: Mean Absolute Error (MAE), Mean Bias Error (MBE), Root Mean Square Error (RMSE), relative RMSE (rRMSE), standard deviation ( $\sigma$ ), coefficient of determination ( $R^2$ ) and forecast skill ( $s$ ). MAE provides the mean of the absolute forecast error, MBE gives information about the bias in forecast error (Measured - Forecast). A negative MBE error indicates over prediction whereas positive MBE indicates under-prediction. RMSE gives information about the spread of the forecast errors, rRMSE is the RMSE normalized by mean annual GHI, and  $R^2$  shows the relation between actual and forecast value. For a perfect forecast  $R^2 = 1$ . Forecast skill is a measure to compare the performance of the proposed model against the benchmark model [92].

#### 4.3.1 Intra-week forecast (day-ahead)

For the day-ahead reforecast model, the Akaike information criterion (AIC) was applied to select the number of lagged inputs for the model and  $n_a = n_b = n_k = 1$  were selected as the model parameters. Using the NAM forecast issued at 00:00 UTC as an input and actual hourly GHI measured at ground, the day-ahead forecasts are issued at midnight for the following day.

As stated above, the NAM model tends to over-predict irradiance. For the chosen location, the performance of the NAM model is even lower than that of the smart persistence model. To correct for this bias, a reforecast model was applied (see figure 4.1). The reforecast model corrects the bias by learning from continuous, structured errors, occurring in the forecast. After applying the reforecast, the bias was reduced by 67.21% and the deviation in forecast errors was reduced from  $151.72\text{Wm}^{-2}$  to  $78.91\text{Wm}^{-2}$ .

The distribution of forecast errors for the hours of day are shown in figure 4.2. An analysis of errors reveals that the NWP tends to overestimated GHI before noon and underestimates the GHI after 14:00 PST. After the application



**Figure 4.1:** Sample day of GHI forecast by NAM and reforecast model. The reforecast model corrects for the bias in the NAM forecast. If ramps are not predicted in NAM, it is unlikely that the reforecast model captures them. Both, NAM and reforecast model, tend to over predict. The bottom plot shows the reduction in absolute error. For clear days, the error is close to zero. Most high magnitude errors occur during overcast and cloudy conditions.

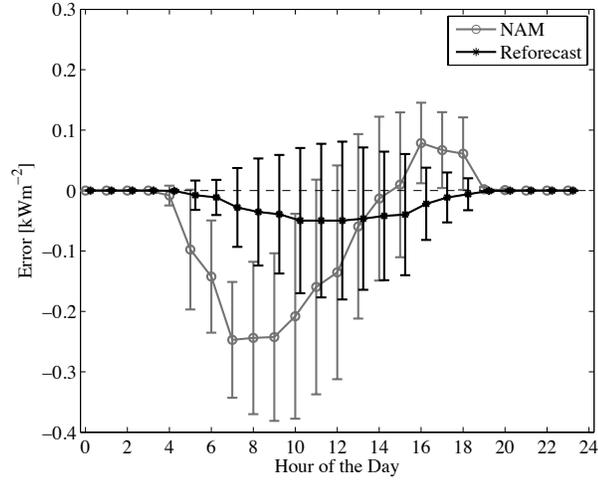
**Table 4.2:** Statistical error metrics for day-ahead GHI forecasts in hourly intervals.

Model	MAE ( $\text{Wm}^{-2}$ )	MBE ( $\text{Wm}^{-2}$ )	RMSE ( $\text{Wm}^{-2}$ )	$\sigma$ ( $\text{Wm}^{-2}$ )	$R^2$ -	Skill s(%)
SP	37.04	-0.69	96.07	96.08	0.91	-
NAM	109.86	-69.65	166.94	151.72	0.72	-
NAM-Reforecast	31.66	-22.84	82.14	78.91	0.93	14.5

of the reforecast model, this bias is reduced. The magnitude of bias in reforecast errors remains unchanged over the day except for sunrise and sunset times. This consistent error can be accounted for during market operations and reserve allocation.

### 4.3.2 Intra-hour forecast

The intra-hour forecasts are implemented using a well known, open source machine learning library (LIBSVM) [93]. For intra-hour forecasts, the optimization is performed once for one step ahead prediction. The features and lag of the



**Figure 4.2:** Mean and standard deviation of day-ahead forecast errors for GHI as a function of hour of the day. The NAM tends to over-predict until noon and then under-predicts later. The reforecast model corrects for the bias in the forecast and achieves consistent performance throughout the day.

**Table 4.3:** Features selected by GA for 15 minute and 5 minute forecast intervals for the SVR-GA model.

15-minute intervals		5-minute intervals	
Features	Lag	Features	Lag
GHI, Backward Average	6	GHI, Backward Average	2
GHI, Lagged values	3	GHI, Lagged Values	1
Image, Blue Average	1	Image, Blue Average	1
Images, nRedBlue Entropy	2	Image, RedBlue Entropy	8
Image, Red Entropy	1	Image, Red Entropy	1
Image, Red Std	3	DNI, Lagged values	1
DNI, Backward Average	3	DNI, Variability	1
DNI, Lagged Value	3	-	-

**Table 4.4:** Statistical error metrics for short-term solar irradiance forecast in 15 minute intervals.

Forecast horizon (minutes)	Persistence				Smart Persistence					SVR-GA					
	MAE ( $Wm^{-2}$ )	MBE ( $Wm^{-2}$ )	RMSE ( $Wm^{-2}$ )	rRMSE (-)	MAE ( $Wm^{-2}$ )	MBE ( $Wm^{-2}$ )	RMSE ( $Wm^{-2}$ )	rRMSE (-)	Skill-P (%)	MAE ( $Wm^{-2}$ )	MBE ( $Wm^{-2}$ )	RMSE ( $Wm^{-2}$ )	rRMSE (-)	Skill-P (%)	Skill-SP (%)
15	24.95	1.59	44.17	9.32	15.01	0.35	41.21	8.70	6.70	13.55	0.97	35.22	7.43	20.26	14.53
30	36.41	2.41	57.69	12.15	19.36	0.58	51.18	10.78	11.28	16.86	0.54	42.56	8.97	26.22	16.64
45	46.63	2.99	68.24	14.36	22.26	0.79	57.20	12.04	16.18	19.01	0.57	46.66	9.82	31.62	18.42
60	56.18	3.35	77.61	16.32	24.38	0.99	61.19	12.87	21.16	20.53	0.04	49.25	10.36	36.54	19.51
75	65.43	3.50	86.63	18.21	26.06	1.22	64.14	13.48	25.96	21.74	0.04	51.24	10.77	40.85	20.11

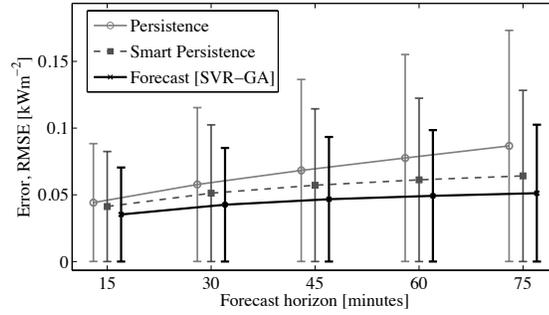
features selected by the GA are given in table 4.3. General practice in forecasting is to propagate the 1 step ahead forecast into the forecast model as an input to forecast the next steps. Doing so, the errors in the forecasts are also propagated into next intervals and the forecasts become correlated. To avoid this correlation between various forecasts, the features selected by the GA are used to train the models specific for each forecast horizon and forecast interval. This ensures that each forecast is independent of each other and they can be studied independently. The persistence and smart persistence models are implemented according to equation 4.4 and 4.5.

### 15-minute interval forecasts

The forecasts in 15 minute intervals are produced for forecast horizon: 15, 30, 45, 60, and 75 minutes. The statistics on the forecast errors are reported in table 4.4. The trend in RMSE and standard deviation of the forecast errors for persistence, smart persistence and state of art forecast model is shown in figure 4.3. With increasing forecast horizon, the RMSE and standard deviation increases linearly for the P model, whereas in case of SP and SVR-GA, after the forecast horizon of 45 minutes, the RMSE converges with small change in magnitude. The skill achieved by SVR-GA model, with respect to the P and SP models, ranges from 20.26-40.86% and 14.53-20.11% respectively.

### 5-minute interval forecasts

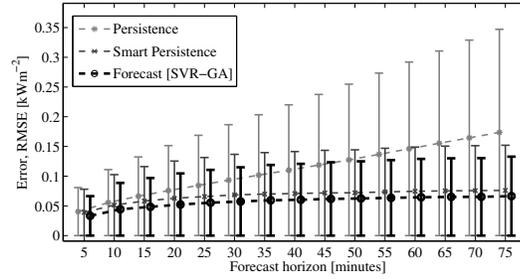
The 5 minute interval forecasts are produced up to a forecast horizon of 5 to 75 minutes. The error statistics for all the forecast horizons are provided in table



**Figure 4.3:** RMSE and standard deviation for short-term GHI forecasts for the 15 minute forecast interval with forecast horizon ranging from 15 to 75 minutes. With increasing forecast horizon, the standard deviation for the P model increases linearly. For the SP and SVR-GA model, the increase in RMSE and standard deviation is very small for the 45 - 75 minute forecast horizon.

4.5. For the P model, the RMSE error increases from  $40.43$  to  $173.61$   $\text{Wm}^{-2}$  as the forecast horizon increases from 5 to 75 minute. The trend in error is linear. In case of SP, there is steep increase in RMSE from 5 to 20 minutes and afterwards the error magnitude starts converging and the increase in error is only  $10$   $\text{Wm}^{-2}$  as the forecast horizon changes from 25 to 75 minute. Similar patterns are observed in the results for SVR-GA. Furthermore, the SP always underestimates GHI irrespective of the forecast horizon whereas the SVR-GA initially underestimates for 5 - 25 min forecast horizon. Afterwards, it overestimates GHI resulting in a negative bias.

Using P as a reference model, the forecast skill achieved by SP and SVR ranges between 3.41 to 56.21% and 17.86-61.74% respectively. Comparing the rRMSE forecast errors to evaluate the performance of P and SP for 5 to 25 minute forecast horizon, the improvements are marginal. As the forecast horizon starts to increase, the improvements start increasing linearly. There is clear benefit in using smart persistence over persistence model and in using SVR-GA over smart persistence. A skill of 14.49% can be expected using SVR-GA over SP. After a forecast horizon of 40 minutes, the improvements achieved by the forecast model are not consistent as the forecast age increases and ground imagery is not sufficient enough to capture cloud dynamics outside this time horizon. For forecast horizon greater than 40 minutes, additional data such as satellite imagery can be taken into account if higher accuracy is required.

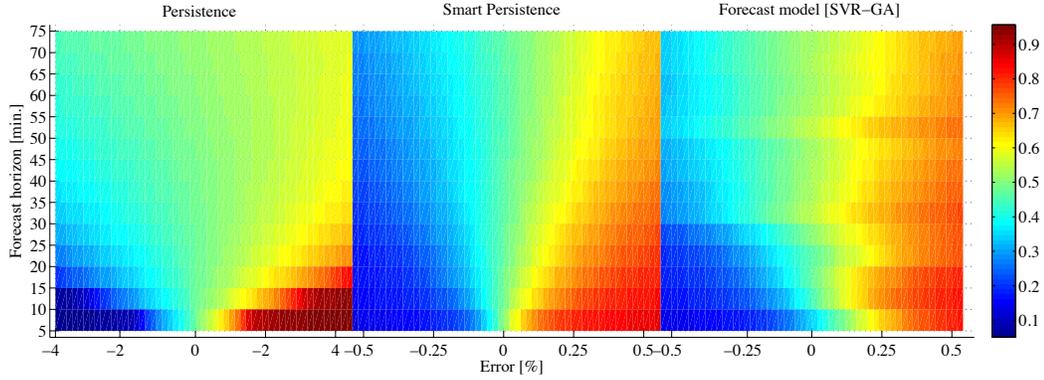


**Figure 4.4:** RMSE and standard deviation for short-term solar forecasts with horizons of 5 to 75 minutes in 5 minute intervals. With increasing forecast horizon the error in Persistence (P) model increases linearly. The error is almost constant after a 25 minute forecast horizon for the Smart Persistence (SP) and SVR-GA forecast model.

**Table 4.5:** Statistical error metrics for short-term solar irradiance forecast in 5-minute intervals.

Forecast horizon (minutes)	Persistence				Smart Persistence					SVR-GA					
	MAE (Wm <sup>-2</sup> )	MBE (Wm <sup>-2</sup> )	RMSE (Wm <sup>-2</sup> )	rRMSE (-)	MAE (Wm <sup>-2</sup> )	MBE (Wm <sup>-2</sup> )	RMSE (Wm <sup>-2</sup> )	rRMSE (-)	Skill-P (%)	MAE (Wm <sup>-2</sup> )	MBE (Wm <sup>-2</sup> )	RMSE (Wm <sup>-2</sup> )	rRMSE (-)	Skill-P (%)	Skill-SP (%)
5	19.81	1.13	40.43	8.54	13.42	0.24	39.05	8.25	3.41	12.62	1.35	33.21	7.02	17.86	14.95
10	32.26	2.06	55.52	11.71	18.87	0.47	51.28	10.81	7.64	17.05	0.89	44.33	9.35	20.15	13.55
15	42.77	2.75	66.25	13.95	22.17	0.68	58.06	12.22	12.36	19.40	0.83	48.45	10.20	26.87	16.55
20	52.58	3.22	75.78	15.94	24.62	0.89	62.74	13.20	17.21	21.44	0.95	52.34	11.01	30.93	16.57
25	61.80	3.47	84.41	17.74	26.33	1.10	65.70	13.81	22.17	22.91	0.11	55.33	11.63	34.45	15.78
30	70.93	3.52	93.32	19.61	27.96	1.33	68.40	14.38	26.70	24.27	-0.37	57.45	12.08	38.44	16.01
35	80.01	3.35	101.71	21.39	29.20	1.56	69.86	14.69	31.31	25.42	-0.50	59.46	12.50	41.14	14.89
40	88.80	2.98	110.05	23.16	30.12	1.81	70.74	14.89	35.72	26.21	-0.72	60.37	12.70	45.14	14.66
45	97.42	2.38	118.75	25.02	30.87	2.04	71.76	15.12	39.57	26.96	-0.87	61.69	13.00	48.05	14.03
50	106.02	1.53	127.38	26.89	31.38	2.23	72.15	15.23	43.36	27.42	-1.47	62.49	13.19	50.94	13.39
55	114.97	0.45	136.73	28.93	32.15	2.39	73.48	15.55	46.26	28.08	-0.96	63.53	13.44	53.54	13.54
60	123.87	-0.86	146.02	30.98	32.81	2.52	74.48	15.80	48.99	28.43	-0.86	64.50	13.68	55.83	13.40
65	132.63	-2.37	155.31	33.06	33.28	2.60	75.30	16.03	51.52	28.99	-0.94	65.07	13.85	58.10	13.59
70	141.24	-4.07	164.50	35.14	33.78	2.65	75.66	16.16	54.00	29.22	-0.46	65.19	13.93	61.37	13.84
75	149.79	-5.96	173.61	37.23	34.16	2.63	76.03	16.31	56.21	30.97	-0.14	66.42	14.24	61.74	12.64

The cumulative frequency distribution (CDF) for the forecast errors is shown in figure 4.5. Using SP and SVR-GA reduce the spread in the forecast model. Comparing all the three models, there is 0.3-0.6 probability that error will be between  $\pm 1\%$  for P,  $\pm 0.1\%$  for SP and 0 – 0.2% for the forecast model. Since positive error means underestimation, there is excess of energy as compared to estimated. This kind of forecast is beneficial for the market participant as they can curtail extra energy. More details about the value of these forecasts for real-time energy imbalance markets is discussed next.



**Figure 4.5:** Cumulative frequency distribution (CDF) of the solar forecast errors in five minute intervals for the Persistence (P), Smart persistence (SP) and forecast model (SVR-GA) for the forecast horizon ranging from 5 minutes to 75 minutes. The color bar represents the CDF ranging from 0 to 1. The spread of errors in the P model is reduced by the SP and SVR-GA. For the 5 minute forecast horizon, there is 0.4 to 0.6 probability that there will be  $\pm 1\%$  error whereas with SP it is only  $\pm 0.1\%$ .

## 4.4 Summary

Day-ahead and short-term solar resource forecasts covering forecast horizons ranging from 1 day to 5 minute in 1 hour, 15 and 5 minute intervals are proposed and validated. State-of-the-art forecasts (with reforecast enhancements) are shown. The reforecast methodology reduces the bias in day-ahead forecast by 67.21% in NAM forecast and achieves a forecast skill of 14.5% over the smart persistence model. The error of the persistence model increases linearly, whereas smart persistence and the SVR-GA model, the error converges by forecast horizon of 25 minutes for 5 minute resolution. The skill achieved by SVR-GA ranges between 14.53-20.11% and 12.64-16.75% for 15 minute and 5 minute forecast intervals (forecast horizons ranging from 5 to 75 minutes). For forecasts, covering greater than or equal to 35 minute forecast horizons, the skill drops implying that more sophisticated methods e.g. including satellite imagery are required in addition to ground data.

## Acknowledgments

This chapter contains work previously submitted to journal:

- A. Kaur, L. Nonnenmacher, H.T.C. Pedro and C.F.M. Coimbra, “Benefits of solar forecasting for energy imbalance market”, *Renewable Energy*, [Submitted, March 2015].

# Chapter 5

## Net load forecasting for power grids with high solar penetration

The power grid is undergoing an inevitable change. With the increasing awareness about the adverse effects of conventional energy resources, policies are being developed to accelerate the penetration of renewable energy. The biggest potential for further capacity installation are intermittent renewable sources, mainly wind and solar. Based on data from the World Wind Energy Council, 320 GW are currently installed with a prospective increase up to 2000 GW by 2030. Due to the stochastic nature of solar and wind power, this increasing renewable penetration presents various types of management and operational challenges for the reliable operation of the electric grid on both production and consumption side [82, 94]. On the demand side, it is challenging for grid operators, (e.g. the Independent System Operators (ISOs) or the utilities) to match variable production from intermittent sources with the net load of the customers or the microgrid. This problem intensifies with customers-generators and microgrids with onsite variable generation since the stochasticity in onsite power generation translates into the net load demand from the macro grid. To mitigate these adverse effects, forecasts for expected power generation and net load are needed. There have been significant advancements in the field of load and renewable energy forecasting. Comprehensive reviews are available on load forecasting [5, 18, 91], solar forecasting [19, 95] and wind forecasting methods [96, 97]. However, the integration of these fore-

casting methods in the operational practices of system operators has gained little attention.

The major contributions of this study are (1) the introduction of the concept of net load forecasting for grids with high renewable energy penetration; (2) the implementation of a solar power prediction algorithm optimized with decision heuristics based on no-exogenous inputs only. This forecast has competitive accuracy in comparison to more complex models and can be used by commercial solar power producers to manage solar power production and plan ahead for the expected ramps in solar energy in data poor environments independently of the integration to a net load forecast; (3) proposition of methods to integrate load and solar forecasts to create the net load forecast. The net load forecast concept can be adapted for wind forecasting as well and (4) validating the beneficial characteristics of net load forecasting with data from a microgrid with high penetration of solar power (up to 33% annually).

The paper is structured as follows: section 5.1 provides the background of load, demand and production forecasts and introduces the concept of net load forecasting. Section 5.2 explains the data sets utilized in this study and why a microgrid is used as a testbed. Section 5.3 contains the proposed methodology. Section 5.4 shows details of the model implementation. Section 5.5 discusses results and uncertainty of the net load forecast. Section 5.6 presents the relationship between the occurring solar and net load forecast errors. Final conclusions are drawn in section 5.7.

## 5.1 Background

This section gives a short review of relevant previous work, categorized into solar power generation and load forecasting for power grids. The concept of net load forecasting merging production and load forecasts is presented and its technical and economical benefits for current and future grids are discussed.

### 5.1.1 Solar power generation forecasting

Various solar irradiance forecasting techniques using Artificial Neural Networks, sky imagery, numerical weather prediction, etc., have been proposed [24, 22, 21, 74, 98, 72, 99, 100]. An extensive review on solar forecasting techniques can be found in [19, 95]. While all solar irradiance forecasting methods can be used as an input to forecast solar power output, there have been studies directly forecasting output of solar power plants. Application of regression methods to forecast solar power using weather forecast as an exogenous input was shown in [101, 102, 103] and [101] concluded that the accuracy of solar power forecasts can be increased by 10% by using more accurate weather forecast. Moreover, [102] showed that the past values of solar power contribute to the accuracy of forecast model with up to 2 hour forecast horizon, thus the use of weather forecasts as an input is recommended for forecast horizons greater than 24 hours. A Kalman filter was designed to forecast solar power for cloudy days in [104]. A methodology to predict solar power forecasting up to 2 days forecast horizon using European Centre for Medium-Range Forecasts (ECMWF) as an input was presented in [105] and results showed that the proposed methodology adapted to changing weather conditions but overestimated solar production for the snow cover on the modules.

Furthermore, Artificial Neural Networks (ANNs) based on self-organizing maps using weather forecasts and past power generation as inputs were applied in [106]. Support Vector Machines (SVM) were used in [107] for solar power forecasting by classifying the days as clear, cloudy, foggy, and rainy day. Fuzzy theory and ANN based method was proposed in [108] and the results were validated through computer simulation. Similarly, weather based hybrid method consisting of self-organizing maps and linear vector quantization networks was proposed in [109] for day ahead hourly solar power prediction. The results showed that hybrid method outperformed simple SVR and traditional ANN methods. All the forecast methods discussed above use exogenous inputs like weather forecasts, sky imagery, etc. Various methods with no exogenous inputs were investigated in [23]. The study concluded that the ANN based method outperformed all other methods i.e. persistent model, Autoregressive Integrated Moving Average model (ARIMA), and

k-Nearest Neighbors (kNNs). Significant improvements can be achieved by optimizing ANN parameters with Genetic Algorithms. All mentioned studies related to solar power forecasting are listed in table 5.1.

None of these studies take into account soiling effects, varying aerosol content in the atmosphere and efficiency degradation of the solar panels over time [110, 111]. In section 5.3, we propose a solar power output forecast model that includes heuristics to account for the changing solar power profiles due to change in seasons, aerosol content in the atmosphere, soiling, etc.

### 5.1.2 Load forecasting

Most previously proposed techniques for load forecast for power grids are based on artificial intelligence [7], ensemble methods [91, 57, 27], Support Vector Regression [112, 113, 114], and hybrid models [115]. Optimization techniques are applied to select the input variables and hyper-parameters for the forecast model. In many recent studies, application of biologically inspired optimization algorithms are shown for load forecasting problem [116, 117]. The focus of all these studies has been to provide more accurate and reliable load forecasts. The problem of forecasting net load for power grids with high renewable energy penetration has not been addressed, yet. In [20] it was shown that onsite solar PV generation impacts the load forecast accuracy when conventional methods are used. An accuracy drop of 3% and 9% for 1-hour and 15-minute forecast horizon respectively has been reported driven by the variability of the solar resource rather than the solar penetration level. Thus, current industrial forecasting practices have to be updated to accommodate increasing renewable energy penetration.

### 5.1.3 Net load forecasting

A recent study investigating the interconnection of distributed generation in Massachusetts by [118] highlights the need for more accurate intra-day and day-ahead resource forecasting capabilities and conclude that load and generation forecasts could be combined to reflect the net load for areas. [119] quantified the

**Table 5.1:** Forecast models proposed for solar power forecasting.

Ref.	Inputs	Forecast models	Forecast horizon	Data forecast	Location
[101]	Weather forecast and temperature forecast	Regression method	1h	Solar power	Expo 205, Aichi Japan
[102]	Weather forecast	Autoregressive and Autoregressive with exogenous input	upto 36h	Solar power from 21 PV stationson rooftops	Denmark
[105]	Forecasts from European Center for Medium-Range Forecasts (ECMWF)	Physical model	upto 2d	Solar power	Oldenburg, Germany
[106]	Past measurements and meteorological forecasts of solar irradiance, relative humidity and temperature	Self-organized map (SOM) and ANN	24 h	Solar power	Huazhong, China
[107]	Temperature	Support Vector Machine	1 day	Solar power	China
[23]	Time-lagged inputs, no-exogenous inputs	Persistent model, Autoregressive Integrated Moving Average model (ARIMA), k-Nearest Neighbor (kNNs), ANN and ANN-GA	1h and 2h	Hourly averaged solar power data from 1 MW solar farm	Merced CA, USA
[108]	Weather reported data i.e. clouds, humidity and temperature	Fuzzy theory and Recurrent Neural Networks	24h	Hourly solar power simulations	-
[109]	Historical PV and weather prediction by Taiwan Central Weather Bureau (TCWB) e.g. temperature, probability of precipitation and solar irradiance	Weather-based hybrid method consisting of Self-Organizing Map (SOM), Learning Vector Quantization (LVQ), Support Vector Regression (SVR) and fuzzy inference method	1-d; every 3-h	Hourly solar power	Taiwan

uncertainty of net load caused by inaccurate wind power output predictions. In this study, we pick up the concept of net load uncertainty from [119, 20] and combine it with the suggestion from [118] to aggregate several forecasts to a net load forecast. While this study is based on data from a microgrid with high solar penetration (section 5.2.1), the concept of net load forecasting is equally valuable for all kind of power grids with high renewable energy penetration from intermittent generators, explicitly also for interconnected grids with distributed generation from wind and solar energy converters.

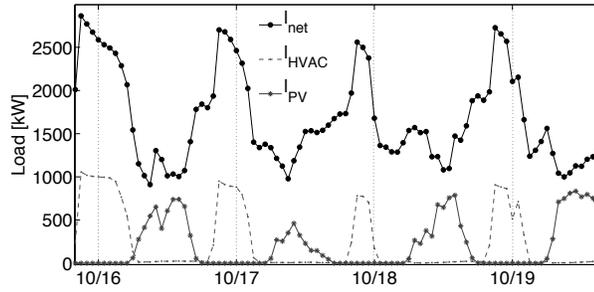
In general, the need and value of net forecasting heavily relies on the interconnection regulations and tariffs under which distributed resources or microgrids are tied to a macrogrid or transmission grid. For example, the California Public Utility Commission (CPUC) currently only recognizes three types of tariffs for microgrid interconnection: (1) net-metering, (2) self-generation (impedes export of generated energy and is usually combined with a time-of-use (TOU) tariff when energy has is purchased from the macrogrid) and the (3) wholesale distribution access tariff (WDAT) [120]. All of these interconnection options impede to take full advantage of beneficial technical capabilities a microgrid can provide in the energy system since they do not facilitate a bidirectional flow of energy. Many studies highlight the need for better interconnection regulations and tariffs that enable a better integration of microgrids within the macrogrid, while sharing costs and benefits fairly [121, 122, 120, 123, 124]. Therefore, under current conditions, net load forecasting solely creates economic value by reducing energy purchasing costs for the microgrid operator (optimized load shifting, see section 5.2 for details). Under future scenarios, with regulations in place that enable microgrids to draw energy from and provide services to the macrogrid, accurate net load prediction becomes important since the load uncertainty at the point of common coupling is an important variable for all interconnection regulations. Market participation and the necessity to minimize the uncertainty introduced by large fleets of interconnected microgrids also requires accurate net load forecasting.

## 5.2 Microgrids as testbeds

Various operational decisions of any power grid rely on forecasting. While the findings of this study are generally valid for all power grids with high penetration from intermittent energy sources, the validation of the proposed methods relies on data and findings from microgrids since they provide an excellent testbed for future utility-scale power grids with high renewable generation (e.g. the UC Merced microgrid, see section 5.2.1). Experience from existing microgrids and proof-of-concept studies show the need for accurate forecasting of several variables such as demand, available demand response capacity and power-generation for optimized operations [125, 118, 126]. For example, in the case study of Borrego Springs, a microgrid installed and operated as described in [125], customer load could be curtailed when the forecasting algorithm found benefits for curtailment. Interconnected load forecast is a parameter driving the optimization of microgrid controls and the energy management system [126]. They mention a campus microgrid system with forecasting based optimized resource dispatching, self generation, and grid purchases at Princeton University, New Jersey. Using the load and price forecasts, the mentioned microgrid can buy electricity from the macrogrid based on the hourly wholesale electricity market prices. Optimized purchasing during low energy price times resulted in \$2.5 to \$3.5 million annual savings. As another example, they mention that the Burrstone Energy Center ( $3.6MW_p$  generation) operates under similar conditions to maximize the economic value of their microgrid. The details about the microgrid used in this study are provided in the next subsection.

### 5.2.1 Testbed data

The proposed methodology is applied to forecast solar power and net load demand of University of California, Merced (UCM) situated in San Joaquin valley (see figure 5.1). This community is an ideal test bench to study prospective micro-grids with high solar penetration because it meets 33% of its annual and 3-55% of it's daily power demand by solar energy produced by an onsite 1 MW



**Figure 5.1:** Block diagram for net load for UC Merced system,  $l_{net}$  is the net load demand of the campus from the grid,  $l_{HVAC}$  is the heating, ventilation and air-conditioning load; and  $l_{PV}$  is the solar power output. The intermittence observed in solar power is translated into load demand. Furthermore, at the end of the day the sudden increase in the load demand, also known as duck curve is a major concern for the utilities.

single axis tracking solar power plant [20]. The Heating Ventilation and Air Conditioning (HVAC) load for the campus is a time-independent load. Under current market conditions as discussed above, the advantageous characteristics of net load forecasting as proposed in this study root in the opportunity under tariff option (2) to shift load (e.g. for Heating Ventilation and Air Conditioning (HVAC)) to off-peak hours due to TOU pricing which are lowest at night-time. Hourly data sets consisting of solar energy production, HVAC load and load demand from the grid are utilized for this study.

The data was time synchronized and pre-processed to remove outliers. Data for the year 2010 consisting of 5546 data-points are considered as a training set and data for the year 2011 with 7548 data-points are considered as a testing set. Using data sets for the whole year as training and test sets encompass all seasonal variations for the given location.

### 5.3 Proposed methodology

The net load from the grid  $l_{net}$  for any given time  $t$  can be expressed as,

$$l_{net}(t) = l_T(t) - l_{PV}(t), \quad (5.1)$$

where  $l_T(t)$  is the total UCM load,  $l_{PV}$  is the load demand met by an onsite solar generation i.e.,  $l_{PV}$  is equal to the onsite solar power generation  $p_{PV}$ . Since, thermal storage plant for HVAC load is operated manually, it is assumed to be a deterministic load for this study. Thus, total load demand from the grid  $l_T(t)$  can be decomposed into deterministic and stochastic part  $l_s(t)$ ,

$$l_T(t) = l_s(t) + l_{HVAC}(t). \quad (5.2)$$

Comparing equation 5.1 and 5.2, forecasting net load simplifies to forecasting stochastic part which is equivalent to,

$$\hat{l}_{net}(t) = \hat{l}_s(t) + l_{HVAC}(t) - \hat{l}_{PV}(t), \quad (5.3)$$

where  $\hat{\cdot}$  represents the forecast of  $\cdot$ . The algorithms to forecast solar power and net load are discussed in the next subsection.

### 5.3.1 Solar power forecast

Solar power  $p_{PV}(t)$  at any time  $t$  can be considered as a sum of deterministic clear sky solar power  $p_{CS}(t)$  and stochastic component i.e.,

$$p_s(t) = p_{PV}(t) - p_{CS}(t). \quad (5.4)$$

The clear sky solar power  $p_{CS}(t)$  is a function of day of the year, latitude, longitude of the location which are all deterministic factors. But it is also affected by various daily and seasonal processes due to changing linke-turbidity factor and temporal degradation of solar panel efficiency [110, 111]. To account for these factors, adaptive clear sky solar power identification to update daily clear solar power is presented. To correct for overcast conditions, morning and evening time values heuristics are proposed.

#### Clear sky solar power identification

Clear sky solar irradiance is deterministic and can be modeled as a function of hour of the day  $\tau_D$ , day of the year  $\tau_Y$ , latitude and longitude of the location.

But in case of clear sky solar power  $p_{CS}$  along with the deterministic part,  $p_{CS}$  is continuously affected by the seasonal change, aerosol content in the atmosphere, dust accumulating on the solar panel, temperature dependent efficiency of solar panel, solar panel degradation over the time and so on [110]. This continuous change adds to the forecast errors. These factors can be accounted empirically based on most recent available information about the system. Thus, adaptive clear model  $p_{CS}^a$  that takes into account the recent changes in clear sky solar power is proposed here. This step ensures an accurate separation of deterministic and random components of the solar power after detrending. Solar power based clearness index,  $k_{t,PV} = \frac{p_{PV}}{p_{CS}^a}$  is defined and the value of  $k_{t,PV}$  range between 0 and 1. At the end of the day, five clear sky criteria introduced by [127] are applied to check if the clear sky model should be updated or not. The five criteria are briefly defined below.

1. Mean solar power value during the time period,

$$\bar{G} = \frac{1}{N} \sum_{t=1}^N p_{PV}(t). \quad (5.5)$$

2. Maximum irradiance value in the time-series,

$$M = \max\{p_{PV}(t)\} \quad \forall t \in \{1, 2, \dots, N\}. \quad (5.6)$$

3. Length of the line formed by  $p_{PV}$  values in the time-series,

$$L = \sum_{t=1}^N \sqrt{(p_{PV}(t + \Delta t) - p_{PV}(t))^2 + (\Delta t)^2}. \quad (5.7)$$

4. Maximum deviation from the clear sky slope,

$$S = \max\{|s(t) - s_c(t)|\} \quad \forall t \in \{1, 2, \dots, N\}, \quad (5.8)$$

where,

$$s_c(t) = p_{CS}(t + \Delta t) - p_{CS}(t). \quad (5.9)$$

**Table 5.2:** Thresholds for clear sky solar power identification.

$G_t$	$M_t$	$L_t$	$S_t$	$\sigma_t$	$k_t$	N	$\Delta t$
150	220	220	120	0.12	0.90	3	1 h

5. Variance in the time-series,

$$\sigma = \frac{\sqrt{\frac{1}{N-1} \sum_{t=1}^{N-1} (s(t) - \bar{s})^2}}{\frac{1}{N} \sum_{t=1}^N p_{PV}(t)}, \quad (5.10)$$

where,

$$s(t) = p_{PV}(t + \Delta t) - p_{PV}(t) \quad \forall t \in \{1, 2, \dots, N\}, \quad (5.11)$$

and

$$\bar{s} = \frac{1}{N-1} \sum_{t=1}^{N-1} s(t). \quad (5.12)$$

To make the identification criteria more robust, after checking for threshold, the measured values of the clearness index are also considered. If the clearness index values for the day time are greater than 0.90 then the day is considered as clear day and the clear sky solar power model is updated. The steps for clear solar power identification are given in the algorithm 1.

## Heuristics

Solar power forecast is produced using a base model. In this study we consider Support Vector Regression model as a base model. The forecast model produces de-trended solar output and the adaptive clear sky solar power is added at the end i.e.,

$$\hat{p}_{PV}(t) = \hat{p}_S(t) + p_{CS}^a(t). \quad (5.13)$$

Adding clear solar power always tends to overestimate solar irradiance for overcast conditions (see figure 5.3 for early morning period on 03/02/2011 and 03/06/2011). To correct for this issue, heuristics based on persistence assumption are applied for solar elevation angle  $\alpha_s > -2$ . It assumes that for overcast conditions i.e.  $kt_{pv}(t-1) < 0.30$ , the forecast will be a sum of past values and

---

**Algorithm 1:** Clear sky solar power identification
 

---

**Inputs :** Hourly values of solar power  $p_{PV}$

**Output:** Identified clear sky solar power  $p_{CS}^a$

```

1 initialize  $p_{CS}^a$ ;
2 for all unique days do
3   compute  $G, M, L, S, \sigma$  and  $k_t$  at the end of the day;
4   if  $G < G_t \ \& \ M < M_t \ \& \ L < L_t \ \& \ S < S_t \ \& \ \sigma < \sigma_t$  then
5     | update clear sky model,  $p_{CS}^a$ ;
6   else
7     | if  $k_{t,PV} > 0.90$  then
8       | update clear sky model,  $p_{CS}^a$ ;
9     | end
10  end
11 end

```

---

current weather conditions times the base model forecast value. For details refer to algorithm 2.

The base model used in this paper depends on the past lagged values. Since, at the beginning of the day (sunrise) the inputs are past night time values which are equal to zero, there is a discontinuity in data as night values do not give any useful information about the first hour of the sun rise. Most of the solar irradiance forecast studies ignore the solar irradiance/power values for solar zenith angle less than 5 or 15 degrees [74] because the values of solar irradiance/solar power are negligible as compared to rest of the day time values. But in the case of net load forecasting, a continuous forecast of solar power for all the hours is needed. Therefore, we assume the first morning value to be the clear sky solar value. Similarly, the last value before the sunset is so small that the affect of atmospheric condition is negligible and hence, it is assumed to be equal to the clear sky value.

---

**Algorithm 2:** Heuristics
 

---

**Inputs :**  $\hat{p}_{PV}, k_{t,PV}, \alpha_s$ 
**Output:** Updated solar forecast,  $\hat{p}_{PV}^h$ 

```

1 initialization;
2 while  $\alpha_s > -2$  do
3   if  $kt_{PV}(t-1) < 0.30$  then
4     |  $\hat{p}_{PV}^h(t, \alpha_s) = \hat{p}_{PV}k_{t,PV}(t-1) + p_{pv}(t-1)$ ;
5   else
6     |  $\hat{p}_{PV}^h(t, \alpha_s) = \hat{p}_{PV}$ ;
7   end
8 end

```

---

### 5.3.2 Net load forecast

We compare two approaches to perform net load forecasting : additive and integrated model. In case of additive model, the net load forecast is performed as,

$$\hat{l}_{net}(t) = \hat{l}_T(t) - \hat{l}_{PV}(t). \quad (5.14)$$

Whereas in the integrated model, solar power forecast is used as an input to net load forecasting model. Deterministic  $l_{HVAC}$  is added to the net load forecast at the end. Both the methods are tested by implemented time-series and machine learning based forecast models i.e. Autoregressive model and Support Vector Regression model.

#### Autoregressive model (AR)

AR model is a linear time-series regression model. Using this model, the output can be expressed as linear combination of past outputs/measured values i.e.,

$$A(q)y(t) = e(t) \quad (5.15)$$

where  $q^N$  is shift operator i.e.,  $q^{\pm N} = l_{net}(t \pm N)$ ,  $A(q)$  represents the combination of past values of output using coefficients,  $a_1, a_2, \dots, a_{na}$  and  $A(q) =$

$1 + a_1q^{-1} + \dots + a_{na}q^{-na}$ . The part of the time-series not modeled is represented as  $e(t)$ . For more details refer to [36].

### Autoregressive model with exogenous input (ARX)

ARX model is an extension of AR model with an addition of external inputs. In this study the external input,  $u(t)$  consists of solar power forecast values and past measured solar power. Mathematically, it can be represented as,

$$A(q)y(t) = B(q)u(t - n_k)(t) + e(t) \quad (5.16)$$

where  $B(q) = b_1 + \dots + b_{nb}q^{-nb+1}$ ,  $n_k$  is time delay parameter, it is equal to zero for solar power forecast and 1 for the past values of the measured solar power.

### Support Vector Regression (SVR)

Support Vector Regression technique is based on supervised machine learning algorithm [87, 88]. It has been widely applied for forecasting time-series [128, 129, 130, 114]. Given the training data  $\{(u_1, y_1), \dots, (u_l, y_l)\} \subset U \times \mathbb{R}$  where  $U$  denotes the space of input pattern. The goal is to find a function  $f(u)$  that has at-most  $\epsilon$  deviation from the actually obtained targets  $y_i$ ,

$$f(u) = \langle w, u \rangle + b \quad \text{with} \quad w \in U, b \in \mathbb{R}. \quad (5.17)$$

Support Vector regression solves the following optimization problem,

$$\begin{aligned} \min_x \quad & \frac{1}{2}w^T w + C \sum_{0 < i < m} (\xi_i + \xi_i^*), \\ \text{subject to} \quad & y_i - (w^T \phi(u_i) + b) \leq \epsilon + \xi_i^*, \\ & (w^T \phi(u_i) + b) - y_{ti} \leq \epsilon + \xi_i, \\ & \xi_i, \xi_i^* \geq 0, i = 1, \dots, l, \end{aligned} \quad (5.18)$$

where  $\phi(u)$  maps  $u_i$  to a higher dimensional space using a kernel function and the training errors are subject to  $\epsilon$ -insensitive tube  $y_i - (w^T \phi(u_i) + b) \leq \epsilon$ . Cost of the error,  $C$ , width of the tube and the mapping function  $\phi$  controls the regression quality.

**Table 5.3:** Output and input variables for the SVR model.

Forecast model	$y(t)$	$u(t)$
Solar power	$p_s(t)$	$p_s(t-1), p_s(t-2), \dots, p_s(t-n_{p_s})$
Additive model	$l_t(t)$	$l_t(t-1), l_t(t-2), \dots, l_t(t-n_{l_t})$
Integrated	$l_{net}(t)$	$l_{net}(t-1), l_t(t-2), \dots, l_t(t-n_{l_{net}}), \hat{p}_s(t), \dots, \hat{p}_s(t-n'_{p_s}), p_s(t-1), p_s(t-2), \dots, p_s(t-n''_{p_s})$

For this study, time series formulation is applied i.e.,

$$y_t = f(u(t)). \quad (5.19)$$

The details about  $y(t)$  and  $u(t)$  are given in table 5.3, where  $n_*$  represents the number of lagged values of the variable  $*$ . Because of the use of external input into the SVR model for integrated net load forecasting model, we term it as SVRX to avoid ambiguity.

## 5.4 Model implementation

For the clear sky identification model, the thresholds for  $G_{th}$ ,  $L_{th}$ ,  $M_{th}$ ,  $S_{th}$  and  $\sigma_{th}$  were obtained using the training set given Table 5.2. These thresholds can be updated for other sampling frequencies for both solar power and solar irradiance.

For the forecast model, LIBSVM : A Library for Support Vector Machines [93] was used. Parameter were selected using grid search by  $\nu$ - fold cross validation technique. The SVR optimization problem was simplified to finding  $C$  and  $\gamma$  values as discussed in [114]. The solar power data was scaled linearly in the range of [0.1 1]. For the SVR model, a radial basis kernel function was used. To select the number of lagged inputs Rissanen's Minimum Description Length (MDL) criterion was applied. The model with derived parameters was trained using the whole training set and was validated using the testset.

For integrated net load forecasting, the solar power forecast produced using the above algorithm was used as one of the input parameter for the training model. Also, note that there were some days in the training and test set when net load was negative due to excessive solar generation. To account for negative values, the net load was scaled linearly between [-1 1].

## 5.5 Results and discussion

Various statistical metrics are applied to quantify the results. The forecast results are reported in terms of Mean Absolute Percentage Error (MAPE), Mean Bias Error (MBE), Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and coefficient of determination ( $R^2$ ). MAPE is very sensitive to high magnitude errors when actual value is very small. In this work, the forecast and actual value are removed in computing MAPE when actual value for net load or solar power are less than 0.05 kW.

Furthermore, Persistence model, is implemented based on the assumption of that current conditions are likely to persist in future,

$$\hat{g}(t) = y(t - 1). \quad (5.20)$$

An extension of persistence model is Smart Persistence (SP) model that takes into account for deterministic information available about the system. For this model, the forecast is sum of present stochastic component of solar power that is assumed to persist in future and deterministic future clear sky solar power value i.e.,

$$\hat{p}_{pv}(t) = p_s(t - 1) + p_{CS}(t). \quad (5.21)$$

SP is used as a reference model to validate the goodness of solar forecast models. The performance of the proposed model is compared to that of a SP model in terms of forecast skill ( $s$ ) [69], which is defined as,

$$s = 1 - \frac{RMSE_{model}}{RMSE_{SP}}. \quad (5.22)$$

### 5.5.1 Clear sky identification

The clear sky identification algorithm was applied to identify the clear days and update the clear sky solar irradiance model,  $p_{CS}^a$ . The results for the identification algorithm are given on table 5.4). For the training set, there were total of 228 days, out of which 109 were clear days. The algorithm identified a total of 115 clear days, out of which 101 days were truly clear days, whereas 14 days

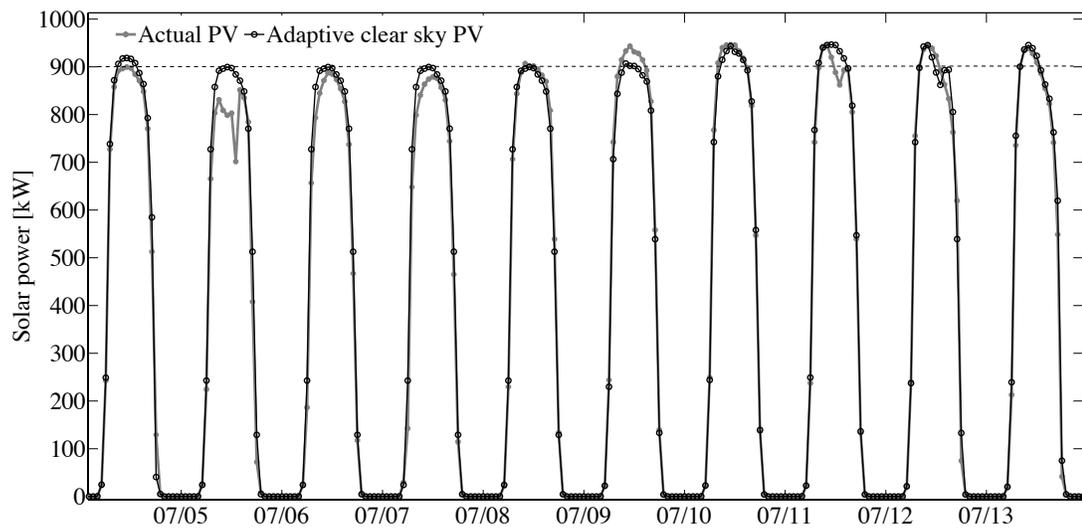
**Table 5.4:** Clear sky identification results

Data set	Total days	Actual clear days	Clear days identified			
			Total	True	False	Missed
Tset	228	109	115	101	14	8
Vset	300	119	126	116	10	3

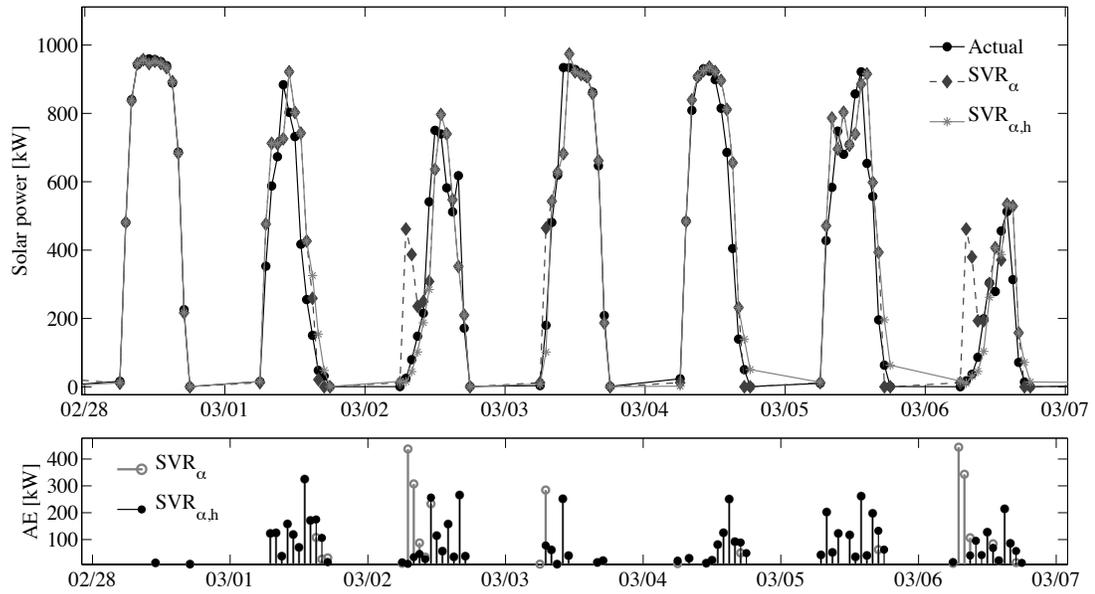
**Table 5.5:** Statistical error metrics for hour-ahead solar power forecast (01-01-2011 to 12-31-2011).

Model	MAPE (%)	MBE (kW)	MAE (kW)	RMSE (kW)	R <sup>2</sup> -	Skill s(%)
Using clear sky model based on $\tau_D$ and $\tau_Y$						
SP	349.08	-9.15	74.44	113.78	0.90	0
SVR	291.18	-12.92	72.08	109.04	0.91	4.17
Using adaptive clear sky model						
SP <sub><math>\alpha</math></sub>	144.78	-9.30	52.29	103.08	0.92	9.40
SVR <sub><math>\alpha</math></sub>	113.47	-16.17	52.19	100.80	0.92	11.41
Applying heuristics						
SP <sub><math>\alpha,h</math></sub>	101.36	-2.37	44.76	88.06	0.94	22.61
SVR <sub><math>\alpha,h</math></sub>	101.12	-5.82	44.33	86.24	0.94	24.20

were falsely identified as clear and 8 days were missed. Given that the accuracy of the algorithm is defined as percentage of actual clear in total number of clear days identified, for the training set the accuracy is 87.82% and for the testing set it is 84.12%. Incorrect identification happens for the days with very small ramps that do not exceed the threshold range as shown in figure 5.2. The disadvantage of identifying incorrect days is that unnecessary ramps in the solar power are introduced which affect the accuracy of the forecast. This does not happen very often as the model auto-corrects itself (e.g. in figure 5.2 it can be observed that 07-11-2011 was identified incorrectly as clear day and a false ramp was introduced for 07-12-2011, but this was autocorrected by 07-13-2011). The errors introduced by identifying incorrect days can be ignored because they are small in magnitude as compared to the improvements achieved in forecasting as discussed in the next section.



**Figure 5.2:** Time series for the solar power generated for the 10 consecutive days from the year 2011. Dashed line indicates a reference level at 900 kW and it can be observed that after 07/08 the maximum solar power exceeds the reference level. Adaptive clear sky model takes into account these kind of changes and updates the clear sky model. Even though 07/11 is a cloudy day, it was identified as a clear day. However, the adaptive clear sky identification algorithm autocorrects itself and it was updated by another clear sky model for 07/13.



**Figure 5.3:** Time series for the actual solar power and forecast for 1-h forecast horizon (*top*) and absolute error, AE (*bottom*) with night values removed. Here we can compare the results from  $SVR_{\alpha}$  and  $SVR_{\alpha,h}$  model. For the overcast period,  $SVR_{\alpha,h}$  is able to correct for the over-predicted solar power by  $SVR_{\alpha}$ . For a cloudy day with ramps (03/01/2011 and 03/05/2011), both the models have similar errors. The  $SVR_{\alpha,h}$  model works better than  $SVR_{\alpha}$  model in detecting overcast conditions and correcting for errors in the morning time (03/02/2011 and 03/06/2011).

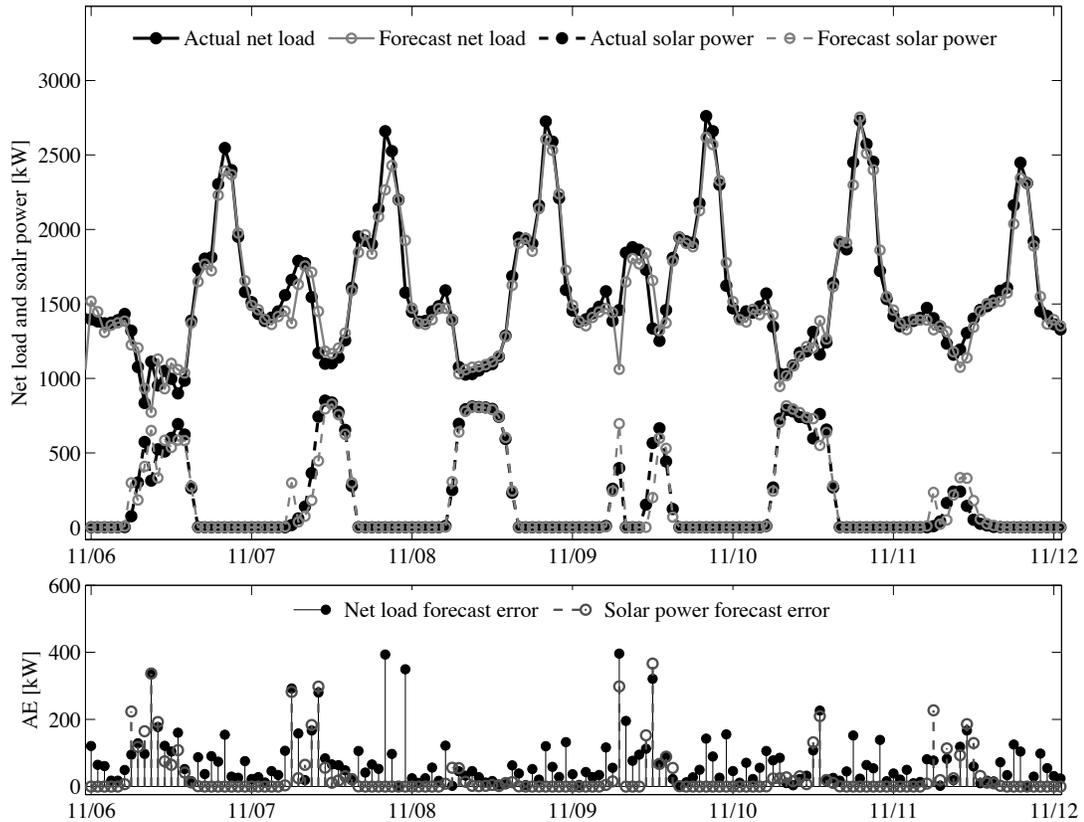
### 5.5.2 Solar power forecast

An hour ahead solar forecast was implemented using SP and SVR as base model. Firstly, the forecast models based on clear sky solar power using  $\tau_D$  and  $\tau_Y$  were implemented. In the next step, the basic clear sky model was replaced with the adaptive clear sky model as explained in section 5.3.1. The improved models were termed as  $SP_\alpha$  and  $SVR_\alpha$ . Finally, heuristics as proposed in section 5.3.1 were applied to the models and termed as  $SP_{\alpha,h}$  and  $SVR_{\alpha,h}$ . All the results and corresponding improvements are listed in table 5.5. MAPE significantly reduces for models with adaptive clear sky model and heuristics. MBE gives the information about the bias in the error. For all the results reported in this study MBE error is negative which suggests that the forecast models always overestimate the power forecast. Furthermore, the MAE gives information about the net error in forecast which is about 44.33 kW i.e., 4.4% of the maximum rating capacity of the power plant. The deviation in forecast values as compared to the actual values is given by the RMSE. It is a scale dependent measure and gives the information in terms of standard deviation w.r.t. the mean.

An adaptive clear sky model ensures that daily variability is taken into account. Thus, by its application the RMSE reduces from 113.78 kW to 103.08 kW for SP and 109.04 kW to 100.80 kW for SVR which is an improvement of 9.40% for SP model and 7.56% for SVR model. Since, the night values give no information about the overcast in the morning, major error was observed in the morning. To correct for such error heuristics were applied and an improvement of 14.57% for SP and 14.44% for SVR model was observed. The statistical metrics show the improvements achieved by using the adaptive clear sky model and then further possible improvements by applying heuristics. Since all these results are achieved without using any exogenous input, the proposed technique can serve as reference to compare the forecast models with exogenous inputs.

### 5.5.3 Net load forecast

The net load forecast was implemented using: 1) additive model where solar and load forecast were produced individually and then combined at the end



**Figure 5.4:** Time series for the net load and solar power forecast for six consecutive days (11/06/2011 to 11/12/2011) from the testset. The absolute error (AE) in the forecasts is shown in the figure below. It can be noticed that the solar forecast error directly influences the net load forecast. Solar power is always over-predicted for the days with overcast conditions (11/07/2011 and 11/11/2011). Heuristics are introduced to correct for these errors. Magnitude of net load forecast error is less for clear (11/08/2011) and overcast days (11/11/2011) as compared to the cloudy days.

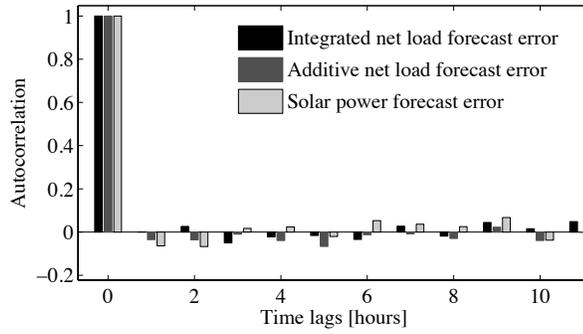
**Table 5.6:** Statistical error metrics for UCM load demand forecast (01-01-2011 to 12-31-2011).

Model	MAPE (%)	MBE (kW)	MAE (kW)	RMSE (kW)	R <sup>2</sup> -
PV forecast including night time values					
SVR <sub><math>\alpha, h</math></sub>	141.07	-2.20	25.59	65.06	0.97
Persistence model for net load forecast					
Persistence	10.93	0.07	152.77	240.92	0.83
Additive model : Model - SVR <sub><math>\alpha, h</math></sub>					
AR	13.98	4.58	64.99	93.83	0.97
SVR	30.47	3.76	63.88	92.48	0.97
Integrated solar power (SVR <sub><math>\alpha, h</math></sub> ) and net load forecast					
ARX	4.60	4.60	57.75	85.06	0.98
SVRX	5.47	5.47	54.74	82.59	0.98

and 2) integrated model where the solar power forecast was used as input into the load forecast model. The forecast error statistics for the additive model and the proposed integrated net load forecasting model are listed in table 5.5.2. The stationarity of errors is shown in figure 5.5 as there is no correlation over the hourly time lags. This validates the model identification because all the information in the time-series has been captured.

The results show that the integrated model outperforms the additive models marginally in terms of all error metrics (see figure 5.6). In case of Autoregressive model, integrated ARX model performs 9.35% better than the additive AR model and SVRX performs 10.69% better than the SVR model in terms of RMSE. Figure 5.6 shows that the spread of additive model forecast errors is more than the integrated model. This validates the lower RMSE for integrated model as compared to the additive model. Thus, integrated model should be preferred for the grid applications.

If compared in terms of MAPE and MBE, time-series based AR model always perform better than the SVR model. Whereas in case of MAE and RMSE, SVR model outperforms the AR model for both additive and integrated case. Sam-

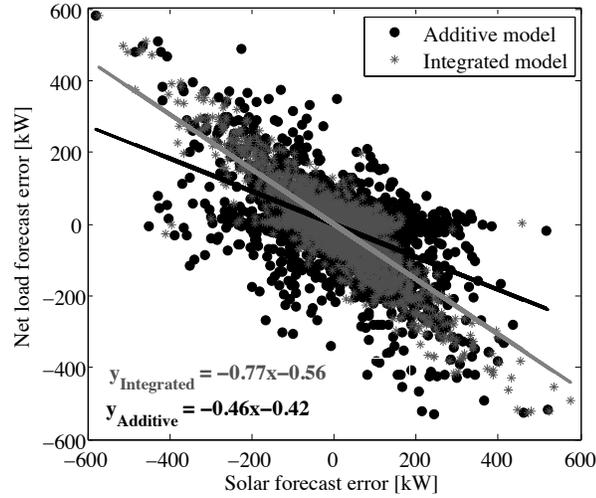


**Figure 5.5:** Error correlation for net load forecast models. After zero lag, there is no correlation in forecast errors. This establishes that all the information in time-series have been captured by the forecast model and the forecast residues are randomly distributed.

ple results are shown in figure 5.4. For overcast days (11-07-2011 (early morning) and 11-11-2011), it can be observed that the solar power model tends to over-predict the initial value. This is due to addition of clear sky solar power and discontinuity in solar data at early morning hours. The forecast error uncertainty is quantified in the next section.

#### 5.5.4 Assessment of forecast uncertainty

The forecast uncertainty can be quantified using the 95% confidence interval. Using the inverse Cumulative Distribution Frequency, the 95% confidence interval corresponds to 2.5 to 97.5 percentile of the distribution reflecting the uncertainty. Freedman-Diaconis rule is applied to define the number of bins for the data sets and the results are shown in figure 5.7. Based on the previous discussions and comparison, it is expected that the uncertainty range for the additive model will be larger than the integrated model. The results show that the uncertainty range for additive model is between -218 kW to 241 kW and for the integrated model the range is from -214.5 kW to 198.6 kW. The maximum value of the net load during the daytime is 2.07 MW. Taking the absolute values, the uncertainty increases by 46 kW for the additive model as compared to the integrated model which is equivalent to 2.2% of maximum net load. Furthermore, for the solar fore-

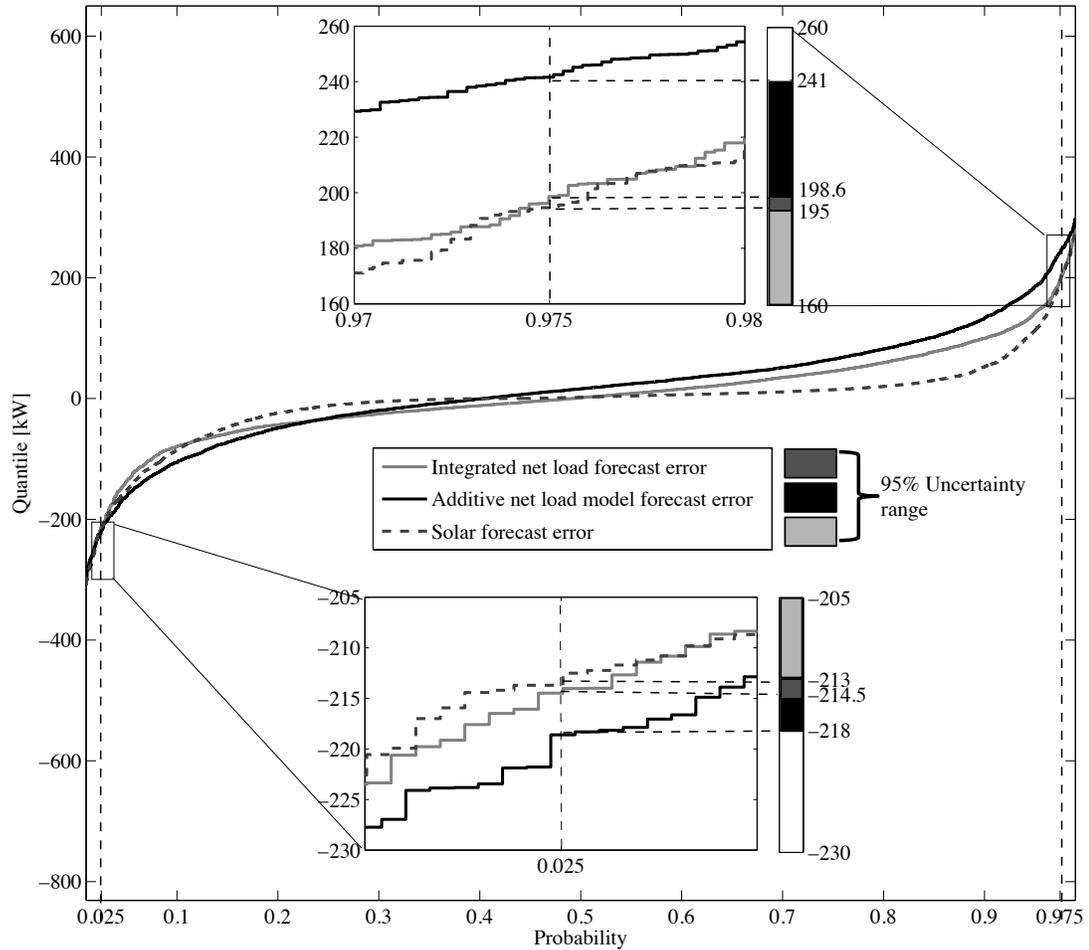


**Figure 5.6:** Comparison of net load forecast errors for both additive and integrated models w.r.t. the solar forecast errors from  $SVR_{\alpha,h}$  model during daytime. Night time values have been removed for this plot and analysis. The net load forecast errors are inversely proportional to solar forecast errors. The linear fit predicts 98% variance of the integrated forecast errors and 83.59% variance of the additive model forecast errors.

casting errors, the uncertainty range is between -213 kW to 195 kW. This is very close to the range of integrated model and it could be a good approximation for the net load forecast errors. The relationship between the solar and net load forecast errors is established in the next section.

## 5.6 Solar and net load forecast errors

For future planning and modeling for the grids with expected high solar penetration, it is important to quantify the relationship between the solar and net load forecast errors. However, as shown in figure 5.5 and discussed previously, both solar and net load errors are stationary and yet figure 5.6 shows that the solar forecast and net load forecast errors are inversely proportional to each other. To test for hidden correlation between these time-series, cointegration test is applied and discussed below.



**Figure 5.7:** Net load and solar forecast uncertainty assessment by applying inverse CDF during daytime. The quantile represents the forecast errors in kW. The upper and lower inset plots show upper and lower bound for the 95% confidence interval for the forecast errors. The 95% confidence interval for the additive net load forecast model ranges between -218 kW to 241 kW (black), for the integrated model this range is -214.5 kW to 198.6 kW (dark grey) and for the solar forecast errors the range is -213 kW to 195 kW (light grey). Thus, the uncertainty range decreased by 46 kW by using the integrated model.

### 5.6.1 Cointegration of solar and net load forecast errors

The concept of spurious regression and cointegration was first introduced by [131, 132]. It is used to define statistical properties of the time-series. Time-series are cointegrated if they share a common stochastic drift. Two random variables are cointegrated if one random variable,  $x(t)$  can be expressed linearly in terms of second random variable,  $w(t)$  using some coefficient  $\beta$  i.e.,

$$x(t) - \beta w(t) = e(t) \quad (5.23)$$

such that the residue,  $e(t)$  after fitting is stationary [133, 134, 135].  $e(t)$  is also known as cointegrating relation. Here, we apply this concept on net load and solar forecast errors. This step ensures that the correlation between two random variable is not spurious and furthermore, error-correction models can be applied for modeling such pair of time-series.

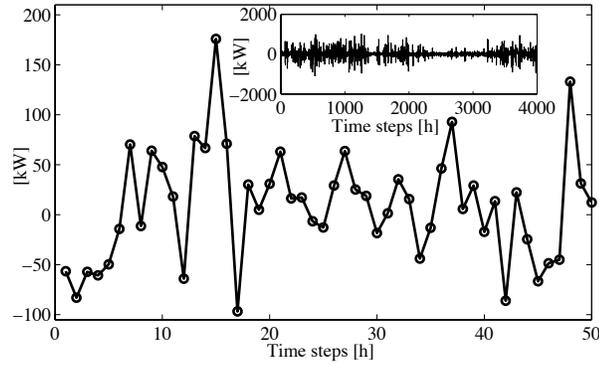
The Engle-Granger cointegration test was applied with the null hypothesis that the net load and solar forecast error time-series are not cointegrated. Results (see figure 5.8) were against the hypothesis with stationary cointegrating relation. Hence, net load and solar forecast errors are co-integrated time-series. This validates the correlation presented in figure 5.6. Therefore, the relationship between net load and solar forecast errors can be represented as

$$\hat{e}(t)_{integrated} = -0.77e(t)_{solar} - 0.56, \quad (5.24)$$

and

$$\hat{e}(t)_{additive} = -0.46e(t)_{solar} - 0.46. \quad (5.25)$$

Using the equations presented above, 98% of the variance of integrated net load forecast errors can be predicted using solar forecast errors whereas only 83.59% of the variance of additive net load forecast errors can be predicted using solar forecast errors.



**Figure 5.8:** Cointegrating relation between the hourly integrated net load and solar forecast errors during the daytime for the first consecutive 50 hours. Inset plot shows the cointegrating relation for the whole time-series. The combination is indeed stationary, which validates the cointegration of the time-series and negates the possibility of spurious regression.

## 5.7 Summary

This concept of net load forecasting is introduced and its technical and economical benefits for interconnected grids are discussed. An exemplary implementation of the proposed concept is shown for a microgrid with high solar penetration. Two different net load forecasting approaches using load and a solar power output forecast are implemented and evaluated: integrated and additive.

To predict solar power a heuristics based approach with no exogenous input is presented. The proposed approach takes changing atmospheric clearness and efficiency degradation of PV panels into account using adaptive clear sky model and heuristics. The accuracy of clear sky solar identification was found to be 84.12%. Adaptive clear sky solar power showed an improvement of 9.4% and the heuristics proposed in this study further showed an improvement of 22.61% for the smart persistence model. For the SVR model, the improvement was 11.41% for the adaptive clear sky model and 24.20% after the heuristics were applied. Thus, the adaptive clear sky model and the heuristics proposed in this study can be applied to any solar forecast algorithm to enhance the forecast accuracy.

As stated above, the solar power forecast is applied for two different net load

forecasting approaches. The integrated solar and load forecast model outperformed the additive model by 10.69 % in terms of Root Mean Square Error for the SVR and SVRX models. The implemented forecast model tends to over-predict solar power for overcast period on early morning time and hence, under-predict net load for the corresponding time. Over the day, frequent forecast errors were observed during cloudy periods as compared to overcast periods which is in agreement with [20].

The uncertainty ranges for the net load and solar power forecast errors were analyzed. The 95% confidence interval for the additive model is larger than the the integrated forecast model by 2.2% of the maximum net load demand. The 95% confidence interval of solar forecast can be used as an approximation for the expected accuracy of the net load forecasts. There is high correlation between the net load forecast errors and solar forecast errors. To validate the correlation between the solar and net load error time-series, the Engle-Granger cointegration test was applied. The two stationary time-series are indeed cointegrated and hence, share the common stochastic drift. Using solar forecast errors, 98% variance of net load forecast errors can be predicted. Thus, solar power time-series is sufficient to provide necessary information to characterize the expected variance and uncertainty in the net load time-series.

Therefore, for utility scale grid and microgrid applications, it is suggested to use an integrated net load forecast model to reduce the uncertainty at the point of coupling between the interconnected grids. The net load forecasting model can also be adapted for grids with high wind energy penetration using a wind forecast method as an input. The proposed concept will enable the grid operators to efficiently manage grids with high intermittent renewable energy penetration and participate in electricity market for economical benefits.

## Acknowledgments

This chapter contains work previously submitted to journal:

- A. Kaur, L. Nonnenmacher and C.F.M. Coimbra, “Net load forecasting for

high renewable energy penetration grids”, *Energy*, [Submitted, December 2014].

# Chapter 6

## Benefits of forecasting

The electricity system is undergoing an inevitable change to address increased variability in generation and net load, introduced by intermittent generators, mainly wind and solar. Many approaches to mitigate the adverse effects of ramping have been proposed, e.g. increased storage capabilities, resource and net load forecasting, demand response, etc. The core of all solutions for integrating higher levels of variable wind and solar generation is to increase the flexibility options available in the grid [136, 137]. Recently, regulating authorities in several jurisdictions reorganized the market environments to allow flexible energy trading schedules, designed to better exploit spatial and temporal diversity in generation and demand. Historically, this reorganization started in Northern Europe by allowing short-term, cross-border electricity trading, driven by the need to integrate increasing shares of variable wind generation.

In October 2014, the Energy Imbalance Market (EIM) in the Western Interconnection was opened in United States of America interconnecting over 30 participating balancing authorities (BAs) in USA and Canada. This allows for generation and demand balancing across Balancing Authority Areas (BAA) on 15 minute and 5 minute time-scales with California Independent System Operator (CAISO) oversight. Previous to the opening, all the Balancing Authorities (BA) were responsible to balance generation and demand for their own area. Now, the ISO can dispatch and share resources across the participating BAAs to balance energy of all BAs. All EIM market participants are mandated to provide a continuous

feed of specified forecasts to the ISO. Energy imbalances caused by errors in the forecasts and bidding are settled by the defined settling regulation e.g. according to the United States Federal Energy Regulatory Commission (FERC) Order 890, for intermittent renewable generators, imbalances greater than 7.5% or 10 MW are settled at 125 percent incremental cost or 75 percent decremental cost of providing the imbalance energy. In contrast to the dominance of wind as intermittent generator in Northern Europe, solar is the dominating intermittent energy source in many regions in the Western Interconnection with tremendous expected growth rates (e.g. California).

In this context, we aim to analyze and quantify the benefits of solar forecasting for EIM operations. The solar forecasts are implemented to cover all necessary forecast horizons for EIM operations. All the implemented forecasts are state-of-the-art methodologies based on broadly available methods, relying on low-cost instrumentation and publicly available data. The contributions of this study are: 1) reforecast methodology to forecast day-ahead global irradiance 2) features based optimized short-term solar forecasting 3) analysis of solar forecast errors for the forecast horizons related to electricity markets, especially the short-term EIM market in the Western Interconnection and 4) detailed analysis on the role of solar forecasting in EIM in terms of uncertainty and estimation of flexibility reserves from the perspective of market operator and participants.

More detailed discussion on the EIM and previous work on solar resource forecasting are provided in section 6.1, the data sets used are described in previous chapter section 4.1, methods for solar forecasting and corresponding results are provided in previous chapter sections 4.2 and 4.3, the value of forecasting for EIM is shown in section 6.2 and the conclusions are drawn in section 6.3.

## **6.1 Energy imbalance markets**

### **6.1.1 Goals**

The main objective for the introduction of EIMs is to reduce imbalances between demand and generation without ancillary services or additional reserves

by enabling regulated, short-term energy trading between interconnected balancing areas. Without EIMs, resources were not shared between the balancing areas. The individual balancing area authorities had to schedule and keep operating reserves to handle imbalances. For the EIM in the Western Interconnection, after scheduling for 15 minute market, the 5 minute market is executed to automatically procure resources to balance expected imbalances between generation and demand in 5 minute time intervals. Taking the advantage of increased geographical diversity in generation and load profiles, the main benefits of this market are reduced operating reserves capacity, enhanced reliability, reduced costs and automatic dispatch, and real-time visibility.

### **6.1.2 Previous work**

This section covers a short summary of previous work, relevant for EIMs. A review on real-time markets is presented in [138]. An overview of previous EIM studies can be found in [139]. They include a comparison of market regulations based on assumptions, annual benefits, and geographic scope. The study includes benefits of the implemented EIM between ISO and PacifiCorp. The impact for EIM, for grids with high levels of wind penetration, was studied in [140]. They show that the introduction of EIMs enables reserve requirement reductions which is beneficial for all EIM participants. Furthermore, they show that the failure or refusal of participation by as little as one entity can reduce the benefits for all other participants in the market. Using forecasts as a decision variable the bidders and market operator can commit or de-commit in case of high or low energy production [141]. An evaluation of energy balance and imbalance settlements in Europe is presented in [142].

A general framework for analyzing various components of market participation for wind generators was proposed in [143]. They discuss the value of information contained in forecasts for grids with high wind penetration. Conclusions cover that forecasting has a high economic value for variable wind energy sources. For current status of wind penetration in CAISO, investment in short-term wind forecasting is precarious, whereas in future scenarios with high wind penetration

levels, forecasting can evolve into an important decision variable for real-time market operation, e.g. economic dispatch in the CAISO area [144]. A detailed analysis on organized markets in the Western Interconnection can be found in [145]. It highlights the factors influencing the success of EIMs, such as cost allocation, transmission rights, participation of various BAAs, stakeholders and discusses the alternatives to organized markets, for instance Intra-hour Transaction Accelerator Platform, the Dynamic Scheduling System, Balancing Authority Reliability-based Control, Area Control Error Diversity Interchange, Enhance Curtailment Calculator, etc. While these alternative market setups might be beneficial in certain cases, the regulating authorities decided to operate an EIM in the Western Interconnection. The focus of our work is the EIM in the Western Interconnection in the United States.

Most of the studies on the impact of EIM on Western and Eastern Interconnection assume forecasts to be persistent [141]. However, there has been tremendous progress in the field of solar energy resource forecasting over the past decade. Hence, previous studies provide a conservative estimate of reserves. In this study, we seek to quantify the benefits of state-of-art-solar solar forecasts for the EIM. The next section covers key design variables of EIMs and current solar forecasting methods that can be utilized for EIM participation and operation.

### 6.1.3 Market design variables

The two fundamental concepts for energy imbalance markets are balance responsibility and imbalance settlements [142]. Balancing responsibility covers the processes from market opening, to binding and market execution. The key variables for balance responsibilities are: (1) program time unit (PTU), defined as the time window for which bids are submitted and base schedules are awarded. (2) Scope of balancing, defined as the magnitude of necessary generation change. (3) Gate closure time ( $GCT_{p/o}$ ); defining the time when the option to submit or modify a bid expires. The  $GCT_p$  is for market participants while the  $GCT_o$  is time when schedules are binded from the operator. (4) Types of imbalances, depending on if over- or under-generation occurs. (5) Closed (zero imbalance) or open (occurring

imbalance) portfolio positions and (6) look ahead time (LAD), defining the ahead time horizon considered for running the optimization to schedule awards.

The design of imbalance settlements define the detailed setup of penalties associated with wrong forecasts and market bids. Details about imbalance settlements can be found in [142]. In general, it covers the frequency of settlements, regulations and pricing of imbalances for each market participant.

The discussed variables allow for broadly varying market designs. The specific regulation of EIMs vary greatly for different world regions. For instance, in Norway, the first GCT of market execution is 7 pm local time on the day before the market is executed. In Sweden, it is 4 pm and in Finland it is 4:30 pm. The PTU in these regions ranges between 60 and 15 minutes. In the Western Interconnection the first GTC before market execution is 40 minutes and PTUs are 15 to 5 minutes. More details for European EIMs can be found in [142] and for the Western Interconnection, the in-depth details are provided below.

The PTU and GCT are the key technical drivers of imbalance markets.

#### **6.1.4 EIM in Western Interconnection**

The Western Interconnection Energy Imbalance Market (WI EIM) is a centralized and coordinated real-time energy market, operating at 15 and 5 minute time intervals. Before the introduction of the WI EIM, resources were not shared between the participating BAAs. Each BA had to independently schedule operating reserves and backup resources. With the introduction of the WI EIM with over 30 participating BAAs, generation and demand can be exchanged between participating entities with CAISO oversight. Hence, imbalances can be corrected for within almost the full Western Interconnection. For instance, under-forecasts of generation by wind in the Seattle area could be balanced with over-forecasts of power output from solar plants in the Mojave Desert.

The following groups actively participate in the WI EIM: EIM entity (over 30 participating BAAs) represented by an entity scheduling coordinator (ESC), the market operator (CAISO), and participating resources represented by a scheduling coordinator. Non-Participating resources are also playing a role in the EIM since

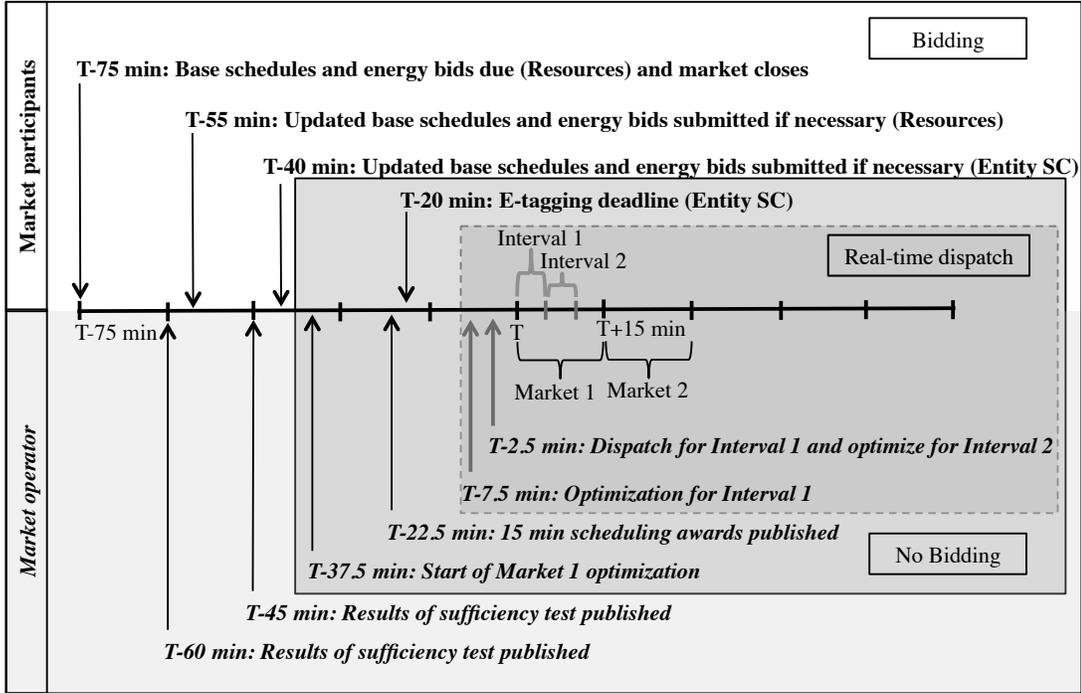
it was shown that the highest benefits are achieved if all resources participate in the market [140].

### **EIM WI market operation**

To determine the most economic dispatch, CAISO automatically accesses the following variables: generation and demand forecasts, outages, resource schedules, economic bids, dynamic contingencies and interchange schedules to run a multi-objective optimization for future planning and scheduling. The detailed time lines of WI EIM market operations are visualized in figure 6.1. The WI EIM process starts 7 days before the operating day,  $D$ . The ESC starts submitting the base schedules and can update the base schedule up to a day before the operating day i.e.  $D - 1$ . On  $D - 1$ , 10:00 am local time, ISO evaluates the base schedules to determine the energy balance, congestion and flexible ramping capacity. This serves as an advisory information for the day-ahead market (DAM). At  $D - 1$ , 1:00 pm local time, ISO runs the sufficiency test and provides the information to the scheduling coordinators. The base schedules and operating bids from the scheduling operators are due 75 minutes before the operating hour i.e.  $T - 75$  minutes where  $T$  represents the start of the operating hour. At  $T - 60$  minutes, ISO runs the sufficiency test and publish results. Based on these results, the updated base schedules and energy bids are due by  $T - 55$  minutes. Again, a sufficiency test is run and results are published. The final base schedules and the energy bids are due by  $T - 40$  minutes. The optimization for the 15 minutes energy market starts at  $T - 37.5$  minutes. Scheduling awards are published at  $T - 22.5$  minutes. The deadline for confirmation between the generator and the market operator (referred to as e-tagging) is due at  $T - 20$  minutes. This process is repeated for every operating hour.

Of all the timelines discussed above,  $T - 75$  minutes is crucial because by this time all interested market participants are required to submit their bids. The  $T - 40$  minutes horizon is the final opportunity ( $GCT_p$ ) for the participants to update their bids based on the sufficiency test results.

For the 5-minute real-time dispatch market, the optimization is run at  $T -$



**Figure 6.1:** Timeline for the real-time Energy Imbalance Market in the Western Interconnection for the operating hour beginning at time  $T$  (Market 1). First the schedules are awarded for 15 minutes market and then 7.5 minutes before the operating hour another optimization is executed and energy is dispatched at  $T - 2.5$  minutes to balance the difference in demand and production for Interval 1.

7.5 minutes interval and at  $T - 2.5$  minutes the resources are dispatched to balance the demand and generation for  $[T, T + 5)$  minute interval. This process is repeated for every 5 minute time intervals.

### 6.1.5 Forecasting for EIMs

The decisions for the market operations are based on forecasts. The optimization problem for the 15 minute market looks ahead up to 4.5 hours in 15 minute time intervals. For the 5 minutes market, the optimization looks ahead up to 65 minutes in 5 minute intervals. These forecasting horizons are of crucial importance for efficient EIM market operation and management. Thus the timeline at which forecast is issued before  $GTC_{p/o}$  and then applied by the vendor or market operator to bid into the market or award base schedules for the given PTU

i.e. the interval of forecast highly influence the dynamics of the market.

In this study we focus on  $k = 24$  hours, 75 – 5 minutes,  $fi, fr = 1$  hour, 15 and 5 minutes. In general, for market application  $k \geq GTC_{p/o}$  and  $fi \geq PTU$ .

To be able to participate in EIM, all the participants are required to provide their resource and load forecasts to the ISO. Since, EIM is designed to balance demand and resource at shorter time scales, in this study we focus on variable energy resource i.e. solar energy that is likely to play an important role in EIM at high penetration levels.

## 6.2 Implications on EIMs

To analyze the implications of solar forecasting for EIM, the forecasts discussed in chapter 4 are used henceforth. For more details about the method, datasets, forecast models and solar forecasting refer to chapter 4.

### 6.2.1 Reserve scheduling

In the United States, all ISOs are required by FERC to keep additional operating reserves to account for errors and any sudden changes in load. Reserve requirements also apply for EIMs and the costs for backup resources effect all market participants. There are various methods in practice to compute reserves [146, 147, 139].

The dominant approach is the n-sigma method, where reserves are calculated by assessing the standard deviation in generation and demand forecast. Based on the forecast above, we can quantify the required resources necessary to cover the uncertainty introduced into the EIM by solar generators with the P, SP and SVR-GA models. Figure 6.2 shows the standard deviations and respective error reduction achieved by SP and SVR-GA model for 1 day and 75 to 5 minute forecast horizons in the intervals of 1 hour, 15 minute and 5 minute averages. As expected, the standard deviation for 15 minute averages is lower than for 5 minute averages. A 15 minute market operation disregards large shares of variance occurring under 5 minute market operation, suggesting that lower reserves have to be

scheduled. This is counter intuitive but can be explained by the fact that temporal smoothing disregards large shares of variability (the solution is discussed further in section 6.2.3). This shows that the dynamics and operations of the market are highly influenced by the PTUs and GCTs.

Depending on GCT, reserves are scheduled. For  $GCT > 75$  minutes, required relative reserve are defined. For  $GCT < 75$  minutes, flexibility reserves are scheduled.

### 6.2.2 EIMs with GCT greater than 75 minutes

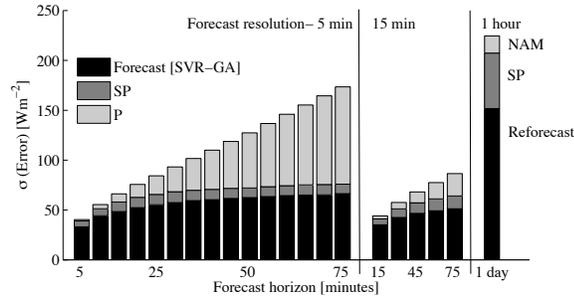
The scheduled relative reserve  $R_r$  for  $GCT > 75$  minutes is by definition a function of standard deviation of forecast errors  $\sigma(e)$  occurring for a given PTU, normalized by the mean annual solar irradiance ( $\overline{GHI}$ ) received on a given location i.e.,

$$R_r = \frac{\sigma(e)}{\overline{GHI}}. \quad (6.1)$$

This approach enables to compare reserve required for different resource forecast approaches. In practice, the standard deviation of power forecast errors is considered (this requires knowledge of specific system characteristics). Resource forecasts directly translate into power forecasts which validates our use of relative reserve as an estimation for operating reserves. For large GCT, NWP predictions have to be used for the resource forecast. As shown above, reforecasting significantly enhances the performance of NWP based GHI prediction. Hence, required reserves can be significantly reduced by the proposed reforecast method (see section 4.2.4). Using the SP model the required reserve is 0.39. Using the NWP reforecast model it decreases to 0.32 which is a reduction of 17.84%.

### 6.2.3 EIMs with GCT lower than 75 minutes

For markets with  $GCT \leq 75$  minutes, the concept of flexibility reserves allocation is used. Flexibility reserves are scheduled as a function of change in standard deviation with respect to the magnitude of solar power production [141]. To



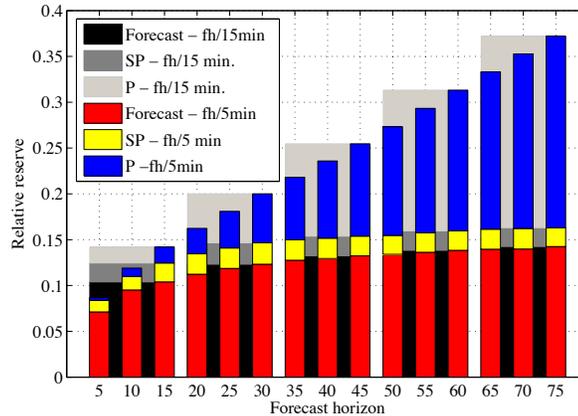
**Figure 6.2:** Standard deviation in forecast error for forecast horizons ranging from 1 day to 5 minutes ahead forecast horizons in 1 hour, 15 minute and 5 minute resolution. The total height of the bar represents the standard deviation in the P forecast errors and then the stacks within a bar shows the reduction in  $\sigma$  by the SP and SVR-GA model. For the one day-ahead forecast, the standard deviation is the highest for the plain NAM forecast, followed by P model. The best performance is achieved by the reforecast model. This chart can be used for any solar forecast application study to design the modeling parameters for the uncertainty at various time-horizons.

show the comparison with  $GCT > 75$  minutes, relative reserves are also computed.

As discussed above, to account for the problem of temporal smoothing the resolution of the forecast  $fr$  has to be considered i.e., for a given  $fr$  and  $fi$  there will be  $n$  number of forecasts such that  $n = \frac{fi}{fr}$ . The forecast errors for all these  $n$  forecasts are represented as  $e_l$  such that  $l \in \{1, 2, \dots, n\}$ . Then, the standard deviation used to compute the relative reserve has to be greater than or equal to the maximum of standard deviation of  $e_l$  forecast errors,

$$\sigma(e)^{fr} \geq \max_l(\sigma(e_l)), \quad (6.2)$$

where  $\sigma(e)^{fr}$  is the standard deviation in  $n$  forecast errors occurring in a given PTU with a forecast resolution  $fr$ . For example, if the market operates on  $PTU = 15$  minutes, and the reserves allocated need to cover the variance up to 5 minute time resolution, then  $n = 3$ . Figure 6.3 shows the relative reserves allocation for various forecast-horizons in 5 and 15 minute intervals, considering the 5 minute variance. The reduction of reserve required using P model compared to SP and SVR-GA is shown. The grey-scale bar represents the reserve allocation for 15 minute interval. Colored bars represent the reserve allocation for the 5 minute

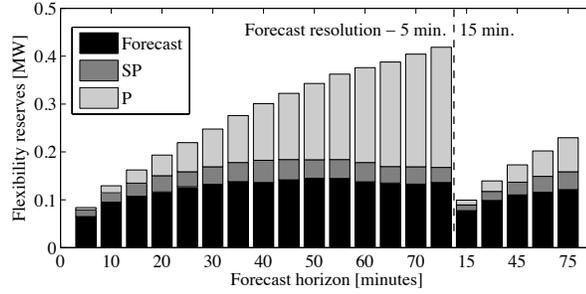


**Figure 6.3:** Benefits of 5 minute versus 15 minute reserve allocation for real-time markets considering 5 minute data interval. The grey-scale bar plots shows the relative reserve required for the 15 min-time intervals whereas the colored bar represents the reserves required for the 5 minute time-intervals. The total height of the bar represents the total reserve required using the persistence model and reduction in the magnitude of reserves is shown by the stacks within each bar.

time interval. It is beneficial for the market operator to allocate resources on shorter time horizons. For short-term forecast horizon (5-15 minute), the reduction is smaller whereas for the higher forecast horizons, the benefits of using SP and forecast model become apparent. The relative reserved can be reduced by 28.5% on average. For 45 minutes forecast horizon, the relative reserve required is 0.25 and it can be reduced by 40% using the SP model.

### Market operator - schedulers

The computed flexibility reserves are shown in figure 6.4). The reduction of flexibility reserves using the P model, compared to SP and SVR-GA are summarized in table 6.1. Using the SP model the improvement ranges between 5.83% to 59.97% for the 5 to 75 minutes forecast horizon, respectively. The reductions using SVR-GA ranges between 22.62% and 67.47%, respectively. Using SVR-GA instead of SP, the flexibility reserves can be reduced by 21%, on average. However, reductions are lower for the 15 minutes forecast intervals. This is expected because at 15 minute resolution the resource variability is lower than for 5 minutes.



**Figure 6.4:** Flexibility reserve required for 1 MW solar plant for 5 minute intervals. The total magnitude of the bar represents the flexibility reserve needed using the Persistence model and the inner stack represents the magnitude reduction by Smart persistence and the forecast model. Beginning from forecast horizon of 5 minutes to 55 minutes, the improvements in reducing flexibility reserves increase and for the forecast horizons greater than 55 minutes, the improvements starts decreasing due to reduced forecast skill.

### Market participant - power generators

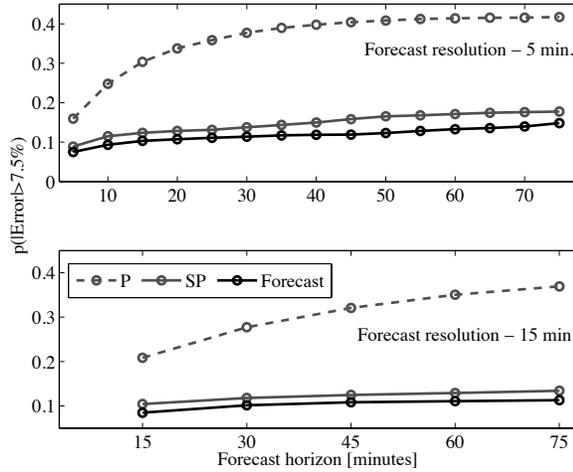
The technical requirements of the power grid depend upon very small discrepancies between generation and load. This drives the need for imbalance settlements. For intermittent renewable generators, FERC defines errors in market bids larger than 7.5 percent as a benchmark for penalties. Hence, we define errors larger than 7.5% as occurring imbalance. To quantify the likelihood that imbalances will occur, the probability of wrong bids through errors in forecasts for 5 minute intervals for a 1MW solar plant are calculated. For 5 to 75 minutes forecast, the probability ranges between 0.15 to 0.41, 0.08 to 0.17, and 0.07 to 0.14 for persistence, smart persistence and forecast model, respectively (see Figure 6.5). Thus, the probability of imbalance can be reduced by 66.93% using SVR-GA instead of P (or 19.65% instead of SP model). Similar results are achieved for the 15 minute forecast resolution, where the probability of an imbalance is reduced by 15.12% by replacing SP with SVR-GA.

### 6.2.4 Implications for the Western Interconnection EIM

Most imbalance markets around the world are operated at 15 minute or greater PTU, however the Energy Imbalance Market in the Western Interconnec-

**Table 6.1:** Summary of benefits of solar forecasting for the real-time energy markets with respect to the Persistence (P) and Smart persistence (SP) forecast model using solar power data for 1MW plant.

Forecast horizon (minutes)	Reduction w.r.t. P model [%]				Reduction w.r.t. SP model [%]	
	Flexibility reserve		$p(\text{Error} > 7.5\%)$		FR	$p(\text{Error} > 7.5\%)$
	SP	SVR-GA	SP	SVR-GA	SVR-GA	SVR-GA
	<i>fr</i> – 5 minutes					
5	5.83	22.62	44.38	53.04	17.83	15.64
10	11.23	26.41	53.66	62.40	17.11	18.85
15	16.99	33.89	59.32	66.03	20.36	16.48
20	22.18	39.76	62.01	68.20	22.59	16.30
25	27.65	41.93	63.43	69.06	19.73	15.38
30	31.80	46.63	63.47	69.80	21.74	17.32
35	35.51	50.02	63.19	69.91	22.49	18.25
40	39.31	54.70	62.35	70.20	25.36	20.85
45	42.89	56.01	60.79	70.55	22.97	24.89
50	46.49	57.69	59.51	68.89	20.94	24.82
55	49.24	60.10	59.26	68.89	21.40	25.64
60	52.69	63.43	58.65	67.88	22.71	23.32
65	56.27	65.16	58.02	67.32	20.32	22.15
70	58.19	67.16	57.62	66.39	21.45	20.68
75	59.97	67.47	57.47	64.45	18.75	16.41
	<i>fr</i> – 15 minutes					
15	10.20	22.29	50.02	59.30	13.47	18.57
30	15.72	29.27	57.43	63.38	16.09	13.96
45	20.91	36.32	61.16	66.25	19.48	13.11
60	26.19	42.67	63.12	68.42	22.34	14.33
75	30.99	47.19	63.72	69.40	23.48	15.65

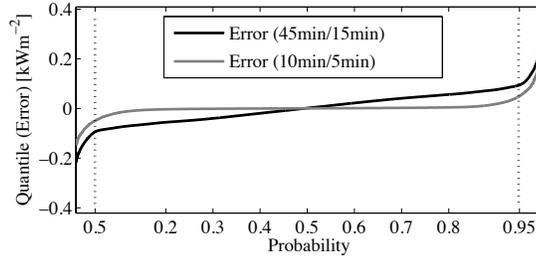


**Figure 6.5:** Energy imbalances (probability for the error to exceed 7.5%) for 1 MW solar plant for 5-75 minutes and 15-75 minutes forecast horizon with 5 minute and 15 minute resolution. The probability of imbalance greater than 7.5% can be reduced by an average of 19.65% and 15.12% by using a forecast model instead of a SP model.

tion is a unique platform where after a 15 minute market, another 5 minute market is operated. This section present results with respect to the timescales for Western Interconnection from the perspective of both, the market operator and market participants.

Let assume the market operator starts the optimization to execute the WI EIM at  $T - 37.5$  minutes to plan for the interval  $[T, T + 15)$  using the state-of-the-art forecast issued at  $T - 45$  minute. Then, in this time interval the forecast error lies between  $-0.8$  to  $0.09$   $\text{kWm}^{-2}$  for a 90% confidence interval. Therefore, if another state-of-the art forecast for a forecast horizon of  $T - 10$  minutes is added to the system before the optimization runs for  $T - 7.5$  minute, the uncertainty of over-prediction can be reduced from  $-0.6$  to  $0.04$   $\text{kWm}^{-2}$ , representing a 25% reduction and 41% overall reduction of variance (see Figure 6.6).

For WI EIM participation, market participants are mandated to submit a bid at  $T - 75$  minute for the  $[T, T + 15)$  time-interval. Using a forecast model instead of SP model, the chances of getting penalized reduces by 15.65%. Furthermore, if the bid is updated  $T - 45$  minute, there is a 13.11% reduction in chances of getting penalized using SVR-GA as compared to the SP model.



**Figure 6.6:** Inverse cumulative frequency distribution of the expected forecast errors that market operator has to optimize for in 5-minute market after scheduling the awards for 15 minute market using the forecast model. Results show that if a forecast model is applied at 10 minutes forecast horizon in five minute resolution the uncertainty can be reduced from  $-0.08$  to  $-0.06$   $\text{kW m}^{-2}$  and from  $0.09$  to  $0.04$   $\text{kWm}^{-2}$  for 90% of the time.

### 6.3 Summary

A review about energy imbalance markets is presented. The key design variables are explained and their significance is discussed. It is shown that the program time unit as well as the gate closure time are the key variables controlling the market dynamics. The process of imbalance settlements are the key regulatory and economical drivers for market participants and the operator.

The application of the discussed forecasts reduce the relative reserve requirement for participation in imbalance markets with day-ahead gate closure time by 17.84%. Similarly for EIMs with shorter gate closure time, the flexibility reserves for an explanatory 1 MW solar plant can be reduced by 21% on average, for the 5 minute forecast-intervals. For 15 minute forecasts, the improvements are 16.14% compared to the smart persistence model. The improvements for the 15-minute intervals are smaller than for 5 minute intervals due to variability smoothing. For participating commercial solar producers, the probability of causing an imbalance is reduced by 19.65% and 15.12% for 5 minute and 15 minute forecast intervals. The improvements achieved through forecasting depend on market operation time scales.

## Acknowledgments

This chapter contains work previously submitted to journal:

- A. Kaur, L. Nonnenmacher, H.T.C. Pedro and C.F.M. Coimbra, “Benefits of solar forecasting for energy imbalance market”, *Renewable Energy*, [Submitted, March 2015].

# Chapter 7

## Conclusions

The electric power industry is changing rapidly, mainly driven by new technologies, consumer demand and new environmental regulations. Solar and wind power plants are appearing everywhere, conventional vehicles are being replaced with electric plug-in vehicles, etc. With all these changes, our net load demand profiles are transforming with a lot of uncertainty. For a smooth transition and reliable integration of various renewable technologies, there is a need to quantify the effects of these technologies to develop solutions. This study focused on electric grids with high solar penetration and proposed various forecasting methods to reduce the associated uncertainty in net load and solar generation.

We reviewed the current state of the art for load and solar forecasting and proposed ensemble re-forecasting methods that further refine the existing forecasts by learning from the past structured errors. Correlation analysis of the forecasts issued by Independent System Operators revealed presence of structured errors, especially at off-peak hours. The proposed ensemble re-forecast HAM predictions show an improvement of 47% and 36% in terms of Mean Absolute Percentage Error over the forecasts provided by CAISO and ERCOT. For DAM, the improvements are 34% for CAISO and 47% for ERCOT. Temporal analysis comparing the internal forecast produced by the ISOs and re-forecast shows significant improvement during off-peak hours and small improvement for on-peak hours.

The application of our proposed ensemble reforecast on solar irradiance previously predicted by NAM model achieved a skill of 14.5% at 1-24 hours lead

times. Results validate the potential of the proposed methodology to enhance the forecast accuracy, independent of time-series profile (load and solar) or forecast horizon.

Next, the models for short-term solar forecasting are proposed and validated using ground data and imagery as inputs. The skill achieved by features-based forecasting model ranges between 14.53-20.11% and 12.64-16.75% for 15 minute and 5 minute forecast intervals (forecast horizons ranging from 5 to 75 minutes). For forecasts with forecast horizon greater than or equal to 35 minutes, the skill drops implying that more information is required e.g. satellite imagery.

For remote solar power plants with no additional inputs, heuristics based approach is proposed. The proposed method takes into account the changing atmospheric clearness and efficiency degradation of PV panels due to dust, high temperature, etc. The adaptive clear sky model and heuristics proposed can be applied to any solar forecast algorithm to enhance the forecast accuracy. For instance, when applied to smart persistence, the adaptive clear sky and heuristics enhanced forecast accuracy by 9.4% and 22.6% respectively, and the improvements for SVR based model were 11.41% and 24.20% respectively.

Furthermore, we quantified the impact of solar power on our net load demand and deduced that it is solar variability that drives the forecast error for high penetration communities. We showed that the forecast error distribution for grids with solar generation is best characterized as a t-distribution, where heavy tails reflect errors due to high ramping events in the solar power output.

To mitigate this effect of increasing solar penetration, we propose integrating solar and load forecasts within a single platform. Optimal methods to combine solar and load forecasting are compared: additive and integrated net load forecast models. For utility-scale grids and microgrid based applications, an integrated forecast model is recommended because the 95% confidence interval of the net load errors for the additive model is greater than the integrated model by 2.2% of the maximum net load demand. The net load and solar forecasting errors exhibited correlation which was validated with Engle-Granger cointegration test. Indeed, the two time-series are cointegrated with common stochastic drift. Therefore, so-

lar time-series suffice to characterize the expected variability and uncertainty in the net load and solar power generation.

The benefits of the proposed solar forecasting techniques are quantified for real-time energy imbalance market. The benefits of 5-minute forecast are higher than the benefits of 15-minute forecast due to higher variability at 5-minute resolution and better information at shorter forecast horizon. Thus, the benefits of forecasting for grid applications are dependent upon the market operation timescales.

In a nutshell, reforecast is a powerful technique to enhance forecast accuracy for any time-series independent of forecast horizon. The solar resource variability has a more pronounced effect on forecast errors rather than the penetration level. Our capabilities to mitigate intermittence are influenced by local solar micro-climate. Hence, for future planning and siting of local solar farms, the solar variability of a given location should be considered in addition to solar resource assessment. The proposed forecast methods will enable the grid operators and power generators to efficiently manage the grids with increasing intermittent renewable generation and participate in electricity markets for economical benefits. The design variables regulating the electricity markets, power grid time lines, the bidding and the binding schedules, govern the dynamics of power grids and drive the benefits of forecasting. Therefore, having the flexibility to operate on shorter time-horizons is the key to optimally manage such grids with high intermittent renewable penetration.

## Acknowledgments

This chapter contains work from previously published and submitted journal articles:

- A. Kaur, H.T.C. Pedro, and C.F.M. Coimbra, “Impact of onsite solar generation on system load demand forecast”, *Energy Conversion and Management* 75 (0) (2013) 701 - 709.
- A. Kaur, H.T.C. Pedro, and C.F.M. Coimbra, “Ensemble re-forecasting meth-

ods for enhanced power load prediction”, *Energy Conversion and Management* 80 (0) (2014) 582 - 590.

- A. Kaur, L. Nonnenmacher and C.F.M. Coimbra, “Net load forecasting for high renewable energy penetration grids”, *Energy*, [Submitted, December 2014].
- A. Kaur, L. Nonnenmacher, H.T.C. Pedro and C.F.M. Coimbra, “Benefits of solar forecasting for energy imbalance market”, *Renewable Energy*, [Submitted, March 2015].

# Bibliography

- [1] EIA, Annual energy outlook 2015 with projection to 2040, Report DOE/EIA-0383, U.S. Energy Information Administration (April 2015).
- [2] EPIA - European Photovoltaic Industry Association, Global market outlook for photovoltaics 2014-2018 (2014).
- [3] WECC Standard BAL-STD-002-0 Operating Reserves (2010).
- [4] 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, 2006.
- [5] L. Suganthi, A. A. Samuel, Energy models for demand forecasting a review, Renewable and Sustainable Energy Reviews 16 (2) (2012) 1223 – 1240.
- [6] P. D. Matthewman, H. Nicholosl, Techniques for load prediction in the electricity-supply industry, Proceedings of the IEEE 115 (10) (1968) 1451–1457.
- [7] H. S. Hippert, C. E. Pedreira, R. C. Souza, Neural networks for short-term load forecasting: A review and evaluation, IEEE transactions on Power Systems 16 (1) (2001) 44–55.
- [8] CAISO - Today's Outlook, <http://www.caiso.com/SystemStatus.html>, [cited July 2013].
- [9] ERCOT, <http://www.ercot.com/content/cdr/html/loadForecastVsActualCurrentDay.html>, [cited July 2013].
- [10] H. Willis, J. Northcotegreen, Spatial electric-load forecasting-a tutorial review, Proceedings of the IEEE 71 (1983) 232–253.
- [11] 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.

- [12] I. Moghram, S. Rahman, Analysis and evaluation of five short-term load forecasting techniques, *Power Systems, IEEE Transactions on* 4 (4) (1989) 1484 –1491.
- [13] 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.
- [14] S. Tzafestas, E. Tzafestas, Computational intelligence techniques for short-term electric load forecasting, *Journal of intelligent and robotic systems* 31 (1/3) (2001) 7–68.
- [15] H. K. Alfares, M. Nazeeruddin, Electric load forecasting: literature survey and classification of methods, *International Journal of Systems Science* 33 (1) (2002) 23–34.
- [16] M. Adya, F. Collopy, How effective are neural networks at forecasting and prediction ? a review and evaluation, *Journal of Forecasting* 17 (1998) 481–495.
- [17] S. F. Crone, M. Hibon, K. Nikolopoulos, Advances in forecasting with neural networks? Empirical evidence from the NN3 competition on time series prediction, *International Journal of Forecasting* 27 (3) (2011) 635 – 660.
- [18] 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.
- [19] R. H. Inman, H. T. C. Pedro, C. F. M. Coimbra, Solar forecasting methods for renewable energy integration, *Progress in Energy and Combustion Science*.
- [20] 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 (0) (2013) 701 – 709.
- [21] R. Marquez, H. T. C. Pedro, C. F. M. Coimbra, Hybrid solar forecasting method uses satellite imaging and ground telemetry as inputs to {ANNs}, *Solar Energy* 92 (0) (2013) 176 – 188.
- [22] R. Marquez, C. F. M. Coimbra, Intra-hour DNI forecasting based on cloud tracking image analysis, *Solar Energy* 91 (2013) 327 – 336.
- [23] H. T. C. Pedro, C. F. M. Coimbra, Assessment of forecasting techniques for solar power production with no exogenous inputs, *Solar Energy* 86 (7) (2012) 2017 – 2028.

- [24] R. Marquez, V. G. Gueorguiev., C. F. M. Coimbra, Forecasting of global horizontal irradiance using sky cover indices, *Journal of Solar Energy Engineering* 135 (1) (2012) 011017–011017.
- [25] I. Drezga, S. Rahman, Short-term load forecasting with local ANN predictors, *Power Systems, IEEE Transactions on* 14 (3) (1999) 844 –850.
- [26] J. W. Taylor, R. Buizza, Neural network load forecasting with weather ensemble predictions, *Power Systems, IEEE Transactions on* 17 (3) (2002) 626 – 632.
- [27] J. W. Taylor, R. Buizza, Using weather ensemble predictions in electricity demand forecasting, *International Journal of Forecasting* 19 (1) (2003) 57 – 70.
- [28] R. E. Abdel-Aal, Improving electric load forecasts using network committees, *Electric Power Systems Research* 74 (1) (2005) 83 – 94.
- [29] S. Fan, L. Chen, W.-J. Lee, Short-term load forecasting using comprehensive combination based on multimeteorological information, *Industry Applications, IEEE Transactions on* 45 (4) (2009) 1460 –1466.
- [30] 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.
- [31] K. Siwek, S. Osowski, R. Szupiluk, Ensemble neural network approach for accurate load forecasting in a power system, *International Journal of Applied Mathematics and Computer Science* 19 (2) (2009) 303–315.
- [32] Y. Li, D.-F. Wang, P. Han, Selective ensemble using discrete differential evolution algorithm for short-term load forecasting, in: *Machine Learning and Cybernetics, 2009 International Conference on*, Vol. 1, 2009, pp. 80 –85.
- [33] M. Matijas, J. A. K. Suykens, S. Krajcar, Load forecasting using a multivariate meta-learning system, *Expert Systems with Applications* 40 (11) (2013) 4427 – 4437.
- [34] P. Ansarimehr, S. Barghinia, H. Habibi, N. Vafadar, Short term load forecasting for Iran national power system using artificial neural network and fuzzy expert system, in: *Power System Technology, 2002. Proceedings. PowerCon 2002. International Conference on*, Vol. 2, 2002, pp. 1082–1085 vol.2.
- [35] G.-C. Liao, T.-P. Tsao, Application of a fuzzy neural network combined with a chaos genetic algorithm and simulated annealing to short-term load forecasting, *Evolutionary Computation, IEEE Transactions on* 10 (3) (2006) 330–340.

- [36] L. Ljung, *System Identification: Theory for the User*, Prentice Hall PTR, New Jersey, USA, 1999.
- [37] K. M. Passino, *Biomimicry for Optimization, Control, and Automation.*, Springer-Verlag, London, U.K., 2005.
- [38] CAISO, <http://www.caiso.com>, [cited July 2013].
- [39] FERC, <http://www.ferc.gov/market-oversight/mkt-electric/california.asp>, [cited July 2013].
- [40] ERCOT, <http://www.ercot.com>, [cited July 2013].
- [41] California Public Utilities Commission, Renewables portfolio standard, quarterly report, 1st quarter 2011 (2011).
- [42] D. Lew, N. Miller, K. Clark, G. Jordan, Z. Gao, Impact of high solar penetration in the western interconnection, Tech. Rep. NREL/TP-5500-49667, NREL (December 2010).
- [43] P. Denholm, R. M. Margolis, Evaluating the limits of solar photovoltaics(pv) in electric power systems utilizing the energy storage and other enabling technologies, *Energy Policy* 35 (2007) 4424–4433.
- [44] P. N. Price, Methods for analyzing electric load shape and its variability, Report LBNL-3713E, LBNL,CEC (August 2010).
- [45] D. Lew, G. Brinkman, E. Ibanez, M. Hummon, B. M. Hodge, M. Heaney, Sub-hourly impacts of high solar penetrations in the western united states, Tech. Rep. NREL/CP-5500-56171, NREL (September 2012).
- [46] P. Denholm, M. Hand, Grid flexibility and storage required to achieve very high penetration of variable renewable electricity, *Energy Policy*.
- [47] G. Gross, F. Galiana, Short-term load forecasting, *Proceedings of the IEEE* 75 (12) (1987) 1558–1573.
- [48] G. E. Oamek, B. C. English, A review of load forecasting methodologies, Center for Agricultural and Rural Development, Iowa state University 84-WP-5.
- [49] M. A. Abu-El-Magd, N. K. Sinha, Short-term load demand modeling and forecasting: A review, *Systems, Man and Cybernetics, IEEE Transactions on* 12 (3) (1982) 370–382.
- [50] J. Taylor, P. E. McSharry, Short-term load forecasting methods: An evaluation based on european data, *IEEE transactions on Power Systems* 22 (2008) 2213–2219.

- [51] Y. Chakhchoukh, P. Panciatici, L. Mili, Electric load forecasting based on statistical robust methods, *Power Systems, IEEE Transactions on* 26 (3) (2011) 982 –991.
- [52] C. Garcia-Ascanio, C. Mate, Electric power demand forecasting using interval time series: A comparison between var and imlp, *Energy Policy* 38 (2) (2010) 715 – 725.
- [53] M. Lopez, S. Valero, C. Senabre, J. Aparicio, A. Gabaldon, Application of som neural networks to short-term load forecasting: The spanish electricity market case study, *Electric Power Systems Research* 91 (0) (2012) 18 – 27.
- [54] W. C. Hong, Electric load forecasting by seasonal recurrent svr (support vector regression) with chaotic artificial bee colony algorithm, *Energy* 36 (9) (2011) 5568 – 5578.
- [55] X. Yang, J. Yuan, J. Yuan, H. Mao, An improved wm method based on pso for electric load forecasting, *Expert Systems with Applications* 37 (12) (2010) 8036 – 8041.
- [56] C. Xia, J. Wang, K. McMenemy, Short, medium and long term load forecasting model and virtual load forecaster based on radial basis function neural networks, *International Journal of Electrical Power and Energy Systems* 32 (7) (2010) 743 – 750.
- [57] M. De Felice, X. Yao, Short-term load forecasting with neural network ensembles: A comparative study [application notes], *Computational Intelligence Magazine, IEEE* 6 (3) (2011) 47 –56.
- [58] E. Elattar, J. Goulermas, Q. Wu, Generalized locally weighted gmdh for short term load forecasting, *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on* 42 (3) (2012) 345 –356.
- [59] J. W. Taylor, Short-term load forecasting with exponentially weighted methods, *Power Systems, IEEE Transactions on* 27 (1) (2012) 458 –464.
- [60] R. Mamlook, O. Badran, E. Abdulhadi, A fuzzy inference model for short-term load forecasting, *Energy Policy* 37 (4) (2009) 1239 – 1248.
- [61] M. Rejc, M. Pantos, Short-term transmission-loss forecast for the slovenian transmission power system based on a fuzzy-logic decision approach, *Power Systems, IEEE Transactions on* 26 (3) (2011) 1511 –1521.
- [62] M. Alamaniotis, A. Ikonopoulou, L. Tsoukalas, Evolutionary multiobjective optimization of kernel-based very-short-term load forecasting, *Power Systems, IEEE Transactions on* 27 (3) (2012) 1477 –1484.

- [63] H. Kebriaei, B. Araabi, A. Rahimi-Kian, Short-term load forecasting with a new nonsymmetric penalty function, *Power Systems, IEEE Transactions on* 26 (4) (2011) 1817–1825.
- [64] H. Bludszuweit, J. A. Dominguez-Navarro, A. Llombart, Statistical analysis of wind power forecast error, *IEEE Transactions on Power Systems* 23 (08858950) (2008) 983–991.
- [65] B. M. Hodge, M. Milligan, Wind power forecasting error distributions over multiple timescales, *IEEE (1944-9925)* (2011) 1–8.
- [66] M. Lange, On the uncertainty of wind power predictions-analysis of the forecast accuracy and statistical distribution of errors, *Journal of Solar Energy Engineering* 127 (2005) 177–184.
- [67] GE, Analysis of wind generation impact on ERCOT ancillary services requirements, Report, GE Energy, Electric Reliability Council of Texas (March 2008).
- [68] R. O. Duda, P. E. Hart, D. G. Stork, *Pattern Classification*, John Wiley & Sons, Inc, 2000.
- [69] R. Marquez, C. F. M. Coimbra, A proposed metric for evaluation of solar forecasting models, *Journal of Solar Engineering* 135.
- [70] B. Hodge, D. Lew, M. Milligan, Short-term load forecasting error distributions and implications for renewable integration studies, Tech. Rep. NREL/CP-5500-57340, NREL (January 2013).
- [71] R. Marquez, B. Davis, A. Kaur, C. F. M. Coimbra, 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, 2012.
- [72] L. Nonnenmacher, C. F. M. Coimbra, Streamline-based method for intra-day solar forecasting through remote sensing, *Solar Energy* 108 (0) (2014) 447–459.
- [73] 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.
- [74] Y. 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.

- [75] 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.
- [76] Y. Chu, B. Urquhart, S. M. I. Gohari, H. T. C. Pedro, J. Kleissl, C. F. M. Coimbra, Short-term reforecasting of power output from a 48 mwe solar pv plant, *Solar Energy* 112 (2015) 68–77.
- [77] B. O. Akyurek, A. S. Akyurek, J. Kleissl, T. S. Rosing, Tesla: Taylor expanded solar analog forecasting, in: 2014 IEEE International Conference on Smart Grid Communications (SmartGridComm), 2014, pp. 127–132.
- [78] D. Bernecker, C. Riess, E. Angelopoulou, J. Hornegger, Continuous short-term irradiance forecasts using sky images, *Solar Energy* 110 (0) (2014) 303 – 315.
- [79] A. Zagouras, H. T. C. Pedro, C. F. M. Coimbra, On the role of lagged exogenous variables and spatiotemporal correlations in improving the accuracy of solar forecasting methods, *Renewable Energy* 78 (0) (2015) 203 – 218.
- [80] R. Perez, E. Lorenz, S. Pelland, M. Beauharnois, G. V. Knowe, K. H. Jr., D. Heinemann, J. Remund, S. C. Mller, W. Traunmller, G. Steinmauer, D. Pozo, J. A. Ruiz-Arias, V. Lara-Fanego, L. Ramirez-Santigosa, M. Gaston-Romero, L. M. Pomares, Comparison of numerical weather prediction solar irradiance forecasts in the us, canada and europe, *Solar Energy* 94 (0) (2013) 305 – 326.
- [81] P. Mathiesen, J. Kleissl, Evaluation of numerical weather prediction for intraday solar forecasting in the continental united states, *Solar Energy* 85 (5) (2011) 967 – 977.
- [82] L. Nonnenmacher, A. Kaur, C. F. M. Coimbra, Verification of the suny direct normal irradiance model with ground measurements, *Solar Energy* 99 (0) (2014) 246 – 258.
- [83] 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 (0) (2015) 770 – 782.
- [84] P. Ineichen, Comparison of eight clear sky broadband models against 16 independent data banks, *Solar Energy* 80 (4) (2006) 468–478.
- [85] P. Ineichen, R. Perez, A new airmass independent formulation for the Linke turbidity coefficient, *Solar Energy* 73 (3) (2002) 151–157.

- [86] C. Gueymard, Clear-sky irradiance predictions for solar resource mapping and large-scale applications: Improved validation methodology and detailed performance analysis of 18 broadband radiative models, *Solar Energy* 86 (8) (2012) 2145 – 2169.
- [87] V. Vapnik, S. E. Golowich, A. Smola, Support vector method for function approximation, regression estimation, and signal processing, *Advances in neural information processing systems* (1997) 281–287.
- [88] A. J. Smola, B. Scholkopf, A tutorial on support vector regression, *Statistics and Computing* 14 (3) (2004) 199–222.
- [89] J. H. Holland, *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence.*, U Michigan Press, 1975.
- [90] P. Castillo, J. Merelo, A. Prieto, V. Rivas, G. Romero, G-prop: Global optimization of multilayer perceptrons using gas, *Neurocomputing* 35 (1) (2000) 149–163.
- [91] A. Kaur, H. T. C. Pedro, C. F. M. Coimbra, Ensemble re-forecasting methods for enhanced power load prediction, *Energy Conversion and Management* 80 (0) (2014) 582 – 590.
- [92] R. Marquez, C. F. M. Coimbra, Proposed metric for evaluation of solar forecasting models, *ASME Journal of Solar Energy Engineering* 135 (2013) 0110161–0110169.
- [93] C.-C. Chang, C.-J. Lin, LIBSVM: A library for support vector machines, *ACM Transactions on Intelligent Systems and Technology* 2 (2011) 27:1–27:27, software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [94] A. Zagouras, R. H. Inman, C. F. M. Coimbra, On the determination of coherent solar microclimates for utility planning and operations, *Solar Energy* 102 (0) (2014) 173 – 188.
- [95] E. W. Law, A. A. Prasad, M. Kay, R. A. Taylor, Direct normal irradiance forecasting and its application to concentrated solar thermal output forecasting a review, *Solar Energy* 108 (0) (2014) 287 – 307.
- [96] A. Tascikaraoglu, M. Uzunoglu, A review of combined approaches for prediction of short-term wind speed and power, *Renewable and Sustainable Energy Reviews* 34 (0) (2014) 243 – 254.
- [97] J. Jung, R. P. Broadwater, Current status and future advances for wind speed and power forecasting, *Renewable and Sustainable Energy Reviews* 31 (0) (2014) 762 – 777.

- [98] S. Quesada-Ruiz, Y. Chu, J. Tovar-Pescador, H. T. C. Pedro, C. F. M. Coimbra, Cloud-tracking methodology for intra-hour {DNI} forecasting, *Solar Energy* 102 (0) (2014) 267 – 275.
- [99] A. Mellit, M. Benghaneim, S. Kalogirou, An adaptive wavelet-network model for forecasting daily total solar-radiation, *Applied Energy* 83 (7) (2006) 705 – 722.
- [100] B. Amrouche, X. L. Pivert, Artificial neural network based daily local forecasting for global solar radiation, *Applied Energy* 130 (0) (2014) 333 – 341.
- [101] M. Kudo, A. Takeuchi, Y. Nozaki, H. Endo, J. Sumita, Forecasting electric power generation in a photovoltaic power system for an energy network, *Electrical Engineering in Japan* 167 (4) (2009) 16–23.
- [102] P. Bacher, H. Madsen, H. A. Nielsen, Online short-term solar power forecasting, *Solar Energy* 83 (10) (2009) 1772 – 1783.
- [103] C. Tao, D. Shanxu, C. Changsong, Forecasting power output for grid-connected photovoltaic power system without using solar radiation measurement, in: *Power Electronics for Distributed Generation Systems (PEDG)*, 2010 2nd IEEE International Symposium on, 2010, pp. 773–777.
- [104] M. Hassanzadeh, M. Etezadi-Amoli, M. Fadali, Practical approach for sub-hourly and hourly prediction of pv power output, in: *North American Power Symposium (NAPS)*, 2010, 2010, pp. 1–5.
- [105] E. Lorenz, T. Scheidsteger, J. Hurka, D. Heinemann, C. Kurz, Regional pv power prediction for improved grid integration, *Progress in Photovoltaics: Research and Applications* 19 (7) (2011) 757–771.
- [106] C. Chen, S. Duan, T. Cai, B. Liu, Online 24-h solar power forecasting based on weather type classification using artificial neural network, *Solar Energy* 85 (11) (2011) 2856 – 2870.
- [107] J. Shi, W. jen Lee, Y. Liu, Y. Yang, P. Wang, Forecasting power output of photovoltaic systems based on weather classification and support vector machines, *Industry Applications, IEEE Transactions on* 48 (3) (2012) 1064–1069.
- [108] A. Yona, T. Senjyu, T. Funabashi, C.-H. Kim, Determination method of insolation prediction with fuzzy and applying neural network for long-term ahead pv power output correction, *IEEE Transactions on Sustainable Energy* 4 (2) (2013) 527–533.

- [109] H. Yang, C. Huang, Y. Huang, Y. Pai, A weather-based hybrid method for 1-day ahead hourly forecasting of pv power output, *IEEE Transactions on Sustainable Energy* 5 (3) (2014) 917–926.
- [110] M. Mani, R. Pillai, Impact of dust on solar photovoltaic (pv) performance: Research status, challenges and recommendations, *Renewable and Sustainable Energy Reviews* 14 (9) (2010) 3124 – 3131.
- [111] F. A. Mejia, J. Kleissl, Soiling losses for solar photovoltaic systems in california, *Solar Energy* 95 (0) (2013) 357 – 363.
- [112] M. D. Felice, A. Alessandri, F. Catalano, Seasonal climate forecasts for medium-term electricity demand forecasting, *Applied Energy* 137 (0) (2015) 435 – 444.
- [113] P.-F. Pai, W.-C. Hong, Support vector machines with simulated annealing algorithms in electricity load forecasting, *Energy Conversion and Management* 46 (17) (2005) 2669 – 2688.
- [114] B.-J. Chen, M.-W. Chang, C.-J. Lin, Load forecasting using support vector machines: a study on eunite competition 2001, *IEEE Transactions on Power Systems* 19 (4) (2004) 1821–1830.
- [115] N. Liu, Q. Tang, J. Zhang, W. Fan, J. Liu, A hybrid forecasting model with parameter optimization for short-term load forecasting of micro-grids, *Applied Energy* 129 (0) (2014) 336 – 345.
- [116] W.-C. Hong, Electric load forecasting by support vector model, *Applied Mathematical Modeling* 33 (5) (2009) 2444 – 2454.
- [117] A. Kavousi-Fard, H. Samet, F. Marzbani, A new hybrid modified firefly algorithm and support vector regression model for accurate short term load forecasting, *Expert Systems with Applications* 41 (13) (2014) 6047 – 6056.
- [118] KEMA, Prepared for Prepared For: Massachusetts Department of Energy Resources and Massachusetts Clean Energy Center, Massachusetts distributed generation interconnection report - final report (2011).
- [119] Y. V. Makarov, Z. Huang, P. V. Etingov, J. Ma, R. T. Guttromson, K. Subbarao, B. B. Chakrabarti, Incorporating wind generation and load forecast uncertainties into power grid operations, Pacific Northwest National Laboratory, 2010.
- [120] C. Villarreal, D. Erickson, M. Zafar, Microgrids: A regulatory perspective, Prepared by: California Public Utilities Commission Policy & Planning Division.

- [121] M. Soshinskaya, W. H. J. Crijns-Graus, J. M. Guerrero, J. C. Vasquez, Microgrids: Experiences, barriers and success factors, *Renewable and Sustainable Energy Reviews* 40 (0) (2014) 659 – 672.
- [122] P. M. Costa, M. A. Matos, J. A. P. Lopes, Regulation of microgeneration and microgrids, *Energy Policy* 36 (10) (2008) 3893 – 3904.
- [123] A. Siddiqui, C. Marnay, Distributed generation investment by a microgrid under uncertainty, *Energy* 33 (12) (2008) 1729–1737.
- [124] A. F. Taha, N. A. Hachem, J. H. Panchal, A quasi-feed-in-tariff policy formulation in micro-grids: A bi-level multi-period approach, *Energy Policy* 71.
- [125] T. Bialek, Borrego springs microgrid demonstration project, Prepared for: California Energy Commission CEC?500?2014?067.
- [126] M. Hyams, et al., Microgrids: an assessment of the value, opportunities and barriers to deployment in new york state (2011).
- [127] M. J. Reno, C. W. Hansen, J. S. Stein, Global horizontal irradiance clear sky models: implementation and analysis, SAND2012-2389, Sandia National Laboratories, Albuquerque, NM.
- [128] C.-J. Lu, T.-S. Lee, C.-C. Chiu, Financial time series forecasting using independent component analysis and support vector regression, *Decision Support Systems* 47 (2) (2009) 115 – 125.
- [129] S. Fan, L. Chen, Short-term load forecasting based on an adaptive hybrid method, *IEEE Transactions on Power Systems* 21 (1) (2006) 392–401.
- [130] C.-L. Huang, C.-Y. Tsai, A hybrid softm-svr with a filter-based feature selection for stock market forecasting, *Expert Systems with Applications* 36 (2, Part 1) (2009) 1529 – 1539.
- [131] C. Granger, P. Newbold, Spurious regressions in econometrics, *Journal of Econometrics* 2 (2) (1974) 111 – 120.
- [132] C. W. J. Granger, Some properties of time series data and their use in econometric model specification, *Journal of Econometrics* 16 (1) (1981) 121 – 130.
- [133] C. W. J. Granger, Developments in the study of cointegrated economic variables, *Oxford Bulletin of Economics and Statistics* 48 (3) (1986) 213–228.
- [134] R. F. Engle, C. W. J. Granger, Co-integration and error correction: representation, estimation, and testing, *Econometrica: journal of the Econometric Society* (1987) 251–276.

- [135] V. A. Muscatelli, S. Hurn, Cointegration and dynamic time series models, *Journal of Economic Surveys* 6 (1) (1992) 1–43.
- [136] L. Bird, M. Milligan, D. Lew, Integrating variable renewable energy: Challenges and solutions, Tech. rep., Technical Report, National Renewable Energy Laboratory, September, accessible at <http://www.nrel.gov/docs/fy13osti/60451.pdf> (2013).
- [137] A. Mills, Integrating solar pv in utility system operations, Tech. rep., ANL/DIS-13/18 Argonne National Laboratory (2013).
- [138] Q. Wang, C. Zhang, Y. Ding, G. Xydis, J. Wang, J. stergaard, Review of real-time electricity markets for integrating distributed energy resources and demand response, *Applied Energy* 138 (0) (2015) 695 – 706.
- [139] R. Orans, A. Olson, J. Moore, J. Hargreaves, R. Jones, G. Kwok, F. Kahrl, C. Woo, Energy imbalance market benefits in the west: A case study of pacificorp and {CAISO}, *The Electricity Journal* 26 (5) (2013) 26 – 36.
- [140] J. King, B. Kirby, M. Milligan, S. Beuning, Flexibility reserve reductions from an energy imbalance market with high levels of wind energy in the western interconnection, Tech. rep., National Renewable Energy Laboratory (NREL), Golden, CO. (2011).
- [141] M. Milligan, K. Clark, J. King, B. Kirby, T. Guo, G. Liu, Examination of potential benefits of an energy imbalance market in the western interconnection, Tech. Rep. TP-5500-57115, NREL (2013).
- [142] R. A. van der Veen, R. A. Hakvoort, Balance responsibility and imbalance settlement in northern european evaluation, in: *Energy Market, 2009. EEM 2009. 6th International Conference on the European*, IEEE, 2009, pp. 1–6.
- [143] E. Bitar, R. Rajagopal, P. Khargonekar, K. Poolla, P. Varaiya, Bringing wind energy to market, *Power Systems, IEEE Transactions on* 27 (3) (2012) 1225–1235.
- [144] B. Hodge, A. Florita, J. Sharp, M. Margulis, D. Macreavy, The value of improved short-term wind power forecasting, Tech. Rep. TP-5D)-63175, NREL (2015).
- [145] A. Cifor, P. Denholm, E. Ela, B.-M. Hodge, A. Reed, The policy and institutional challenges of grid integration of renewable energy in the western united states, *Utilities Policy* (0) (2014) –.
- [146] H. Holttinen, M. Milligan, E. Ela, N. Menemenlis, J. Dobschinski, B. Rawn, R. Bessa, D. Flynn, E. Gomez Lazaro, N. Detlefsen, Methodologies to determine operating reserves due to increased wind power, in: *2013 IEEE Power and Energy Society General Meeting (PES)*, 2013, pp. 1–10.

- [147] B. Mauch, J. Apt, P. M. S. Carvalho, P. Jaramillo, What day-ahead reserves are needed in electric grids with high levels of wind power?, *Environmental Research Letters* 8 (3) (2013) 034013.