

Demand-Side Participation & Baseline Load Analysis in Electricity Markets

by

Nima Harsamizadeh Tehrani

B.Sc., Electrical Engineering, Isfahan University of Technology, 2010

Ph.D., Electrical Engineering, Nanyang Technological University, 2015

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of

MASTER OF APPLIED SCIENCE

in the Department of Mechanical Engineering

© Nima H. Tehrani, 2016
University of Victoria

All rights reserved. This thesis may not be reproduced in whole or in part, by photocopy
or other means, without the permission of the author.

Supervisory Committee

Demand-Side Participation & Baseline Load Analysis in Electricity Markets

by

Nima Harsamizadeh Tehrani

B.Sc., Electrical Engineering, Isfahan University of Technology, 2010

Ph.D., Electrical Engineering, Nanyang Technological University, 2015

Supervisory Committee

Dr. Curran Crawford, (Department of Mechanical Engineering)

Supervisor

Dr. Yang Shi, (Department of Mechanical Engineering)

Departmental Member

Abstract

Supervisory Committee

Dr. Curran Crawford, (Department of Mechanical Engineering)

Supervisor

Dr. Yang Shi, (Department of Mechanical Engineering)

Departmental Member

Demand participation is a basic ingredient of the next generation of power exchanges in electricity markets. A key challenge in implementing demand response stems from establishing reliable market frameworks so that purchasers can estimate the demand correctly, buy as economically as possible and have the means of hedging the risk of lack of supply. System operators also need ways of estimating responsive load behaviour to reliably operate the grid. In this context, two aspects of demand response are addressed in this study: scheduling and baseline estimation. The thesis presents a market clearing algorithm including demand side reserves in a two-stage stochastic optimization framework to account for wind power production uncertainty. The results confirm that enabling the load to provide reserve can potentially benefit consumers by reducing electricity price, while facilitating a higher share of renewable energy sources in the power system. Two novel methods, Bayesian Linear regression and Kernel adaptive filtering, are proposed for baseline load forecasting in the second part of the study. The former method provides an integrated solution for prediction with full accounting for uncertainty while the latter provides an online sequential learning algorithm that is useful for short term forecasting.

Table of Contents

Supervisory Committee	ii
Abstract	iii
Table of Contents	iv
List of Tables	vi
List of Figures	vii
List of Abbreviations and Symbols.....	ix
Acknowledgments.....	xiii
Chapter 1: Introduction.....	1
1. Types of DR programs.....	6
1.1. Non-dispatchable Demand Response Programs (Market-based DRP).....	6
1.2. Dispatchable Demand Response Programs	9
2. Motivation.....	9
2.1. Main Contributions	9
2.2 Thesis Structure	10
Chapter 2: Demand Response Programs.....	11
1. DR programs in Europe	11
2. DR programs in U.S.....	12
Chapter 3: Stochastic Unit Commitment with Demand Response Scheduling	19
1. Solution Method.....	20
2. Problem Formulation	21
2.1. Objective Function.....	22
2.2. Constraints	23
3. Computer Simulation.....	26
3.1. Three-bus System Case Study	26
3.2. IEEE-RTS Case Study	30
4. Summary	35
Chapter 4: Baseline Load Forecasting Using Bayesian Approach	36
1. Data-driven Models and Uncertainty Analysis.....	38
2. Linear Regression Review	39
3. Bayesian Linear Regression.....	40
4. Data Collection and Model Implementation.....	43
4.1. Case study I: Vancouver, British Columbia, Canada	43
4.2. Case study II: Austin, Texas, USA	45
5. Simulation Results and Discussion.....	46
6. Case Study III	55
6.1. Data Cleaning and Filtering.....	56
6.2. Output Process Rate Data Modeling.....	58
6.3. Process Rate Forecast (1-hr ahead).....	61
7. Summary	65
Chapter 5: On-line Baseline Load Learning Using Kernel Adaptive Filtering	67
1. What is Learning? Why Online Learning?	67
2. Linear Adaptive Filters	68
3. Least-Mean-Square (LMS) Algorithm	69
4. Recursive Least-Square (RLS) Algorithm.....	70
5. Nonlinear Adaptive Filters.....	71

5.1. The Kernel Trick.....	71
6. Kernel Adaptive Filtering (KAF).....	73
7. Kernel Least Mean Square Algorithm	74
8. Computer Simulation	76
9. Summary	80
Chapter 6: Conclusions and Future Work.....	82
1. Summary & Discussion	82
2. Future possibilities for current research work	83
Bibliography	86
Appendix A: Customer Baseline Load (CBL) Calculation in NYISO	94

List of Tables

Table 1. Wind power scenarios.....	27
Table 2. Amount of spilled wind power production [MW].....	28
Table 3. Scheduling results [MW] – Case 1	29
Table 4. Scheduling results [MW] – Case 2	29
Table 5. Distribution of the total system demand	31
Table 6. Amount of spilled wind power production.....	34
Table 7. Number of committed units at each hour.....	34
Table 8. Impact of DR level on the system (LA reserve cost is 5 \$/MWh).....	35
Table 9. Performance metrics for BC area aggregate baseline	54
Table 10. Performance metrics for Pecan St. Household baseline	54
Table 11. RMSE for different testing data sizes for LMS and KLMS	78
Table 12. KLMS comparison for different value of step size and kernel parameter (BC Load).....	79
Table 13. KLMS comparison for different value of step size and kernel parameter (Pecan St).....	80

List of Figures

Figure 1. Structure of the electricity market (E: Electrical energy; AS: Ancillary services)	1
Figure 2. . Load aggregator interactions in electricity market	3
Figure 3. Typical load model capable of demand response	5
Figure 4. Hierarchy of DR programs	6
Figure 5. A two-settlement electricity market [29]	8
Figure 6. Scenario tree for the two-stage problem	21
Figure 7. One-line diagram of the three-bus system	27
Figure 8. Total expected cost of the system variations	28
Figure 9. LMP variations for different LA reserve costs	30
Figure 10. One-line diagram of the IEEE-RTS	30
Figure 11. Daily load profile of the IEEE-RTS	31
Figure 12. Simulated ARMA(1,1) process, 20 scenarios, each with 24 observations, dashed line represents the mean wind power production	32
Figure 13. Total expected cost of the system variations	33
Figure 14. Marginal cost variations for different demand side reserve offer costs	34
Figure 15. Load and temperature variation in 2012	43
Figure 16. Load and temperature variation during January 2012, Vancouver BC	44
Figure 17. Load and temperature variation during January 2015, Austin TX	46
Figure 18. Histograms of BLR coefficients for BC area baseline for model 1 . Vertical line is the ordinary least square results	48
Figure 19. Histograms of BLR coefficients for Pecan St. household baseline for model M1 . Vertical line is the ordinary least square results	48
Figure 20. Approximate pdf of the variables of three models for BC area baseline	49
Figure 21. Approximate pdf of the variables of three models for Pecan St. household baseline	49
Figure 22. Trend plots for validation results of M1 , M2 and M3 for the first week of January 2013, 2014 and 2015 respectively for BC area baseline	50
Figure 23. Trend plots for validation results of M1 , M2 and M3 for Pecan St. Household	51
Figure 24. Histogram for validation results of M1 on Jan. 1 st 2013 for each hour of the day. Actual load is shown by red line	52
Figure 25. Histogram for validation results of M2 on Jan. 1 st 2014 for each hour of the day. Actual load shown by red line	53
Figure 26. Asset rate vs. asset electrical demand	57
Figure 27. Output process rate at Shire Oaks May-Aug 2015	58
Figure 28. A closer look at the output process rate at Shire Oaks 1-14 May 2015	59
Figure 29. Filtered output process rate	59
Figure 30. Average daily PR and a Fourier fit with one term	60
Figure 31. Average daily PR and a Fourier fit with two terms	60
Figure 32. Average daily PR and a Fourier fit with three terms	61
Figure 33. Fourier series curve fitting	61
Figure 34. Day-ahead output process rate forecast, red line represents the actual value	63
Figure 35. Current process rate calculation vs. Regression (update every 5 min)	64
Figure 36. Current process rate calculation vs. Regression (1-hr commitment)	64

Figure 37. Residual probability density function for regression and persistent model	65
Figure 38. Basic structure of a linear adaptive filter [79]	68
Figure 39. Basic structure of a nonlinear adaptive filter [79]	71
Figure 40. LMS and KLMS learning curves	77
Figure 41. Effect of regularization parameter on RMSE mean and standard deviation ...	78
Figure 42. Prediction vs. actual time series data for BC aggregate load 2012	79
Figure 43. Prediction vs. actual time series data for an aggregate household in Pecan St.	80

List of Abbreviations and Symbols

Acronyms

ARMA	Autoregressive moving average
BPA	Bonneville Power Administration
CAISO	California Independent System Operator
DISCO	Distribution Company
DR	Demand Response
DOE	Department of Energy
ERCOT	Electric Reliability Council of Texas, Inc.
FERC	Federal Energy Regulatory Commission
GENCo	Generation Company
ISO	Independent System Operator
KAF	Kernel Adaptive Filtering
LA	Load Aggregator
LMP	Locational marginal price
MISO	Midwest Independent System Operator
NYISO	New York Independent System Operator
PJM	Pennsylvania New Jersey Maryland Interconnection LLC
RTS	Reliability Test System
SMIP	Stochastic Mixed Integer Programming
SVM	Support Vector Machine
TRANSCO	Transmission Company
TSO	Transmission System Operator

Symbols

The nomenclature used in this thesis is listed below for clarity.

Indices

i	Generating units $\{1, 2, \dots, N_G\}$, N_G is the total number of generation units
j	Load aggregators $\{1, 2, \dots, N_L\}$, N_L is the total number of load units
k	Load Aggregators' energy blocks

m	Generating units' energy blocks
n, r	Indices of buses
t	Time periods $\{1, 2, \dots, N_T\}$, N_T is the total number of time periods
ω	Wind power scenarios $\{1, 2, \dots, N_\Omega\}$, N_Ω is the total number of scenarios

Constants

C_{it}^{RU}	Generation side up-spinning reserve cost (\$/MWh)
C_{it}^{RD}	Generation side down-spinning reserve cost (\$/MWh)
C_{it}^{NS}	Generation side non-spinning reserve cost (\$/MWh)
C_{jt}^{RU}	Demand side up-spinning reserve cost (\$/MWh)
C_{jt}^{RD}	Demand side down-spinning reserve cost (\$/MWh)
C_t^{WS}	Cost of wind power spillage (\$/MWh)
d_t	Time period t duration (h)
DT_i	Minimum down time (h)
$\lambda_{it}^G(m)$	Marginal cost of the m -th block of energy (\$/MWh)
λ_{jt}^L	Utility of LA (\$/MWh)
λ_{it}^{SU}	Start-up cost in period t (\$)
π_ω	Probability of wind power scenario ω
RD_i	Ramp-down limit (MW/h)
RU_i	Ramp-up limit (MW/h)
SU_i	Start-up ramp limit (MW/h)
SD_i	Shut-down ramp limit (MW/h)
UT_i	Minimum up time (h)
V_{jt}^{LOL}	Value of load shed for LA (\$/MWh)
X_i^{on}	On time at the beginning of scheduling horizon (h)
X_i^{off}	Off time at the beginning of scheduling horizon (h)
$X(n, r)$	Reactance of line (n, r) (per unit)

Variables

$C_{it\omega}^A$	Adjustment cost incurred at scenario ω (\$)
C_{it}^{SU}	Start-up cost (\$).
$C_{it\omega}^{SU}$	Actual start-up cost incurred at scenario ω (\$)
$\delta_{nt\omega}$	Voltage angle (rad)
$f_{t\omega}(n, r)$	Power flow through line (n, r) limited to $f^{max}(n, r)$ (MW)
$L_{jt\omega}^C$	LA power consumption (MW)
L_{jt}^S	Power scheduled for LA bounded by $L_{jt}^{S,min}$ and $L_{jt}^{S,max}$ (MW)
$L_{jt\omega}^{shed}$	Load shedding imposed on LA (MW)
$P_{it}^G(m)$	Scheduled power output from the m -th block of energy limited to $P_{it}^{G,max}(m)$
$P_{it\omega}^G$	Actual power output at scenario ω bounded by P_i^{min} and P_i^{max} (MW)
$P_{t\omega}^{WP}$	Realization of wind power generation in period t and scenario ω (MW)
P_{it}^S	Scheduled power output (MW)
$q_{jt\omega}$	Reserve quantity of LA j in period t and scenario ω (MW)
R_{it}^U	Scheduled up-spinning reserve limited to $R_{it}^{U,max}$ (MW)
R_{it}^D	Scheduled down-spinning reserve limited to $R_{it}^{D,max}$ (MW)
R_{it}^{NS}	Scheduled non-spinning reserve limited to $R_{it}^{NS,max}$ (MW)
R_{jt}^U	Spinning reserve up scheduled for LA limited to $R_{jt}^{U,max}$ (MW)
R_{jt}^D	Spinning reserve down scheduled for LA limited to $R_{jt}^{D,max}$ (MW)
$r_{it\omega}^U$	Generation side deployed up-spinning reserve (MW)
$r_{it\omega}^D$	Generation side deployed down-spinning reserve (MW)
$r_{it\omega}^{NS}$	Generation side deployed non-spinning reserve (MW)
$r_{jt\omega}^U$	Demand side deployed up-spinning reserve (MW)
$r_{jt\omega}^D$	Demand side deployed down-spinning reserve (MW)
$r_{it\omega}^G(m)$	Reserve deployed from the m -th block of energy ω (MW)
$S_{t\omega}$	Wind power spillage (MW)
u_{it}	Binary variable equal to 1 if unit i is scheduled to be committed in period t

$v_{it\omega}$ Binary variable equal to 1 if unit i is online in period t and scenario ω

$x_{jt\omega}$ Binary variable equal to 1 if LA j is online in period t and scenario ω

Acknowledgments

I would like to deeply thank my supervisor, Dr. Curran Crawford, for his guidance during my research and study at the University of Victoria. His spectacular vision, perpetual energy, and enthusiasm for research have motivated all his advisees, including me. Mostly, I thank him for the freedom in thinking and research directions that he provides for his students, which gave me the liberty to explore very interesting subjects within the principal theme of my dissertation. His open and positive approach served my curiosity, and he was my advocate and mentor throughout my studies. He deserves my utmost gratitude and I hope he remains a guiding presence to steer my way in the future.

I am grateful to my dissertation committee members, Dr. Yang Shi and Dr. Lin Cai. Without their invaluable comments, this thesis could not have been more accurate and more clearly expressed.

I would also like to thank all amazing researchers and staff at the Institute for Integrated Energy Systems at the University of Victoria (IESVic) and Mechanical Engineering department, who provided valuable insight and support for my ideas.

Further, I would like to thank CIMTAN and NSERC for providing funding throughout my studies, without which it would not have been possible.

Finally, I would like to thank my friends and family and above all my wife, Yasamin, for always encouraging my studies while keeping me grounded. This thesis is dedicated to her.

Chapter 1: Introduction

The electricity market can be divided into two different types: the *spot market*, where the electrical energy and ancillary services are traded for immediate physical delivery, and the *futures market*, where the delivery is later and normally does not involve physical delivery. The *futures market* is normally used for risk hedging. Ancillary services are functions separated from the electrical energy market, which is used to support reliability and power quality of the power system. One example is the power system reserves. Coexisting with the electricity market, there are also bilateral contracts that the market agents are free to trade. These contracts are normally used to guarantee a certain amount of electrical energy for the demand-side, or to guarantee a certain profit for the supply-side, or used as a risk hedging mechanism. Figure 1 depicts the general time frame of an electricity market.

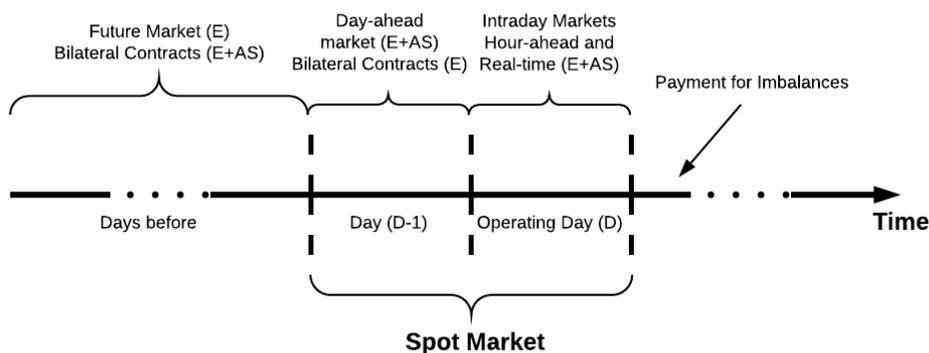


Figure 1. Structure of the electricity market (E: Electrical energy; AS: Ancillary services)

In a typical US market, Independent System Operator (ISO) performs functions of system optimization and market operation control in a competitive power pool. ISO is independent and does not own generation nor transmission or distribution. It makes sure that market information is facilitated to all parties on a non-discriminatory basis.

Today's electric grid is evolving into a "Smart Grid" where computing and communication technology allow assets at all levels of the system to be monitored and controlled. Furthermore, deregulation of the power system has led to competition among

generation companies (GENCOs), transmission owners (TRANSCO) and distribution companies (DISCO). These entities are developing innovative smart grid strategies to improve their reliability and profit. On the other hand, renewable energy resources especially wind and solar power is expected to serve increasing shares of energy requirements in the near future as production costs continue to drop. In fact, renewable energy accounted for almost two-thirds of new U.S. electrical generation put into service during 2015 according to the Federal Energy Regulatory Commission (FERC) [1]. These intermittent resources not only are intrinsically incapable of providing load following but also contribute to increasing imbalances due to forecast uncertainty.

In the Smart Grid along with the increasing share of intermittent sources in the supply chain, end-users are expected to play an active role in grid management via Demand Response (DR). Demand Response (DR) refers to end-use customers reducing their use of electricity in response to power grid needs, economic signals from a competitive wholesale market or special retail rates [2]. DR is defined by the US Department of Energy (DOE) as “a tariff or program established to motivate changes in electric use by end-use customers in response to changes in the price of electricity over time, or to give incentive payments designed to induce lower electricity use at times of high market prices or when grid reliability is jeopardized” [3].

DR potentially reduce energy generation in peak times (reduce the cost of energy and possibly emissions depends on the electricity mix). Hence, investment in peaking units can be avoided. Furthermore, Power system frequency quite often deviates from the nominal value due to supply-demand imbalances. Reserve power is required to deal with this problem but stand-by power reserve supplied by generating units is expensive. To resolve this issue, demand side can be used as an ancillary service provider for reserve and regulation. The benefits are more steady frequency, reduce capacity reserve requirements or increase the reliability of supply, low-cost reserve for daily operation/critical situations and increase penetration level of intermittent renewable resources.

In transmission and distribution, DR can be used to relieve congestion, manage contingencies or avoid outages, reduce overall losses and facilitate technical operation (i.e. keep frequency and voltage levels, balance active and reactive power, control power factor,

phase imbalance correction, etc.). Hence, utilities are able to defer investment in network reinforcement and increase long-term network reliability. From a demand perspective, DR makes consumers more aware of their cost and consumption. Therefore, consumers have options to maximize their utility by trading-off price with flexibility, thereby reducing electricity bills or receiving payments. Price volatility reduction and increased demand elasticity are other DR advantages from the retail viewpoint. [4].

Considering recent developments in real time telemetry, DR has been shifted from load curtailment to *demand dispatch*. In order to contribute services that support the operation of the grid, demand dispatch needs to be used actively at all times. Real-time DR will improve security and reduce reserves by dynamically adapting loads to available generation. The results from demonstration projects like [5], [6] showed that there are no technical barriers in the way of large-scale integration of automated technologies for DR.

In order to make DR compatible with the electricity market framework, an aggregating entity in the command and contracting architecture should be included. The presence of the aggregator provides the ISO with considerable flexibility in the scheduling of units. Thus, the start-up of cycling and peaking units may be delayed or avoided; the availability of reserves is improved and during off-peak and the need for the reserve is reduced [7]. A load aggregator (LA) is capable of bundling loads to form a single controllable power resource. It is responsible for ensuring requested power response is provided and all mechanical and process constraints are respected to avoid equipment damage and fatigue. LA interfaces with the ISO on supply-side and the DISCOs or loads directly on demand-side as shown in Figure 2. Moreover, the LA can manage the contract size to maintain industry standard reliability over the course of each hour with any time resolution depends on the required service [8].

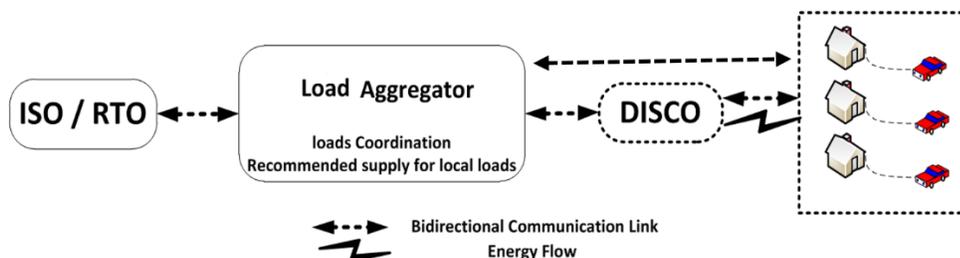


Figure 2. Load aggregator interactions in electricity market

Load aggregation is intrinsically more reliable than the decentralized control where the reliability is entirely dependent on the uncontrolled behavior of the end-use customers. The main goal of an aggregator is to maximize the payment from grid operator by providing DR products for capacity, energy or ancillary services. Ancillary services offered by aggregator include regulation, contingency reserve, renewable firming, peak demand management, fast DR and voltage management (Volt/VAR optimization). Regulation is a variable amount of generation power under automatic control which is independent of economic cost signals and is dispatchable within five minutes. Reserves are additional generation capacity above the expected load. Contingency (primary) reserve, including synchronized and off-line reserve is obtainable within 10 minutes, while the secondary reserve is obtainable within 10-30 minutes following the ISO request [12].

Renewable firming is a service to compensate for the variable output from a renewable power source and maintain a committed power level for a period of time. Peak shaving reduces peak demand to avoid the installation of capacity to supply the peaks of a highly variable load that is only called on infrequently. Fast DR is designed to reduce electricity demand in near real time in response to grid changes when generation is not sufficient to meet peak load, or to handle sudden drops in the wind or solar generation. Voltage management controls voltage magnitude and phase angle at key locations through the distribution system. This minimizes the impact of intermittency on the utility, reduces voltage violations, reduces tap changer operations, minimize reactive power flow and reduces distribution line loss.

Loads participating in a DR program must be capable of deferring their operation with minimal impact on customer comfort. Energy-constrained storage like batteries exhibit significant promise, especially with increasing share of Plug-in Electric Vehicles (PEVs). Moreover, thermostatically controlled loads (TCLs) including refrigerators, heat pumps, air conditioners, hot water heaters, chilled water loops and cold storage facilities are perfect candidates as they are capable of storing thermal energy. A general overview of a load that can work in a DR environment is shown in Figure 3. For each controllable device, there might be an associated input/output storage. The important thing is the capability of modulating device power consumption such that storage constraints are not violated.

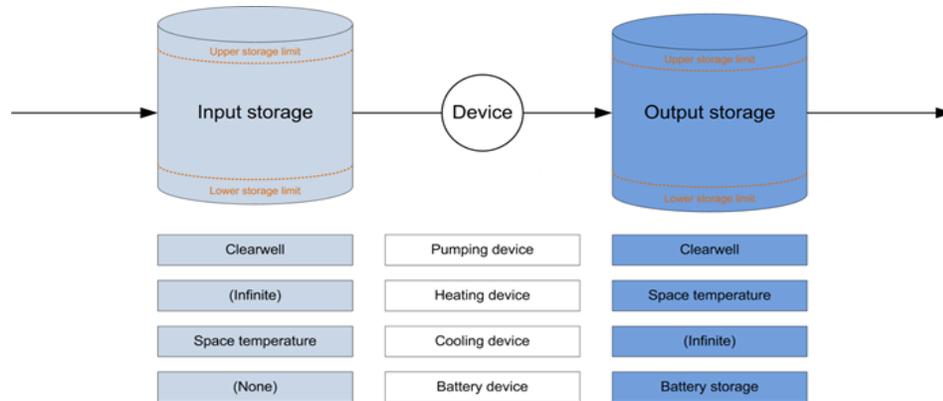


Figure 3. Typical load model capable of demand response

For ancillary services, the aggregator should provide as much power capacity as possible whenever the service is requested by the grid. Revenue of the aggregator could be maximized by performing forecasting and optimization so that each load remains idle whenever the ancillary service price is high. Therefore, the actions of the loads can be optimized by the aggregator to maximize the revenue and eliminate load spikes while meeting load requirements.

Based on the market structure mentioned earlier, LA participates in the competitive electricity pool consists of two successive markets: a day-ahead market and a real-time balancing market. Let us assume LA bids in the day-ahead market for energy and regulation. Deviations from scheduled nominal set point are within the regulation limits. In real-time, aggregator monitors and controls the operation of each asset. Regulation services are shifting to loads where the nominal power level is set at a negative value to provide power. Regulation power consists of down/up-regulation. Down regulation representing an increase in consumption. To provide up regulation, LA decreases the consumption power of loads.

Regulation services are contracted by the ISO/RTO on an hourly basis as needed throughout the day. The actual dispatch of the contracted regulation is normally a few minutes in one direction (up or down) at a time. Payment is typically made for the contracted amount of capacity for a given hour (\$/MW per hour) as well as energy delivered. This price is set hourly through the market clearing process. The aggregator is assumed to be price-taker. Although the large-scale integration of loads in future will affect market prices but eventually it leads to stable electricity prices at a high penetration level.

1. Types of DR programs

In terms of DR control paradigms, following approaches are currently being practiced:

- Direct Load Control: where aggregator controls load. It is versatile, fast and better for power system and requires minimal attention from load perspective since a centralized controller broadcast control signals.
- Indirect Load Control: where load operators (e.g. end point consumers) controls load in response to a control signal for example price. It is less reliable and implausible for fast services. It might induce volatility but consumers can choose degree of involvement. Since it is a self-organizing price-based mechanism, it is effectively a distributed control scheme.

Based on this, DR programs generally can be categorized into two main groups as shown in Figure 4.

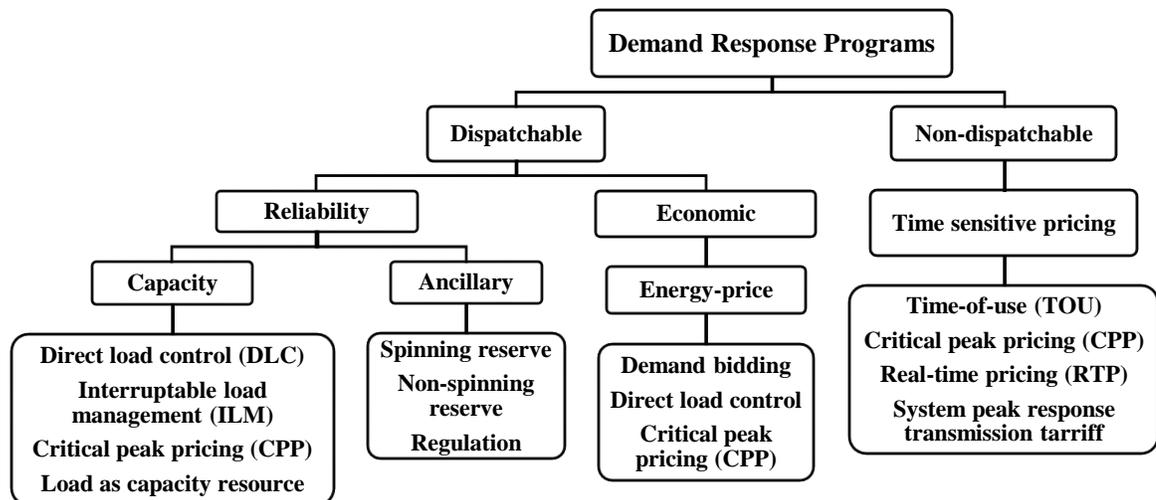


Figure 4. Hierarchy of DR programs

1.1. Non-dispatchable Demand Response Programs (Market-based DRP)

Typically involve changes in customer load in response to prices, whether fixed or dynamic. TOU and CPP are two forms of the pre-defined fixed market-based DR that have been employed by utilities for peak demand management. Recently with advances in real-time telemetry, RTP is getting more attention both in academia and industry. The results from the Olympic Peninsula demonstration project [28] and American Electrical Power gridSMART project [29] showed that customers are willing and able to respond to real-time energy pricing information. In [28], DR saved consumers money on power and

reduced peak load by approximately %15 over the course of one year. Overall, reduction both in household energy consumption and wholesale energy cost is achievable via real-time pricing. Furthermore, construction of new generation, transmission and distribution system can be avoided; with the saving passed along to end-use customers. The key factor in RTP is price elasticity of responsive electrical demand (η) defined as the fractional change in demand (q) to a given fractional change in price (p) [30]:

$$\eta = \frac{\Delta q/q}{\Delta p/p} = \frac{p \Delta q}{q \Delta p}$$

Consider the supply and demand curve in Fig.3. In a typical power pool, the market clearing price (MCP) is established as the intersection of the supply curve (constructed from aggregated supply bids) and the demand curve (constructed from stacked supply bids). As it is shown in Figure 5 (top figure) the cleared market price and quantity are transmitted to price responsive devices in an uncongested situation.

During congestion, the cleared price will increase and the cleared quantity equals the feeder capacity. As a result, the bidding equipment is dynamically encouraged to curtail operations and mitigate congestion limits. The congestion surplus is shown in Figure 5(bottom figure) is rebated back to the consumers who were flexible to price changes; thus, removing the unfair burden of charging price-responsive customers more. Though RTP programs are very appealing due to low communication requirements and private settings, the problem of defining appropriate price signals in order to have efficient and secure grid operation is still a subject of heated debate and research.

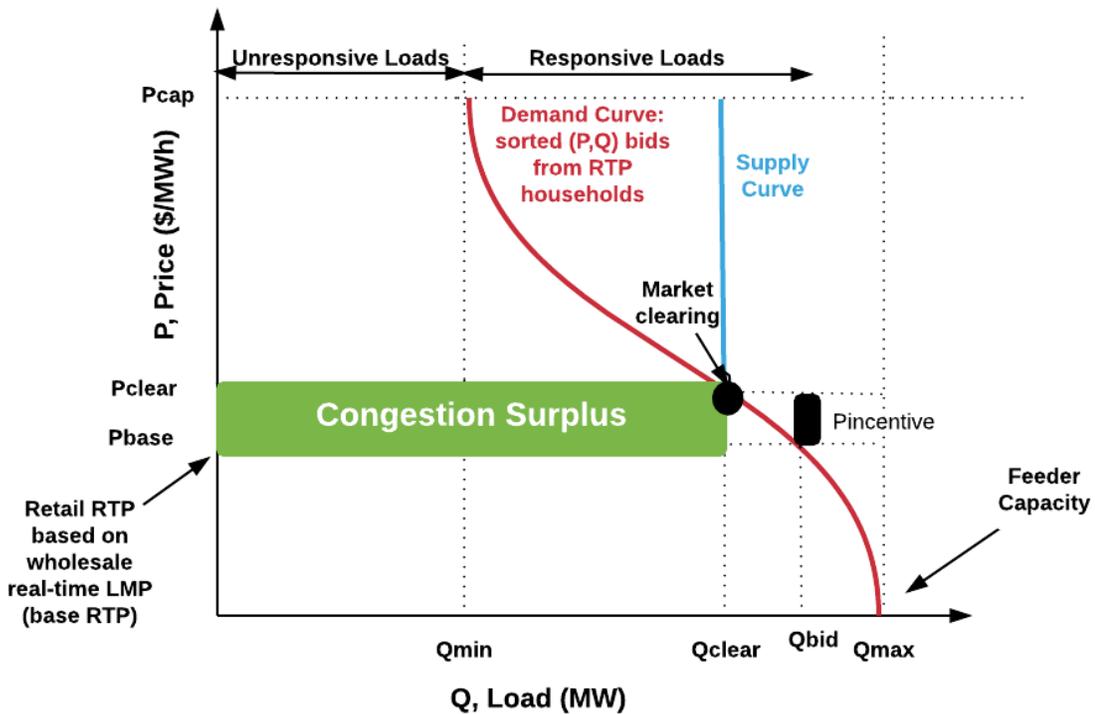
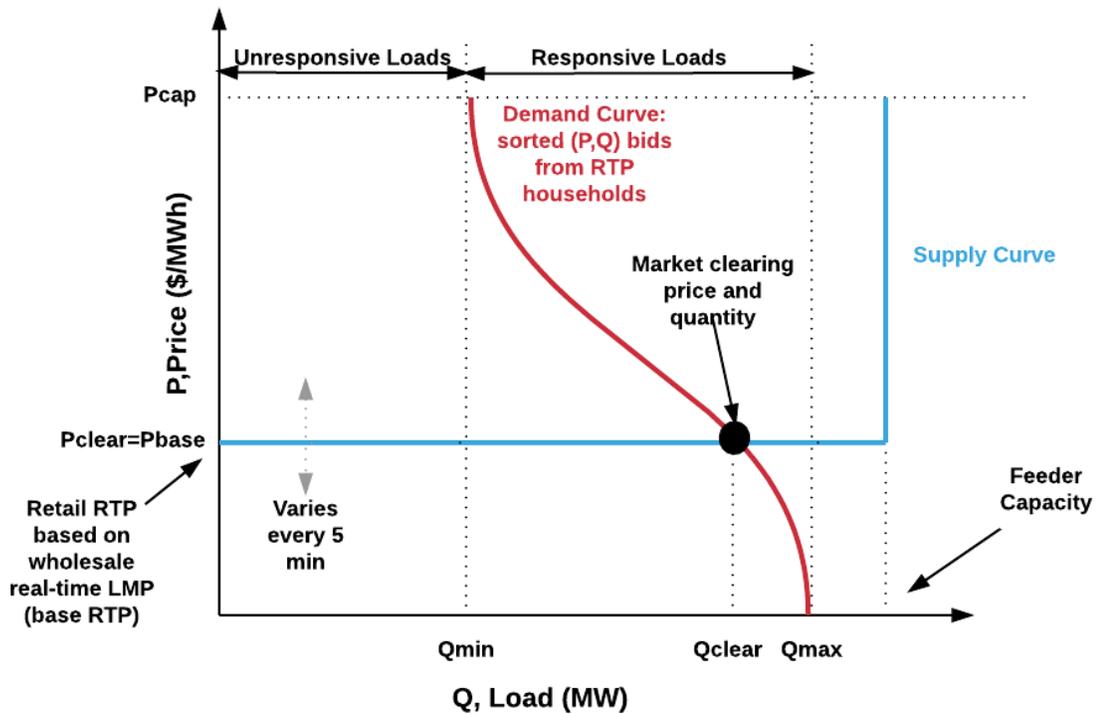


Figure 5. A two-settlement electricity market [29]

1.2. Dispatchable Demand Response Programs

Typically involve customer commitments to modify loads within prior agreed-upon constraints when needed by ISO (either for capacity, energy or ancillary services) and treated much like generation as they are achievable, reliable and capable of responding within ISO/RTO time guidelines. Interruptible Load Management (ILM) and Direct Load Control (DLC) programs have been practiced since the 1970s for peak load management to manage emergency situations. Controlling large industrial/commercial units through ILM and residential TCLs through DLC have been addressed before [4].

Load participation in the wholesale electricity market is another way to achieve DR and is getting more attention from utilities and ISOs. In demand/capacity bidding programs, load aggregator offers demand reductions via price/quantity bids into energy and capacity markets like DADR and EDR programs in NYISO mentioned earlier. If their bids are accepted they must provide demand reductions at specified times for a specific duration. In some cases, loads can also participate in ancillary services markets like NYISO's DSASP.

These programs ensure reliable grid operation but suffer from end-users' hesitation due to privacy concerns and high computational and communication requirements.

2. Motivation

2.1. Main Contributions

With a high penetration level of renewable generation, measures have to be considered to take into account the uncertain nature of these resources. DR is expected to play a major role to mitigate the major shortcoming of harvesting power from the wind: the variability challenge [31], [32]. The motivation for chapter 3 is to investigate the effect of enabling loads to participate in the electricity market clearing process in a stochastic framework.

Furthermore, regardless of the type of DR program employed (dispatchable vs. non-dispatchable), all require analysis to estimate the demand reduction. Baseline models are used for a variety of purposes including DR measurement and verification (M&V), improving DR program design, and operation and financial settlement for DR participants. Therefore, it is in the best interest of the utilities, LAs, and end-users to have as accurate a baseline estimation as possible. Measuring DR performance is of utmost concern for policy

makers and DR program designers, as mentioned in the FERC report that “development of standardized practices for quantifying demand reductions would greatly improve the ability of system operators to rely on demand response programs” and is “central to the issue of measurement is a determination of the customer baseline” [33]. This is a challenge due to limited communication, complex inputs such as weather and uncertain end-use behavior. Chapters 4 and 5 are proposing new online techniques that both utilities and aggregators can use for baseline load forecasting. Chapter 4 introduces the idea of using a recursive Bayesian linear regression approach. Chapter 5 continues the topic of on-line baseline load forecasting using the nonlinear filtering technique.

2.2 Thesis Structure

This thesis proceeds as follows: Chapter 2 provides an overview of existing demand response programs in Europe and U.S. Stochastic unit commitment with demand response scheduling is addressed in chapter 3. In chapter 4, a multiple linear regression model is used in a Bayesian framework to forecast load baseline. Bayesian method offers an integrated approach to inference with a full accounting for uncertainty. In Chapter 5, Kernel adaptive filters are introduced as a new method for short-term baseline analysis. Chapter 6 concludes the thesis, providing a summary of key findings and recommendations for future work.

Chapter 2: Demand Response Programs

Incorporating demand response in the power market has been an active research topic in recent years. A number of grid operation governance policies have been aimed at encouraging DR participation in the power markets. ISOs and utilities around the world have already realized the benefits of relying on DR for ancillary services provision. The success of enabling demand response capable customers to bid into capacity markets has led systems to open up their energy, capacity, and ancillary service markets to entities providing load control [9-13].

1. DR programs in Europe

DR programs in Europe is used to be mainly interruptible tariffs to promote participation by large industrial consumers and time-of-use tariffs for small consumers [14]. However, with the development of Smart Grid this situation is changing in Europe [15]. For example, the TSO of Norway (Statnett) is acquiring DR through market bidding, mainly focused on end-users and independent aggregators [16]. Denmark's TSO (Energinet.dk), in order to promote the participation of small loads in the regulation power market, published a proposal in 2011 which outlines participation in the regulation market and self-regulation [17].

TSOs in Germany call for a joint monthly tender to procure a fixed quantity (1500 MW as of now) of the interruptible load for measures to maintain network and system security. Interruptible loads are seen as large consumption units which are connected to the high and extra high voltage network. The Federal Network Agency defined new conditions for primary and secondary reserve providers in order to facilitate market participation for small generators, loads and storage [18], [19].

The interruptible loads are categorized into two categories: immediately interruptible loads (SOL) and fast turn off loads (SNL). Loads can select any of the following options to serve under:

- at least 15 minutes in each case at any time several times a day at any distance up to the duration of one hour per day at least four times a week

- at least four hours at a time at any time once every seven days
- at least eight hours at a time at any time once every 14 days

In Italy, the transmission system is mainly controlled by a single TSO [20]. TSO is assigned with the responsibility to procure the resources necessary to guarantee the balance of the power system and to release the intra-zonal congestions. These resources are procured through an Ancillary Service Market (MSD). The MSD is cleared through a pay as bid algorithm. TSO is the central counterparty which accepts bids/offers from market participants related to different reserve and balancing services. This market is further divided into two:

Ex-ante MSD: In this market, TSO accepts energy demand bids and supply offers in order to relieve residual congestions and to create reserve margins. There are 4 different scheduling sub-stages. These consecutive sub-stages are instances before starting operation hour that TSO updates reserve requirements for the system. In spite of having 4 different scheduling sub-stages, there is only a single session for bid/offer submission that starts at 12:55 p.m. of the day before the day of delivery and closes at 5.30 p.m. on the same day. The result of each of scheduling sub-stages is declared at the different point of times.

Balancing Market (MB): In this market actual balancing take place. The TSO selects bids/offers in respect of groups of hours of the same day on which the related balancing session takes place. For now, there are 5 balancing sessions. The first session of the MB takes into consideration the valid bid/offers that participants have submitted in the previous ex-ante MSD session. For the other sessions of the MB, all the settings for bid/offer submission open at 10.30 p.m. of the day before the day of delivery (and anyway not before the results of the previous session of the ex-ante MSD are made known) and close 1 hour and a half before the first hour which may be negotiated in each session. TSO accepts energy demand bids and supply offers in order to provide its service of secondary control and to balance energy injections and withdrawals into/from the grid in real time.

2. DR programs in U.S.

Important regulatory decisions in U.S. paved the ground for DR resources participation in wholesale markets and increase the revenue they can generate. As a result, the number

of DR programs integrated into the electricity markets are increasing across the country. In June 2011, DOE and the Federal Energy Regulatory Commission (FERC) jointly submitted the *Implementation Proposal for The National Action Plan on Demand Response* report to Congress. In Ref. [21] the authors proposed a framework for evaluating the cost-effectiveness of DR which was prepared for the national forum on the national action plan on DR. At the state level, at least 28 states require utilities to include demand-side resources in their resource planning for their future energy needs [22].

FERC order 719 (2008) requires ISOs to accept bids from DR resources in their ancillary service markets. It enables DR resources to compete in the market on a basis comparable to other resources [23]. Market structure is very much still evolving to fully unleash the power of DR as evidenced by the debate around FERC Order 745 (2011). In this order, FERC set the compensation for DR at the locational marginal price (LMP) for the place and time the DR is offered [24]. It tried to establish uniform rules for customer engagement in U.S. electricity markets. However, US Court of Appeal in 2014 invalidated Order 745 by ruling that FERC does not have jurisdiction to federally regulate demand response as a tool of the wholesale bulk power market. The challengers also argue that demand response was, in essence, a retail sale and thus not subject to FERC's jurisdiction. In October 2015, U.S. Supreme Court heard oral arguments over whether the FERC had jurisdiction to issue Order 745 and eventually approved the specific rules set in FERC Order 745 in January 2016 [25]. There is also this question that whether this Order effectively ensures double compensation for responsive loads providing DR. One possible approach could be implementing real-time retail pricing at the LMP to eliminate wholesale DR compensation. [26]

In the restructured market of ERCOT, loads directly compete with generators in day-ahead ancillary service market (*load acting as a resource*). Scheduled load resources receive capacity payment regardless of whether they are called. In January 2015, the ERCOT average demand response from load resources was around 1366 MW [9].

The NYISO has four demand response programs: 1) emergency demand response program (EDRP); 2) ICAP special case resources (SCR) program; 3) day ahead demand response program (DADRP); 4) demand side ancillary services program (DSASP). The EDRP and SCR programs are capacity programs in which load resources are curtailed in

energy shortage events in order to maintain a reliable system. The DADRP program allows load resources to bid into NYISO's day-ahead energy market in a method similar to generators. Finally, the DSASP program allows load resources to provide load-following and regulation services. Capacity programs are still NYISO's most popular product for demand response, followed by energy markets, and then ancillary services.

As of July 31, 2014, a total of 1210.7 MW of demand response was enrolled in the NYISO's EDRP and ICAP/SCR program. This corresponded to a 4.6% decrease from the MW enrolled in 2013 and represents 4.1% of the 2014 Summer Capability Period peak demand of 29,782 MW. During the analysis period of August 2013 through July 2014, there were no offers or schedules of DADRP resources. There are three demand side resources actively participating in the DSASP as providers of Operating Reserves. The resources represent 126.5 MW of capability and had an average performance of 154% during the analysis period of May 2014 through October 2014 [10].

Demand response is an integral part of PJM's markets for energy, day-ahead reserve scheduling, capacity, synchronized reserve and regulation. Like NYISO, PJM allows load entities to participate in their capacity, energy, and ancillary service markets. Also like NYISO, demand response participation is almost an order of magnitude greater in PJM's capacity market (8,683 MW) versus its energy market (1,727 MW).

In 2015, the total demand reduction assuming full DR compliance and economic reductions is estimated to be 10,432 MW. But unlike NYISO, the load is an active participant in PJM's ancillary service markets, particularly PJM's synchronized reserve market that provides load-following services. In the first quarter of 2011, demand response provided on average 84,551 MWh of synchronized reserve service in PJM, which translates to an average demand response capacity of 118 MW while in the first month of 2015, DR provided an average synchronized reserve and regulation capacity of 335 MW and 12 MW respectively [11], [12].

The MISO demand response participation is also similar to NYISO and PJM. It allows demand response resources to participate in its capacity, energy, and ancillary service markets. But where MISO differs is that it has 17 MW of the load from an aluminum smelter providing regulation service, making it the only ISO or RTO to have a load entity providing flexible operating reserve service on timescales shorter than 10 minutes. Another

example is the demand response reserves pilot project from ISO-NE for load resources [13].

CAISO is working towards the development of full-fledged Demand Response products [27]. As of December 2015, CAISO has introduced two DR programs: Proxy Demand Resource (PDR) and Reliability Demand Response Resource (RDRR). The prior introduction of these programs on CAISO platform, the development, and implementation of the Demand Response programs were the responsibilities of Investors Owned Utilities (IOUs). In fact, each of the IOUs has their own wide range of DR programs which are open to their customers. Altogether there are 3 IOUs in California: Pacific Gas and Electric (PG&E), San Diego Gas and Electric (SDG&E), and Southern California Edison (SCE).

The current time is seen as a transition phase where California ISO and utilities are moving away from Integrated Demand Side Management (IDSM) approach towards Integrated Demand Side Resources (IDSR) approach. The focus of the IDSM approach was in the development of utility led programs focused on energy efficiency, programs like critical peak pricing, Load Response, Local Generation etc. it focused towards more coherent and efficient optimization of operations and maintenance of these programs. IDSR approach on the other hand shifts focuses towards developments of collective actions to optimize demand response resources rather than utility driven tailored programs.

California Public Utilities Commission (CPUC) has set the deadline of 2018 to enable all DR resources (within CAISO and all utilities) to participate in CAISO programs. CAISO and utilities have been asked to develop guidelines for the enablement of this target. They are still working on defining the rules and regulation for aggregator participation. However, IOU programs have well-defined guidelines and rules for each of their DR programs.

PDR is a load or aggregation of loads that is capable of measurably and verifiably reducing their electric demand. It is treated just like a supply resource and, it can bid economically into following CAISO markets just as other supply sources do:

1. day-ahead energy market
2. 5-minute real-time energy markets,
3. Day-ahead and real-time non-spinning reserves markets.

A PDR must meet a minimum load curtailment defined for different markets. As of January 2016 following are the minimum load curtailments required:

1. 0.1 MW (100 kW) for Day-Ahead and Real-Time energy
2. 0.5 MW (500 kW) for Day-Ahead and Real-Time energy Non-Spinning Reserve

Each aggregation must meet the minimum load curtailment requirement on aggregation level to participate in the market.

RDRR is a wholesale product that enables emergency-responsive DR resources to integrate into the CAISO's economic day-ahead and real-time reliability market. RDRR is a load or aggregation of loads that is capable of measurably and verifiably reducing their electric demand. It relies on the same functionality and infrastructure designed for PDR and is modeled like a supply resource. Resources first offer economic energy in the day-ahead market, then provide the remaining uncommitted capacity as energy in real-time when required under a system or local emergency. An Individual and aggregated demand response resources are eligible to participate if the resource is configurable to offer day-ahead energy and respond to real-time reliability events. Program Specifications are as follows:

1. Minimum Load Curtailment: 500 KW
2. Real-time reliability service must reach full curtailment within 40 minutes
3. Minimum Run Time ≥ 1 hour
4. Maximum Run Time ≤ 4 hours
5. Can opt for voluntary discrete dispatch (all or nothing)
6. Must be available for up to 15 events and/or 48 hours per 6 month period

CAISO is working on a '*bifurcation policy for demand response programs*' which is expected to be implemented fully by 2018. Under this policy, all demand response resources have to be categorized in either of two categories: '*Load Modifying Resource*' and '*Supply Side Resource*'. The objective of this bifurcation is to separate '*event-based dispatchable*' supply side resources from load modifying resources.

Load Modifying Resource: Load-modifying resources (LMRs) refer to non-event demand response programs not integrated into the wholesale market. CAISO emphasizes that the primary purpose of LMRs is to avoid capacity costs (not avoid energy costs), to reduce peak generation and avoid the construction of additional capacity.

Supply Side Resource: Supply-side resources participate in event-based demand response programs and are integrated into the ISO wholesale market. CPUC wants increased participation in this category hence they bound the utilities to come up with pilot projects to competitively solicit supply-side DERs for demand response programs.

CPUC launch a pilot project 'Demand response Auction Mechanism' (DRAM) which allows DR providers, including third-party aggregators, to directly participate in CAISO's day-ahead energy market. The plan calls for utilities to procure resource adequacy (RA) from third-party demand-response providers (e.g., utilities, third party aggregators) on a monthly basis. Through this mechanism, they are creating a path for transitioning away from bilateral utility contracts toward a pay-as-bid auction-based method for securing supply-side resources. DR resources under DRAM will be marketed in CAISO's energy market. DR providers will have to register DERs with CAISO as a Proximity Demand Resource (PDR). The first DRAM mechanism was completed by late 2015. The second phase will seek demand response to meet traditional, system-wide RA needs in 2016.

Distributed Energy Resource Provider (DERP) Framework

CAISO is in process of introducing a new category of market players named DERP with the objective to increase participation in the wholesale market. The DERP will be defined as *'The owner/operator of one or more DERs that participates in ISO markets as an aggregated resource.'* DERP is aimed to alleviate different barriers for small DER's such as 500 kW minimum capacity requirement, strict telemetry and metering requirements directly with the ISO etc. In DERP though revenue-grade metering will be required to be deployed to all resources, however, DERP to ISO communication will be mediated through a scheduling coordinator. The DERP will either have to hire a third-party scheduling coordinator entity or fill the role itself. Broadly four stakeholders will perform following responsibilities:

- **ISO:** ISO will coordinate dispatch with the scheduling CAISO's DERP Framework coordinator. It also holds the authority to audit and test metering facilities, data handling, and processing procedures.
- **Scheduling Coordinator:** SC will be responsible for submitting aggregated settlement quality meter data (SQMD) from all underlying DERs directly to ISO. SCs will also be responsible for performing audits and tests to ensure compliance

with local regulatory requirements, disaggregating resource-level SQMD of a DERP's underlying resources, Scheduling, bidding, real-time telemetry, control signal disaggregation, SQMD submittal and settlement with participating DERs and other related activities.

- **DERP:** The DERP is required to operate and maintain DERs following applicable ISO tariffs. As a DERP will be a scheduling coordinator-metered entity (SCME), it will forward directly metered underlying DER data to the scheduling coordinator. It will also provide the ISO with basic, historical underlying DER information, including resource attributes and meter/telemetry data for settlement and operational purposes. DERPs are not prohibited from also being scheduling coordinators.
- **Resources:** All resources will be required to follow the local regulatory authority requirements. Furthermore, DERs must install revenue quality metering and employ direct meters to measure performance, rather than relying on a baseline methodology.

Chapter 3: Stochastic Unit Commitment with Demand Response Scheduling

Based on the operation of the markets previously mentioned in chapter 1, we represent the day-ahead scheduling and simultaneous energy and reserve market clearing problem that needs to be addressed by the ISO. The most related previous works to the study presented in this chapter are [8], [34], [35] as they address the market-driven power system operation with DR integration into simultaneous energy and reserve market clearing algorithms.

In [8], the market clearing problem is formulated considering random outages of generating units and transmission lines and highlights the benefits of customers' response to a DR program of the ISO. In [34], a DR program is proposed which helps to integrate wind power by reshaping the load of the system and provides a framework to procure load reduction from DR resources in the wholesale energy market. In the most recent study [35], the authors proposed a detailed DR model including load shifting, curtailment, and use of energy storage and on-site generation in the market clearing process but in a deterministic context. Other studies like [36] proposed a day-ahead market clearing model with DR in the hourly solution of security constrained unit commitment. In [37] hourly DR scheduling was proposed considering the ramping costs of generation.

The difference compared to previous studies is that the goal here is to investigate how enabling loads to provide reserve power affect unit commitment, system operation cost, and renewable penetration level. In this study, a stochastic model for operations planning with wind power generation is proposed. The proposed model is formulated as a two-stage stochastic mixed-integer programming (SMIP) problem and it would schedule commitment states of generating units and their scheduled energy along with the reserve provided by generating units and LAs over the scheduling horizon.

The rest of this chapter is organized as follows. First, the structure of the problem is introduced. An SMIP is proposed for the bidding strategies of resources considering DR. The numerical studies conducted on the three-bus system and IEEE-RTS to highlight benefits of the DR program on the power system.

1. Solution Method

In a power pool, the ISO receives energy offers and bids from LAs to determine the power production, the consumption level, and the price. The aim is to maximize the net social welfare in a process known as market clearing [38]. In many markets, the market clearing procedure is a day-ahead procedure, since the ISO needs to verify in advance that the schedule is feasible and the physical constraints of the grid are not violated. The balancing (or real-time) market operates a very short time before the delivery in order to keep the balance between supply and demand to ensure delivery and system reliability. The balancing market complements the day-ahead market but it is not the only technical market. To minimize reaction time in case of a mismatch between supply and demand, the ISO also runs ancillary services which typically involves spinning and non-spinning reserve, up and down regulation, responsive reserve service, black start and reactive services [39].

In this chapter, a stochastic dispatching model that co-optimizes simultaneously day-ahead and balancing markets is presented. This kind of model is appropriate for those power systems with a significant penetration of renewable resources [40]. The model consists of a two-stage stochastic programming problem, whose first stage is day-ahead scheduling, and the second stage is the real-time system operation under a set of plausible scenarios. Both stages are part of the ISO scheduling bids for day-ahead. For example, the ISO clears the market at $t = 0$ each day for the next day ($t = 24$ to $t = 48$). 15 minutes before starting the next day ($t = 23:45$) it adjusts the bids and offers based on new information gained of uncertain processes in the market and thus updates the previous variables at each time and each scenario.

In this manner, each scenario constitutes a possible realization of the stochastic processes together with an occurrence probability [41]. The final output of the two-stage optimization process is the day-ahead schedule, with the second stage as part of that process to consider scenarios of how resources would be dispatched. It still remains to actually move through time and have a balancing market active based on the schedule set the day before.

A scenario tree comprises a set of nodes and branches as shown in Figure 6. The nodes represent the points where decisions are made. In the root node, the first-stage decisions are made. The nodes connected to the root node are the second-stage nodes (leaf node)

and represent the points where the second-stage decisions are made. They constitute the real-time operation of the power system in order to accommodate the specific realization of the wind power production with adequate reserve deployment. A scenario is a single path between the root and a leaf node. The set of scenarios characterizes the stochastic processes considered in this problem are wind power production. In the technical literature is possible to find multiple scenarios generating procedures [42]. The stochastic wind power production is described with a discrete probability distribution [43].

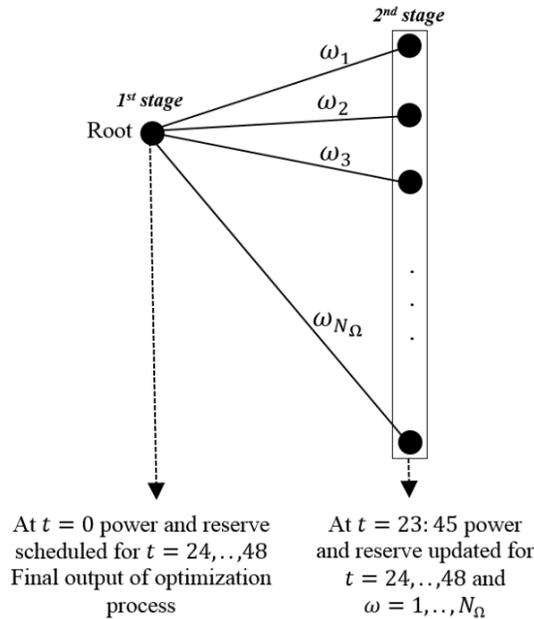


Figure 6. Scenario tree for the two-stage problem

Specifically, the considered decision-making process faced by the ISO is the following: at the beginning of the planning horizon, the day-ahead scheduling of the production for the whole planning horizon is decided for generating units and LAs. Afterward in the second stage for each hour and scenario, the ISO adjusts the real-time dispatch decisions. Note that these decisions depend on the energy schedule previously decided and on the availability of renewable energy. The entire two-stage formulation is solved in one integrated optimization problem.

2. Problem Formulation

The objective function to be minimized separately groups those terms representing the costs pertaining to the day-ahead scheduling and real-time operation of the system. Three sets of constraints are first stage constraints; second stage constraints and finally, the

linking constraints which bind the day-ahead market decisions to the real-time operation of the power system through the deployment of reserves provided by generation units and loads. First stage variables define the day ahead scheduling decisions while the second stage variables are the knobs that we must tweak to meet the constraints of the day-ahead schedule and real-time system operation. The resulting model is formulated as a mixed-integer linear programming problem. The notation and formulation proposed in [42] is used for consistency to show the effect of demand-side reserve participation.

2.1. Objective Function

The objective function seeks to minimize the total expected cost of the system consists of energy production cost, generation-side reserve cost, and demand-side reserve cost. The balancing market is also accounted for in the cost function implicitly in (1). The list of variables is provided in Symbols section at the beginning of the thesis.

$$\text{Min } Z = \text{Min} \left\{ \begin{aligned} & \sum_{t=1}^{N_T} \sum_{i=1}^{N_G} C_{it}^E + \sum_{t=1}^{N_T} \sum_{i=1}^{N_G} C_{it}^R + \sum_{t=1}^{N_T} \sum_{j=1}^{N_L} C_{jt}^L - \sum_{t=1}^{N_T} \sum_{j=1}^{N_L} d_t \lambda_{jt}^L L_{jt}^S \\ & + \sum_{\omega=1}^{N_\Omega} \sum_{t=1}^{N_T} \sum_{j=1}^{N_L} \pi_\omega d_t V_{jt}^{LOL} L_{jt\omega}^{shed} + \sum_{\omega=1}^{N_\Omega} \sum_{t=1}^{N_T} C_t^{WS} S_{t\omega} \end{aligned} \right\} \quad (1)$$

The term $\sum_{t=1}^{N_T} \sum_{j=1}^{N_L} d_t \lambda_{jt}^L L_{jt}^S$ is the consumers' utility function. Due to the correlation between electricity consumption and price, consumer utility maximization promotes the overall system welfare. The term $\sum_{\omega=1}^{N_\Omega} \sum_{t=1}^{N_T} \sum_{j=1}^{N_L} \pi_\omega d_t V_{jt}^{LOL} L_{jt\omega}^{shed}$ accounts for involuntarily load shedding. The energy production cost function in (2) represents start-up and variable costs of generation units including wind power generator (WPG) (assuming zero operating cost component for WPG) besides cost due to change in the start-up plan of units for all scenarios:

$$\sum_{t=1}^{N_T} \sum_{i=1}^{N_G} C_{it}^E = \sum_{t=1}^{N_T} \sum_{i=1}^{N_G} C_{it}^{SU} + \sum_{t=1}^{N_T} \sum_{i=1}^{N_G} \sum_{m=1}^{N_M} d_t \lambda_{it}^G(m) P_{it}^G(m) + \sum_{\omega=1}^{N_\Omega} \sum_{t=1}^{N_T} \sum_{i=1}^{N_G} \pi_\omega C_{it\omega}^A \quad (2)$$

The reserve cost function shown in (3) accounts for generation-side reserve offers consisting of scheduled up, down and non-spinning reserves in addition to deployed reserve in real time:

$$\sum_{t=1}^{N_T} \sum_{i=1}^{N_G} C_{it}^R = \sum_{t=1}^{N_T} \sum_{i=1}^{N_G} d_t \left(C_{it}^{R^U} R_{it}^U + C_{it}^{R^D} R_{it}^D + C_{it}^{R^{NS}} R_{it}^{NS} \right) \\ + \sum_{\omega=1}^{N_\Omega} \sum_{t=1}^{N_T} \sum_{i=1}^{N_G} \sum_{m=1}^{N_M} \pi_\omega d_t \lambda_{it}^G(m) r_{it\omega}^G(m) \quad (3)$$

The demand-side services cost function in (4) includes scheduled up and down reserves in addition to deployed reserve in real time.

$$\sum_{t=1}^{N_T} \sum_{j=1}^{N_L} C_{jt}^L = \sum_{t=1}^{N_T} \sum_{j=1}^{N_L} d_t \left(C_{jt}^{R^U} R_{jt}^U + C_{jt}^{R^D} R_{jt}^D \right) + \sum_{\omega=1}^{N_\Omega} \sum_{t=1}^{N_T} \sum_{j=1}^{N_L} \pi_\omega d_t \lambda_{jt}^L (r_{jt\omega}^U - r_{jt\omega}^D) \quad (4)$$

2.2. Constraints

Three sets of constraints are imposed on the optimization as follows.

2.2.1. First stage constraints

Real power generation and load balance constraints:

$$\sum_{i=1}^{N_G} P_{it}^S = \sum_{j=1}^{N_L} L_{jt}^S, \quad \forall t \quad (5)$$

$$P_i^{\min} u_{it} \leq P_{it}^S \leq P_i^{\max} u_{it}, \quad \forall i, \forall t \quad (6)$$

$$L_{jt}^{S,\min} \leq L_{jt}^S \leq L_{jt}^{S,\max}, \quad \forall i, \forall t \quad (7)$$

$$0 \leq P_{it}^G(m) \leq P_{it}^{G,\max}(m), \quad \forall m, \forall i, \forall t \quad (8)$$

$$P_{it}^S = \sum_{m=1}^{N_M} P_{it}^G(m), \quad \forall i, \forall t \quad (9)$$

Up and down spinning reserve and non-spinning reserve limits for generating units in:

$$0 \leq R_{it}^U \leq R_{it}^{U,\max} u_{it}, \quad \forall i, \forall t \quad (10)$$

$$0 \leq R_{it}^D \leq R_{it}^{D,\max} u_{it}, \quad \forall i, \forall t \quad (11)$$

$$0 \leq R_{it}^{NS} \leq R_{it}^{NS,\max} (1 - u_{it}), \quad \forall i, \forall t \quad (12)$$

Up and down LA reserve constraints:

$$0 \leq R_{jt}^U \leq R_{jt}^{U,max} \quad , \forall j, \forall t(13)$$

$$0 \leq R_{jt}^D \leq R_{jt}^{D,max} \quad , \forall j, \forall t(14)$$

Generating units' start-up cost constraint:

$$C_{it}^{SU} \geq \lambda_{it}^{SU} (u_{it} - u_{i,t-1}) \quad , \forall i, \forall t(15)$$

$$C_{it}^{SU} \geq 0 \quad , \forall i, \forall t(16)$$

Minimum up and down time constraints:

$$[X_i^{on} - UT_i](u_{i,t-1} - u_{it}) \geq 0 \quad , \forall i, \forall t(17)$$

$$[X_i^{off} - DT_i](u_{it} - u_{i,t-1}) \geq 0 \quad , \forall i, \forall t(18)$$

Ramping up and down constraints:

$$P_{it}^S - P_{i,t-1}^S \leq RU_i u_{i,t-1} + SU_i [u_{it} - u_{i,t-1}] + (1 - u_{it}) P_i^{max} \quad , \forall i, \forall t(19)$$

$$P_{i,t-1}^S - P_{it}^S \leq RD_i u_{it} + SD_i [u_{i,t-1} - u_{it}] + (1 - u_{i,t-1}) P_i^{max} \quad , \forall i, \forall t(20)$$

2.2.2. Second stage constraints

Power balance for the nodes at which the WPG is located:

$$\sum_{i:(i,n) \in M_G} P_{it\omega}^G - \sum_{j:(j,n) \in M_L} (L_{jt\omega}^C - L_{jt\omega}^{shed}) + P_{t\omega}^{WPG} - S_{t\omega} - \sum_{r:(n,r) \in \Lambda} f_{t\omega}(n,r) = 0$$

$$n = n', \forall t, \forall \omega(21)$$

Power balance for rest of the nodes:

$$\sum_{i:(i,n) \in M_G} P_{it\omega}^G - \sum_{j:(j,n) \in M_L} (L_{jt\omega}^C - L_{jt\omega}^{shed}) - \sum_{r:(n,r) \in \Lambda} f_{t\omega}(n,r) = 0 \quad \forall n$$

$$\neq n', \forall t, \forall \omega(22)$$

DC power flow equation in steady state:

$$f_{t\omega}(n,r) = \frac{\delta_{nt\omega} - \delta_{rt\omega}}{X(n,r)} \quad \forall (n,r) \in \Lambda, \forall t, \forall \omega(23)$$

Transmission flow limits in the base case:

$$-f^{max}(n, r) \leq f_{t\omega}(n, r) \leq f^{max}(n, r) \quad \forall(n, r) \in \Lambda, \forall t, \forall \omega(24)$$

Constraints (25) and (26) accounts for the decomposition of the deployed reserve into blocks for generating units:

$$r_{it\omega}^U + r_{it\omega}^{NS} - r_{it\omega}^D = \sum_{m=1}^{N_M} r_{it\omega}^G(m) \quad , \forall i, \forall t, \forall \omega(25)$$

$$-P_{it}^G(m) \leq r_{it\omega}^G(m) \leq P_{it}^{G,max}(m) - P_{it}^G(m) \quad , \forall m, \forall i, \forall t, \forall \omega(26)$$

Deployed power generation and ramp limits:

$$P_i^{min} v_{it\omega} \leq P_{it\omega}^G \leq P_i^{max} v_{it\omega} \quad , \forall i, \forall t, \forall \omega(27)$$

$$P_{it\omega}^G - P_{i,t-1,\omega}^G \leq RU_i v_{i,t-1,\omega} + SU_i [v_{it\omega} - v_{i,t-1,\omega}] + (1 - v_{it\omega}) P_i^{max} \quad , \forall i, \forall t, \forall \omega(28)$$

$$P_{i,t-1,\omega}^S - P_{it\omega}^S \leq RD_i v_{it\omega} + SD_i [v_{i,t-1,\omega} - v_{it\omega}] + (1 - v_{i,t-1,\omega}) P_i^{max} \quad , \forall i, \forall t, \forall \omega(29)$$

Involuntary load shedding limit:

$$0 \leq L_{jt\omega}^{shed} \leq L_{jt\omega}^C \quad , \forall i, \forall t, \forall \omega(30)$$

Wind power spillage limit:

$$0 \leq S_{t\omega} \leq P_{t\omega}^{WP} \quad , \forall t, \forall \omega(31)$$

2.2.3. Linking constraints

Deployed up and down spinning reserve and non-spinning reserve constraints for generating units:

$$P_{it\omega}^G = P_{it}^S + r_{it\omega}^U + r_{it\omega}^{NS} - r_{it\omega}^D \quad , \forall i, \forall t, \forall \omega(32)$$

Deployed LA reserve constraint:

$$L_{jt\omega}^C = L_{jt}^S - r_{jt\omega}^U + r_{jt\omega}^D \quad , \forall j, \forall t, \forall \omega(33)$$

Deployed reserve limits:

$$0 \leq r_{it\omega}^U \leq R_{it}^U \quad , \forall i, \forall t, \forall \omega(34)$$

$$0 \leq r_{it\omega}^D \leq R_{it}^D \quad , \forall i, \forall t, \forall \omega(35)$$

$$0 \leq r_{it\omega}^{NS} \leq R_{it}^{NS} \quad , \forall i, \forall t, \forall \omega(36)$$

$$0 \leq r_{jt\omega}^U \leq R_{jt}^U, \forall j, \forall t, \forall \omega (37)$$

$$0 \leq r_{jt\omega}^D \leq R_{jt}^D, \forall j, \forall t, \forall \omega (38)$$

Adjustment cost in the start-up plan:

$$C_{it\omega}^A = C_{it\omega}^{SU} - C_{it}^{SU}, \forall i, \forall t, \forall \omega (39)$$

$$C_{it\omega}^{SU} \geq \lambda_{it}^{SU} (v_{it\omega} - v_{i,t-1,\omega}), \forall i, \forall t, \forall \omega (40)$$

$$C_{it\omega}^{SU} \geq 0, \forall i, \forall t, \forall \omega (41)$$

Each LA reserve quantity is:

$$q_{jt\omega} = r_{jt\omega}^D - r_{jt\omega}^U = q_{jt\omega}(0)x_{jt\omega} + \sum_{k=1}^{N_K} \gamma_{jt\omega}(k) x_{jt\omega}, \forall j, \forall t, \forall \omega (42)$$

$$\gamma_{jt\omega}(k) = q_{jt\omega}(k) - q_{jt\omega}(k-1), \forall j, \forall t, \forall \omega (43)$$

$$u_{it}, v_{it\omega}, x_{jt\omega} \in \{0,1\}, \forall i, \forall t, \forall \omega (44)$$

where $q_{jt\omega}(0)$ is greater than the minimum curtailment level of the DR program specified by the ISO. The LA's bid-quantity offer curve in period t is the same as suggested in [8]. Note that DR costs are assumed here as parameters rather than how to provide the actual DR loads.

3. Computer Simulation

The proposed formulation is demonstrated on a three-bus system [42] and the 24-bus IEEE reliability test system (RTS) [44]. The case studies were solved using the MILP solver of CPLEX 12.2 [45] on a desktop computer with a 3.4-GHz i7 processor and 16 GB of RAM. The computation time is trivial for the three-bus system and less than 30 seconds for the IEEE RTS. The upper bound on the duality gap is set to be zero and the minimum up and down time constraints are not considered in this study.

3.1. Three-bus System Case Study

The three-bus system shown in Figure 7 is used to demonstrate the features of the proposed method. Table 1 provides an example of wind power scenarios. The scheduling horizon spans four time periods. The operator generates three possible wind production

scenarios $\omega \in N_\Omega = \{Forecasted, High, Low\}$ with probabilities 0.6, 0.2 and 0.2, respectively. The ISO then tries to find a strategy that minimizes the *expected cost* (or equivalently maximize the *expected profit*).

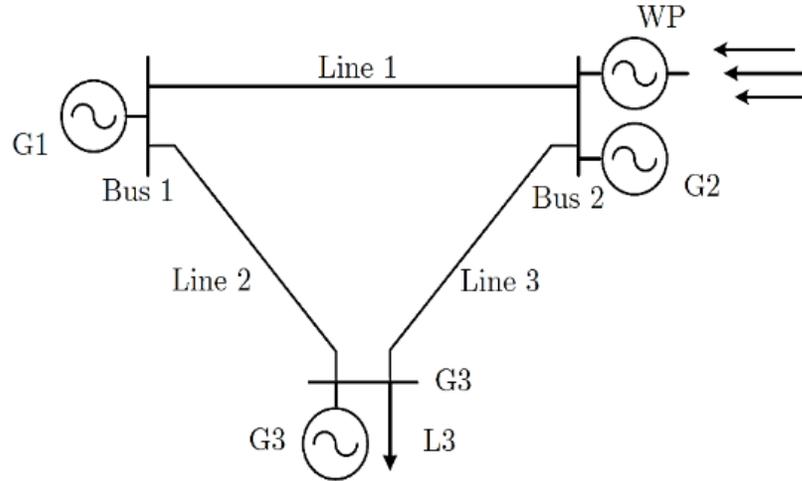


Figure 7. One-line diagram of the three-bus system

Table 1. Wind power scenarios

Period #	$P_{t\omega}^{WP}$ [MW]		
	As forecast	High	Low
1	6	9	2
2	20	30	13
3	35	50	25
4	8	12	6

Figure 8 presents the total expected cost of the system for different demand side reserve offer costs ($C_{jt}^{R^U}$ and $C_{jt}^{R^D}$ assumed to be equal). It highlights the fact that facilitating a higher share for DR in reserve (i.e. decreasing the cost of DR participation) will increase the overall system welfare. Based on this, two case studies were conducted for different demand side reserve costs of 10 \$/MWh and 20 \$/MWh.

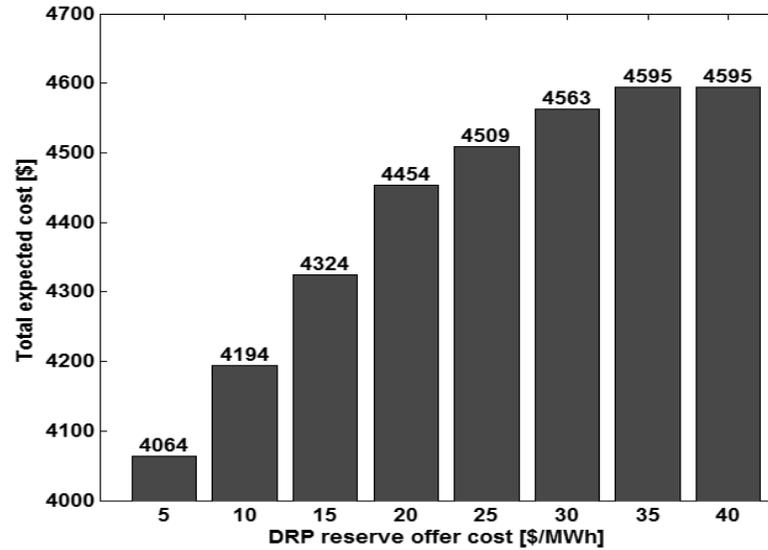


Figure 8. Total expected cost of the system variations

Table 2 shows how utilizing DR reserve decreases the wind power spillage and consequently facilitate higher share of renewable energy integration in the power system. When demand side reserve costs are \$20/MWh, 7 MW of wind power production is spilled in the *High* scenario. Tables 3 and 4 summarize the optimal results for the two cases. In addition, the higher the LA reserve cost, the higher the locational marginal price (LMP).

Table 2. Amount of spilled wind power production [MW]

Demand-side Reserve Cost	WP Scenarios	Period #			
		1	2	3	4
10\$/MWh (Case1)	ω_1	0	0	0	0
	ω_2	0	0	0	0
	ω_3	0	0	0	0
20\$/MWh (Case2)	ω_1	0	0	0	0
	ω_2	3	0	0	4
	ω_3	0	0	0	0

Table 3. Scheduling results [MW] – Case 1

	Unit	Period #			
		1	2	3	4
P_{it}^G	G1	0	0	0	0
	G2	0	0	0	0
	G3	21	42	50	28
R_{it}^U	G1	0	0	0	0
	G2	0	0	0	0
	G3	4	0	0	2
R_{it}^D	G1	0	0	0	0
	G2	0	0	0	0
	G3	0	0	0	0
P_{it}^G	WP	9	38	60	12
R_{jt}^U	LA	3	8	11	4
R_{jt}^D	LA	0	0	0	0

Table 4. Scheduling results [MW] – Case 2

	Unit	Period #			
		1	2	3	4
P_{it}^G	G1	0	0	0	0
	G2	0	0	0	0
	G3	24	42	50	32
R_{it}^U	G1	0	0	0	0
	G2	0	0	0	0
	G3	4	0	0	2
R_{it}^D	G1	0	0	0	0
	G2	0	0	0	0
	G3	0	0	0	0
P_{it}^G	WP	6	38	60	8
R_{jt}^U	LA	0	8	11	0
R_{jt}^D	LA	0	0	0	0

Figure 9 shows the variation of LMPs under first scenario (*Forecasted*) over scheduling horizon when DR reserve cost is \$5/MWh and \$40/MWh. LMPs are the marginal values obtained from power balance equations (21), (22) or the Lagrange multipliers of the dual problem.

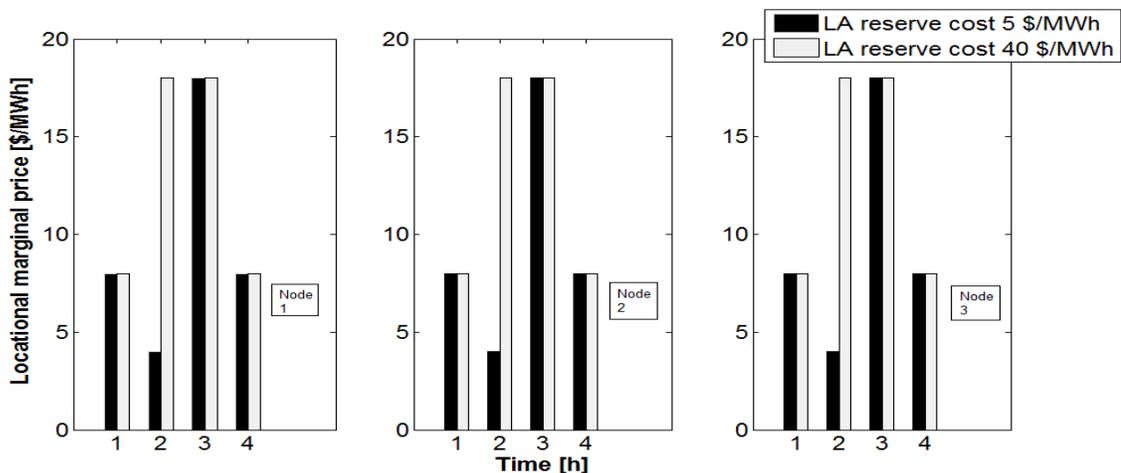


Figure 9. LMP variations for different LA reserve costs

3.2. IEEE-RTS Case Study

The proposed model is applied over a 24-hour horizon to the IEEE-RTS shown in Figure 10.

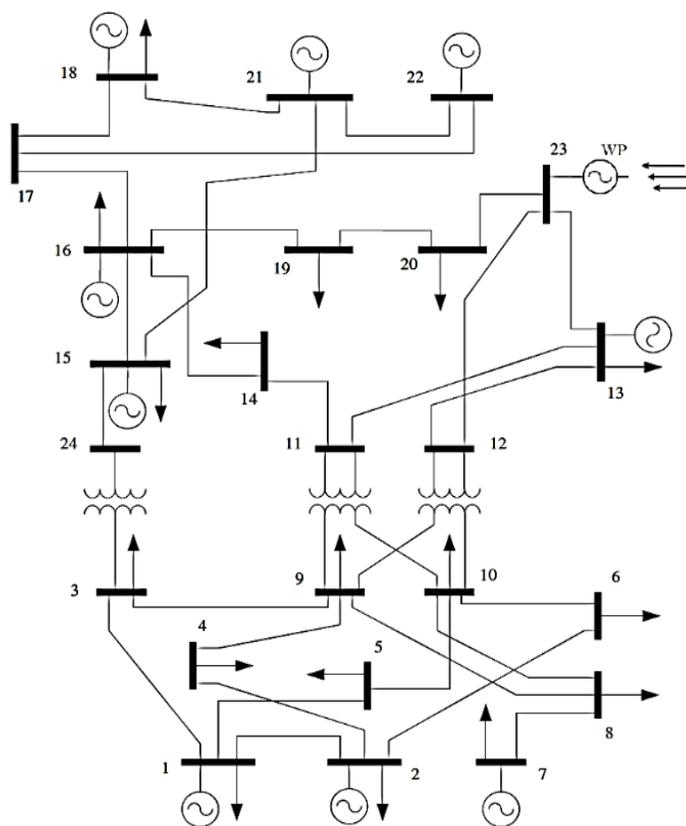


Figure 10. One-line diagram of the IEEE-RTS

The system data are given in [44] including costs. The power system includes 12 generating units. Units 8 and 9 are nuclear power plants and unit 10 is a hydro generator. Unit 12 is a wind power generator whereas the rest are thermal. The wind, hydro and nuclear generators are assumed to be must-run units. The LAs' contribution in percentage to the total system demand is listed in Table 5 [42].

Table 5. Distribution of the total system demand

LA #	% of system load	LA#	% of system load
1	3.8	10	6.8
2	3.4	11	9.3
3	6.3	12	6.8
4	2.6	13	11.1
5	2.5	14	3.5
6	4.8	15	11.7
7	4.4	16	6.4
8	6.0	17	4.5
9	6.1		

The load profile for the 24-hour scheduling horizon is shown in Figure 11 in a stacked bar chart. The shading shows the contribution of each of the 17 LAs based on Table 5. All the LAs offer identical rates for up and down spinning reserves. LAs can be curtailed from their normal levels up to $\alpha\%$ of their scheduled consumption for all hours to provide up spinning reserves. Likewise, LAs can increase by the same quantity to provide down spinning reserve.

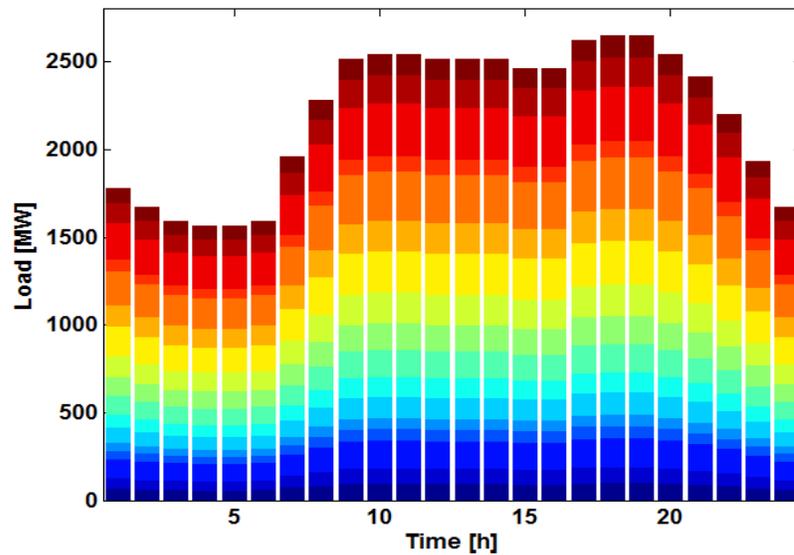


Figure 11. Daily load profile of the IEEE-RTS

For scenario generation the stochastic process wind power (P^{WP}) is determined using an autoregressive moving average (ARMA) model. The resulting model can then be used to simulate the output of the system for a given input or more generally system analysis, prediction and also control design. An ARMA (p, q) process WP is mathematically expressed as:

$$P_t^{WP} = \sum_{j=1}^p \phi_j P_{t-j}^{WP} + \varepsilon_t - \sum_{j=1}^q \theta_j \varepsilon_{t-j} \quad (45)$$

with p autoregressive parameters $\phi_1, \phi_2, \dots, \phi_p$, and q moving average parameters $\theta_1, \theta_2, \dots, \theta_q$. The term ε_t in equation (45) represents an uncorrelated normal stochastic process with mean zero and variance σ_ε^2 and is also uncorrelated with $P_{t-1}^{WP}, P_{t-2}^{WP}, \dots, P_{t-p}^{WP}$. The stochastic process ε_t is also referred to as white noise, an innovation term, or an error term [42]. In this study the following ARMA (1,1) model for characterizing P^{WP} is considered:

$$P_t^{WP} = \phi_1 P_{t-1}^{WP} + \varepsilon_t - \theta_1 \varepsilon_{t-1} + c \quad (45)$$

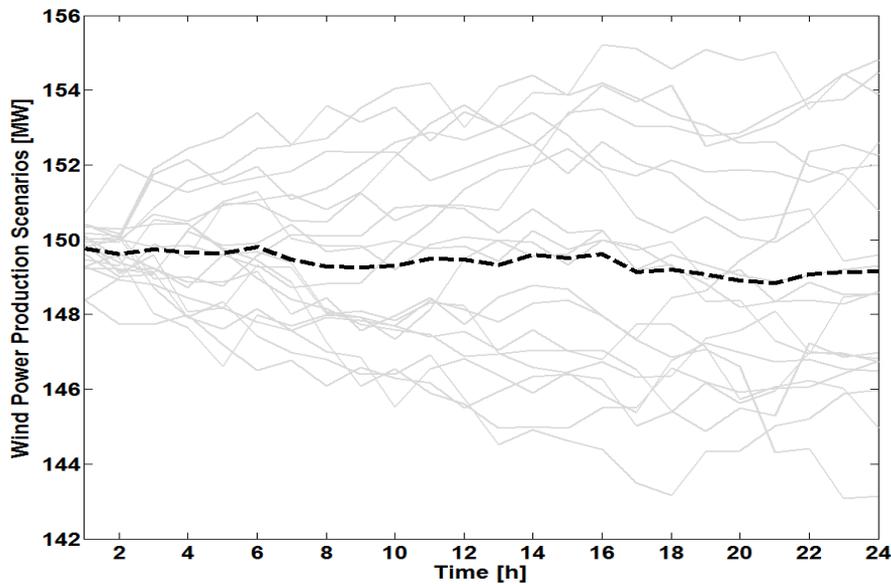


Figure 12. Simulated ARMA(1,1) process, 20 scenarios, each with 24 observations, dashed line represents the mean wind power production

The parameters of the model are $\phi_1 = 0.99$ and $\theta_1 = 0.011$ and constant $c = 0.15$. The standard deviation of the error term is 0.021 based on the scaled data of wind power production in Denmark for the first day of August 2013 [46]. In order to make the

optimization problem tractable, a reduction technique can be applied to obtain a set of reduced scenarios that maintains the statistical properties of the original set to acceptable levels [47]. However, in this study 20 equiprobable scenarios are considered to make the problem tractable as shown in Figure 10 where the dashed line represents the average wind power production. Each path is a single realization of the stochastic process, i.e., one wind power generation scenario for $t = 24$ to $t = 48$.

Different cases are looked at to illustrate the impacts of DR reserve utilization. Figure 13 presents the total expected cost of the system for different demand side reserve offer costs. It again highlights the fact that facilitating a higher share of DR reserve will increase the overall system welfare. Based on this, two cases are examined in the following for different demand side reserve costs of \$5 and \$30 per MWh.

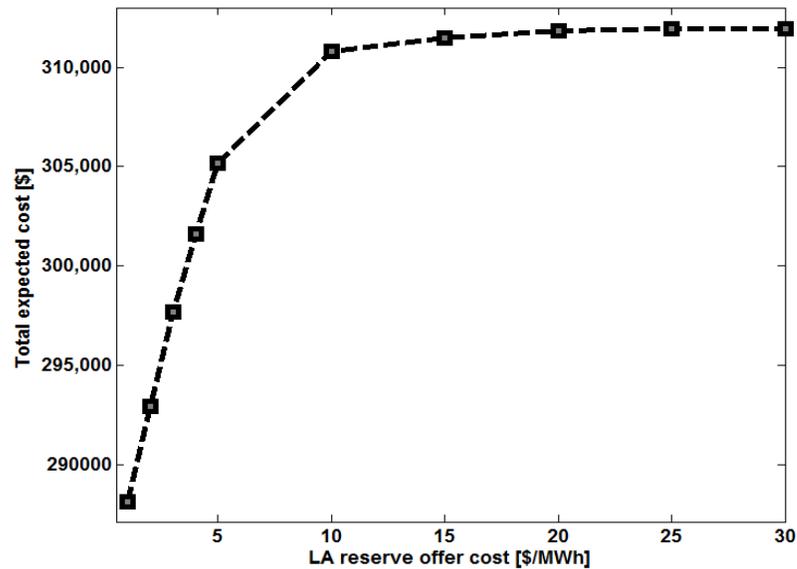


Figure 13. Total expected cost of the system variations

Table 6 shows how utilizing DR reserve decreases the wind power spillage and consequently facilitates a higher share of renewable energy integration in the power system. When demand side reserve cost is 30 \$/MWh, on average 2.3 MW of wind power production is spilled compared to 1.46 MW wind power spillage when the LA reserve cost is \$5/MWh. Note that the total number of events is $24 \text{ hours} \times 20 \text{ scenarios} = 480$.

Table 6. Amount of spilled wind power production

LA reserve offer cost [\$/MWh]	\$5	\$30
Average wind power spillage [MW]	1.46	2.30
Wind power spillage incidents	18	109

DR could also modify the scheduled commitment status of generating units and offer more economic options. As it is shown in table 7, in case 1 (5 \$/MWh) at hour 22 the expensive generating units 1 and 2 would not be committed compared to case 2 (30 \$/MWh).

Table 7. Number of committed units at each hour

Hours	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Case1	8	6	6	6	6	6	9	9	9	10	10	9	9	9	9	9	10	10	10	10	9	7	7	5
Case2	8	6	6	6	6	6	9	9	9	10	10	9	9	9	9	9	10	10	10	10	9	9	7	5

The average marginal price of electricity is shown in Figure 14 for three different cases where the demand-side reserve offer costs are \$3/MWh, \$5/MWh and \$10/MWh. This highlights the fact that deploying DR decreases the average price of electricity.

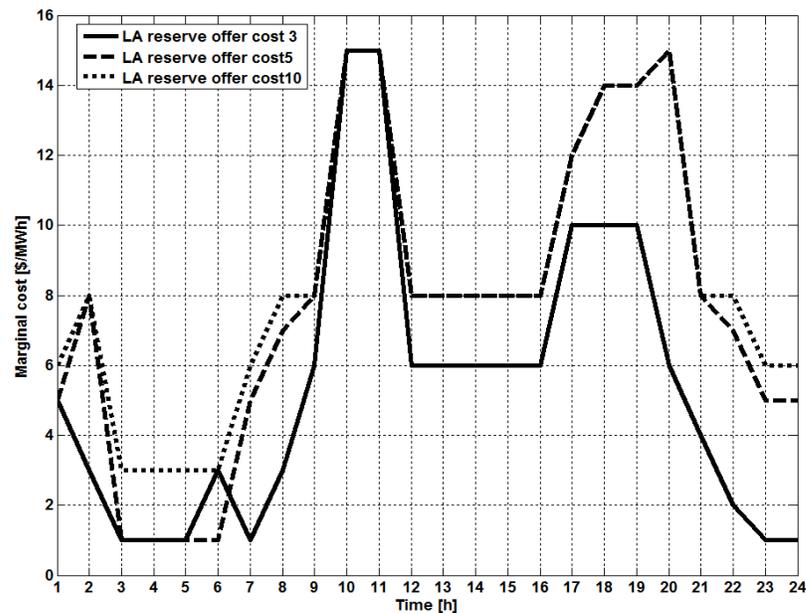


Figure 14. Marginal cost variations for different demand side reserve offer costs

The marginal cost associated with constraint (5) shows how much the optimal value of the objective would increase per unit increase in the amount of resources available. In other

words, the marginal value associated with a resource represents how much more profit ISO would gain by increasing the amount of that resource by one unit.

Table 8 shows the effect of DR penetration level on the system. The higher the penetration level, the higher the performance of the system in terms of reduced total expected cost, average wind power spillage for different scenarios and the marginal cost of the system.

Table 8. Impact of DR level on the system (LA reserve cost is 5 \$/MWh)

% of DR penetration level (α)	Total expected cost [\$]	Average wind power spillage [MW]	Average marginal generation cost [\$/MWh]
5 %	306,660	1.67	7.87
10 %	305,195	1.46	7.79
20 %	305,013	1.32	7.75

4. Summary

The stochastic model has been used to schedule energy and reserve provided by generation and demand-side resources. It is necessary to adopt stochastic approaches for decision making under uncertainty as in the case of clearing electricity markets with high levels of wind generation. This permits the scheduling of loads and services in coordination with the system operator, thus enhancing the power system's efficiency and security while reducing its environmental impact. The proposed stochastic model is formulated as a two-stage SMIP problem. A number of case studies are conducted and the results presented demonstrate the benefits of customers' response for ISO. DR can potentially benefit consumers by offering cheaper electricity while facilitating a higher share of renewable energy sources in the power system.

Chapter 4: Baseline Load Forecasting Using Bayesian Approach

Any demand response program requires a clear procedure for determining customer baseline loads that are intended for use by customers or aggregators enrolled in that program. Aggregators of retail end users may claim load reduction or ancillary service provision from demand side resources with which they have a contractual arrangement. Customer Base Load (CBL) is the average power demand of the resource if the DR had not occurred (not necessarily an hourly period). CBL is used to determine the level of load curtailment provided and thereby compensate the participant.

There are many examples of baseline models in the literature. For example, Coughlin et al [48] compared the performance of predictions produced by different baseline models statistically for commercial buildings. Goldberg et al [49] developed a standardized M&V procedure. Various baseline and M&V methodologies currently developed have been discussed in [50]. Day matching and regression analysis [51, 52] are the two most common techniques for calculating baselines. For example, California electric utilities use methods such as averaging the consumption profiles of the three days with the highest energy usage out of the last ten business days [4]. A rolling average baseline has been used by ISO-NE [53]. In NYISO, there are two sub-categories for CBL calculation: Average Day CBL and Adjusted CBL. It is the responsibility of the customer service provider (CSP) that during the enrolment of any resource in each capability period, it must identify the type of CBL for that resource. The choice of CBL method can be changed when registering for the next capability period. It is also the responsibility of the CSP to provide a complete and accurate CBL calculation to the NYISO, based on the chosen method during capability period registration. The details of how to calculate the CBL in NYISO is explained in Appendix A.

In Ref. [54] the authors investigated the effect of baseline modeling implementation on DR assessment using the linear regression baseline model developed in [55]. The results indicate strong sensitivities to outside air temperature and data filtration method, but weak sensitivities to data resolution. Ref. [56] proposed an exponential smoothing model with weather adjustment to compute customer baseline load. The baseline calculation for

individual or aggregated residential customers has not been addressed so far. Current research has attempted to address it using historical data analysis and ordinary regression techniques at best which do not address uncertainty appropriately. Most of the baseline models discussed are applicable to commercial and industrial customers. Besides, baseline estimation is an evolving process with a limiting factor being data availability. New methods of establishing baselines and measuring their accuracies are required as DR utilizing advanced technologies that will be enabled by Advanced Metering Infrastructure (AMI) is deployed more widely [57].

In this thesis, we address this issue using statistical machine learning. Bayesian Linear Regression (BLR) has been employed due to several advantages over Ordinary Linear Regression (OLR) such as: the use of prior information to update existing parameters, improved performance for small samples, higher flexibility and greater stability, lower influence of outliers and directly estimating uncertainty in the parameters and predicted values [58-60]. In machine learning, offline learning algorithms are used when one has access to the entire training dataset at once in advance. Online learning methods, which update their solution iteratively, are a promising approach towards baseline modeling with improvements in real time. The mapping from the input dataset to the corresponding outputs is updated after the arrival of the new dataset. The recursive Bayesian approach has been used in this work as an online algorithm as new data becomes available in a sequential fashion.

Specifically, we propose a self-learning and updating BLR model to predict power consumption using an autoregressive approach. Two case-studies are explored: Vancouver, BC, Canada where hourly load data is used; and Austin, TX, USA where minute resolution data is used. The performance of the models is then examined and compared before summarizing the main research results.

The rest of this chapter is organized as follows. An overview of data-driven models and the proposed Bayesian Linear Regression model is presented and formulated. Next, data collection and model implementation procedures are explained. The numerical simulation conducted on two case studies are presented to quantify the benefits of the Bayesian approach.

1. Data-driven Models and Uncertainty Analysis

A Baseline model is a type of data-driven model, which is a class of numerical models that create generalized links between input and output datasets [61]. Examples of data-driven models include linear regression, autoregressive models, neural networks, fuzzy regression and fuzzy rule-based systems, model trees, and genetic programming. Data-driven methods have been widely used in many fields because they provide good agreement between observed and modeled data [62]. They are generally easier to calibrate, are based on objective information and require limited assumptions about the physical process being modeled. In addition, data-driven models are useful in solving practical problems, especially when knowledge driven simulation models cannot be constructed due to a lack of understanding of the underlying processes or when existing models are inadequate. Although data-driven models have higher data requirements to calibrate and validate the model, with the advent of the Smart Grid and increased use of real-time monitoring, there exists a great opportunity to utilize high-resolution data that is continuously being collected. Thus, the availability of data which is often cited as a limiting factor in the success of data-driven models is no longer a major issue [59, 60].

A general drawback of numerical representations of physical systems is uncertainty in the input and output data, and in the parameters. Typically, probability-based methods such as Bayesian inference, are used to address this uncertainty in data-driven or conceptual models [63-66]. The advantages of Bayesian analysis, in general, is that it provides a statistical or probabilistic representation of a system, rather than the typical deterministic representation, thus providing confidence intervals of the predictions [67, 68].

Consequently, the uncertainty in a model is represented by including probability distributions of various parameters in the model. The use of Bayesian-based methods has increased significantly in recent years, due to the fact that computing speed has increased significantly [69]. The required numerical solutions (Monte Carlo iterations) required for Bayesian problems can now be found quickly when no analytical solutions exist. Furthermore, there is no need to make assumptions and build-up a model from governing equations that do not really exist in many cases [70]. With this approach, we can naturally address issues like regularization, model selection or comparison without the need for a separate cross-validation data set.

The basic principle of Bayesian applications in uncertainty analysis in numerical models is that any prior information (*i.e.* from previous research, or expert knowledge) can be included in the current analysis to estimate or update current model parameters [66], [71, 72]. This prior is then combined with existing data to calculate the posterior predictive probability of parameters [58-60]. In general, in cases where uncertainty in model parameters is important, a Bayesian analysis is preferred over a simple deterministic analysis [73].

2. Linear Regression Review

Regression analysis generates an equation to describe the statistical relationship between predictor variables and the response variable. The equation or hypothesis is a linear function in terms of predictor variables weights as shown in (1) where \mathbf{X} is a matrix of independent vector variables and the $\boldsymbol{\beta}$ are the corresponding regression coefficients.

$$h(x) = \sum_{i=0}^d \beta_i x_i = \mathbf{X}\boldsymbol{\beta} \quad (1)$$

The goal is to minimize the difference between the hypothesis and actual output shown in (2), basically, we try to minimize the residual sum of squares (RSS):

$$E_{in}(h) = \frac{1}{N} \sum_{n=1}^N (h(x_n) - y_n)^2 = \frac{1}{N} \sum_{n=1}^N (\boldsymbol{\beta}^T x_n - y_n)^2 = \frac{1}{N} \|\mathbf{X}\boldsymbol{\beta} - \mathbf{y}\|^2 \quad (2)$$

Equation (2) can be solve using gradient descent algorithm but it also has a close form solution by setting the gradient of $E_{in}(h)$ to zero. The optimal weight vector is

$$\boldsymbol{\beta}^* = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (3)$$

Equation (3) is called least-square estimates.

To prevent over-fitting one can add a regularization term to the objective function. Two possible approaches are:

- Ridge Regression (a.k.a L2 Regularization) : $RSS(w) + \lambda \|w\|_2^2$

$$\boldsymbol{\beta}^* = (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T \mathbf{y} \quad (4)$$

Coefficient estimates for multiple linear regression models rely on the independence of the model terms. When terms are correlated and the columns of the design matrix \mathbf{X} have an approximate linear dependence, the matrix $(\mathbf{X}^T \mathbf{X})^{-1}$ becomes close to singular. As a result, the least-squares estimate becomes highly sensitive to random errors in the observed

response y , producing a large variance. *Ridge regression* addresses the problem by estimating regression coefficients using β^* where λ is the *tuning parameter* and I is the identity matrix. While biased, the reduced variance of ridge estimates often result in a smaller mean square error when compared to least-squares estimates.

- Lasso Regression (a.k.a L1 Regularization): $RSS(w) + \lambda \|\beta\|_1$

Same as ridge regression, the solution is governed by the continuous parameter λ . It is a convex objective function but the derivative does not exist. Even if we can compute the derivative, there is no closed form solution. One can solve the problem using sub-gradient descent like coordinate descent algorithm. Lasso identifies and removes the redundant predictors and hence leads to a sparse solution.

3. Bayesian Linear Regression

Bayesian linear regression is an approach to ordinary linear regression (OLR) within a Bayesian framework, which can be represented as a Normal probability model:

$$\mathbf{y} \sim N(\mathbf{X}\beta, \sigma^2 \mathbf{I}) \quad (5)$$

where \mathbf{X} is a matrix of independent vector variables, limited to $[\mathbf{x}_1 \mathbf{x}_2 \mathbf{x}_3]$ for this study with $\mathbf{x}_1 = \mathbf{I}$ to allow for an intercept, $[\mathbf{x}_2 \mathbf{x}_3]$ are the observations (selected features to model the output), the β are the corresponding regression coefficients, σ^2 is the variance of the dependent variable y given the observations (*i.e.* $\mathbf{y}|\beta, \mathbf{X}$), and lastly, \mathbf{I} is the identity matrix.

The posterior distribution, the resulting prediction of variables of interest according to Bayes rule, is proportional to the product of a likelihood function and the prior distribution. This means that any calculated model parameters are a function of the likelihood function (calculated from the current data) and the prior (from any previous data, assumptions, etc.). Thus, the regression coefficients can be estimated via:

$$\mathbf{p}(\beta, \sigma^2 | \mathbf{y}, \mathbf{X}) \propto \mathbf{p}(\mathbf{y} | \beta, \sigma^2, \mathbf{X}) \mathbf{p}(\beta, \sigma^2) \quad (6)$$

where the term on the left-hand side represents the probability distribution of the posterior, and the right-hand side is the product of the likelihood and the prior. In the currently proposed Bayesian framework, $(\mathbf{y}|\beta, \sigma^2, \mathbf{X})$ corresponds to the following likelihood function which is a multivariate Normal:

$$p(\mathbf{y}|\boldsymbol{\beta}, \sigma^2, \mathbf{X}) \propto (\sigma^2)^{-N/2} \exp\left(-\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right) \quad (7)$$

It is often parameterized as:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (8)$$

$$\mathbf{s}^2 = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) / \nu \quad (9)$$

where $\hat{\boldsymbol{\beta}}$ is the ordinary least squares solution to estimate the coefficient vector using the Moore-Penrose pseudo-inverse. Using this formulation, the predictive distribution of \mathbf{y} can be found analytically and is a multivariate function \mathbf{t} :

$$(\mathbf{y}^*|\mathbf{y}) = \mathbf{t}(\mathbf{X}^* \hat{\boldsymbol{\beta}}, \mathbf{s}^2 [\mathbf{I} + \mathbf{X}^* (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^{*T}], \nu) \quad (10)$$

where \mathbf{X}^* are new observations used for prediction and \mathbf{y}^* are the predicted values. The derivation for equations (8) and (9) is not given here, but is available in many textbooks (for example [66], [69], [71]). This particular formulation is for the *non-informative* prior case (details below) which means that these values are identical to the least-squares estimates of degrees of freedom ($\nu = N - k$) (with N the total number of observations and k is the number of regression coefficients), $\hat{\boldsymbol{\beta}}$ and the standard error \mathbf{s}^2 . This prior assumes complete ignorance of prior values of the parameters, essentially a uniform distribution with an extremely large variance. This formulation was used to forecast hourly baseline loads (Vancouver, BC) and minute aggregated household load (Austin, TX) for the first batch of data where no prior information was available (hence the use of a non-informative prior). The model was validated using subsequent batches of data.

For these subsequent models, *independent and informative* priors were used as formulated below. For the calibration procedure, the posterior estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\tau}$ (where $\tau = 1/\sigma^2$) from the previous model are used as the priors to estimate $\boldsymbol{\beta}$ and $\boldsymbol{\tau}$ for the next model. If the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\tau}$ are assumed to be independent, *i.e.* $p(\boldsymbol{\beta}, \boldsymbol{\tau}) = p(\boldsymbol{\beta}) p(\boldsymbol{\tau})$, then the independent priors can be expressed as:

$$\boldsymbol{\beta} \sim N(\boldsymbol{\beta}, \mathbf{V}) \quad (11)$$

$$\boldsymbol{\tau} \sim \Gamma(\mathbf{s}^{-2}, \nu) \quad (12)$$

Multiplying these priors with the likelihood function results in an expression for $(\boldsymbol{\beta}, \boldsymbol{\tau} | \mathbf{y})$ that is unfortunately not in the form of a known, analytically defined density function. Thus, an analytical solution for the joint distribution for both posterior parameters and the predictive distribution for \mathbf{y}^* is not possible. However, a numerical solution procedure using a specialized type of Monte Carlo integration (specifically Gibb's sampling algorithm [69]) is used instead. The conditional posterior distribution for $\boldsymbol{\beta}$ and $\boldsymbol{\tau}$ can be calculated as:

$$(\boldsymbol{\beta} | \boldsymbol{\tau}, \mathbf{y}) \sim N(\bar{\boldsymbol{\beta}}, \bar{\mathbf{V}}) \quad (13)$$

$$(\boldsymbol{\tau} | \boldsymbol{\beta}, \mathbf{y}) \sim \Gamma(\bar{s}^{-2}, \bar{\nu}) \quad (14)$$

where the hyper-parameters can be calculated as follows:

$$\bar{\mathbf{V}} = (\mathbf{V}^{-1} + \boldsymbol{\tau} \mathbf{X}^T \mathbf{X})^{-1} \quad (15)$$

$$\bar{\boldsymbol{\beta}} = \bar{\mathbf{V}}(\mathbf{V}^{-1} + \boldsymbol{\tau} \mathbf{X}^T \mathbf{y}) \quad (16)$$

$$\bar{\nu} = \nu + N \quad (17)$$

$$\bar{s}^{-2} = ((\mathbf{y} - \mathbf{X}\bar{\boldsymbol{\beta}})^T (\mathbf{y} - \mathbf{X}\bar{\boldsymbol{\beta}}) + \nu s^{-2}) / \bar{\nu} \quad (18)$$

The bar above the parameter denotes the ‘‘posterior’’ value of the hyper-parameters. Gibb's sampling algorithm can be used to sample from the posterior conditionals to find the joint posterior distribution $(\boldsymbol{\beta}, \boldsymbol{\tau})$ numerically (adapted from [70]):

- 1) A random value of $\boldsymbol{\tau}$ is selected and is used to calculate one realization of $\bar{\mathbf{V}}$ then $\bar{\boldsymbol{\beta}}$.
- 2) A value of $\boldsymbol{\beta}$ is then sampled from $(\boldsymbol{\beta} | \boldsymbol{\tau}, \mathbf{y}) \sim N(\bar{\boldsymbol{\beta}}, \bar{\mathbf{V}})$
- 3) This value is used to sample $(\boldsymbol{\tau} | \boldsymbol{\beta}, \mathbf{y}) \sim \Gamma(\bar{s}^{-2}, \bar{\nu})$.
- 4) The process is repeated until a desired number of samples are achieved and typically an initial number of samples considered to be ‘‘burn-in’’ are discarded (10% has been used for burn-in).

Each pair of sampled values of $\boldsymbol{\beta}$ and $\boldsymbol{\tau}$ represents an approximation of the joint posterior. The predictive probability of \mathbf{y} can be estimated numerically; for each pair of $\boldsymbol{\beta}$ and $\boldsymbol{\tau}$, a sample for \mathbf{y}^* can be drawn from the predictive distribution:

$$(\mathbf{y}^*) \sim N(\mathbf{X}^* \boldsymbol{\beta}, \mathbf{1}/\boldsymbol{\tau}) \quad (19)$$

where \mathbf{X}^* are new observations used for prediction and \mathbf{y}^* are the predicted values. The posterior conditional distributions from the Gibb's sampling routine can then be used as the priors for any subsequent models by approximating the numerical posterior results with a kernel estimator. Thus, in this way data from each model is used to add information into subsequent forms of the model. The hourly (minute) predictions are obtained by averaging the samples of the predictive distributions.

4. Data Collection and Model Implementation

4.1. Case study I: Vancouver, British Columbia, Canada

BC Hydro is the third largest utility in Canada and serves 95 percent of British Columbia's population. BC Hydro's total energy requirements, including losses and sales to other utilities and non-integrated areas, were 57,083 GWh in 2012. Hourly control area load data can be downloaded from the BC Hydro website [74]. In addition, temperature data is publicly available from the Environment Canada website [75]. Figures 15 and 16 represent the BC load area and temperature variations in 2012 and January 2012 respectively which shows the strong correlation between load and temperature.

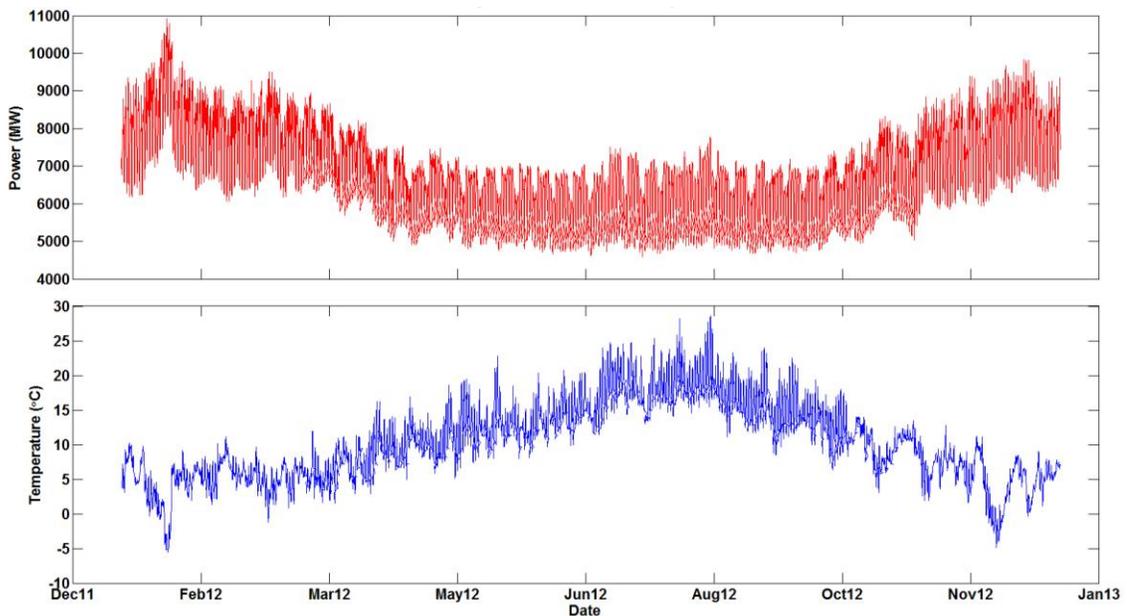


Figure 15. Load and temperature variation in 2012

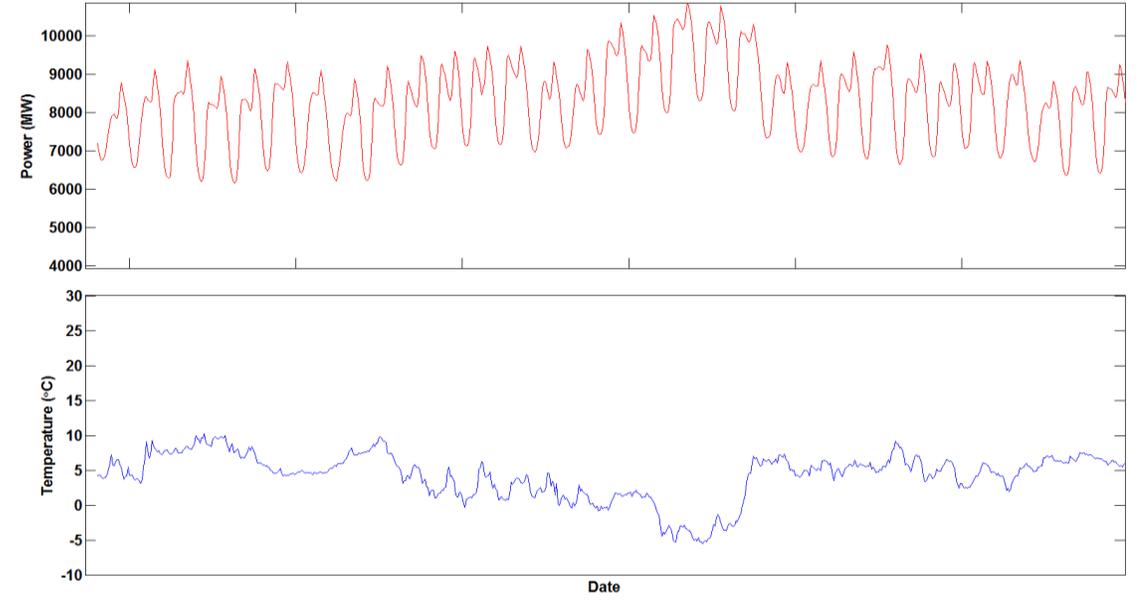


Figure 16. Load and temperature variation during January 2012, Vancouver BC

Note that the load data for the whole province is taken. Given the size of the province, temperature variation is expected and it is challenging to choose one particular station to correlate for the entire province. However, the VANCOUVER HARBOUR CS station was chosen for temperature recordings due to a large population in Lower Mainland around Vancouver, making it an important site to test this method.

A preliminary analysis of the hourly power consumption shows that the load is highly serially correlated with air temperature and 1-day lag load. This correlation was stronger than any other relationship explored (for e.g. 2-day lag load, previous day average load). Thus, for this research, a model was sought in the following form:

$$P(t) = f(P(t - 24), T(t)) \quad (20)$$

where $P(t)$ is the hourly power consumption at time t . $P(t - 24)$ is the previous day same hour load and $T(t)$ is the temperature at time t . As mentioned above, the Bayesian regression technique was explored to investigate the applicability of updating data-driven models for short-term model predictions. The following model was used:

$$P = \beta_0 + \beta_1 T + \beta_2 P_L \quad (21)$$

where P is the observed hourly power consumption, T is the hourly air temperature, P_L is the lagged load, and $\beta_0, \beta_1, \beta_2$ are the regression coefficients.

Typically, in a numerical modeling set-up, a portion of the available data is used for model training, and the remainder is used for model validation. For this research, a quasi-real-time model updating algorithm was used to train and validate the model. First of all, one year's worth of data (2012) was used to calibrate the regression model, while data from the following year (2013) was used to validate the model ($\mathcal{M}1$). Then, two years of data were used to calibrate the model (2012 and 2013) and the next year (2014) for validation ($\mathcal{M}2$). Then, three years of data were used to calibrate the model (2012, 2013 and 2014) and the first five months of the next year (Jan-May 2015) for validation ($\mathcal{M}3$). This structure is used to simulate a real-time recursive algorithm, where when more data is available, the model updates its parameters. By doing so, any changes in the system are implicitly captured by the data-driven model as each subsequent year is added to the dataset.

4.2. Case study II: Austin, Texas, USA

For the second case study, power signals recorded from residential houses in Austin, Texas by Pecan Street Inc. [76] were collected in which not only the household aggregate power but also the individual appliance power demands are monitored at 1-minute intervals. The installations began in January 2011. Each house aggregated power signal is generally a combination of various appliances in the house, such as EV, AC, furnace, dryer, oven, dishwasher, cloth-washer, refrigerator, microwave, bedroom lighting, and bathroom lighting, etc. One house was randomly chosen (ID: 22) to test the algorithm's performance in a real-world application.

Due to the high resolution of recorded signal (1-min), only one month worth of data (January 2015) was considered as shown in Figure 17. First of all, the first week was used to calibrate the regression model, while data from the second week was used to validate the model ($\mathcal{M}1$) using the non-informative prior model. Then, first two weeks of data were used to calibrate the model and the third week was used for validation ($\mathcal{M}2$). Finally, the first three weeks of data were used to calibrate the model and the last week of January 2015 was used for validation ($\mathcal{M}3$). These applications use the independent and informative priors.

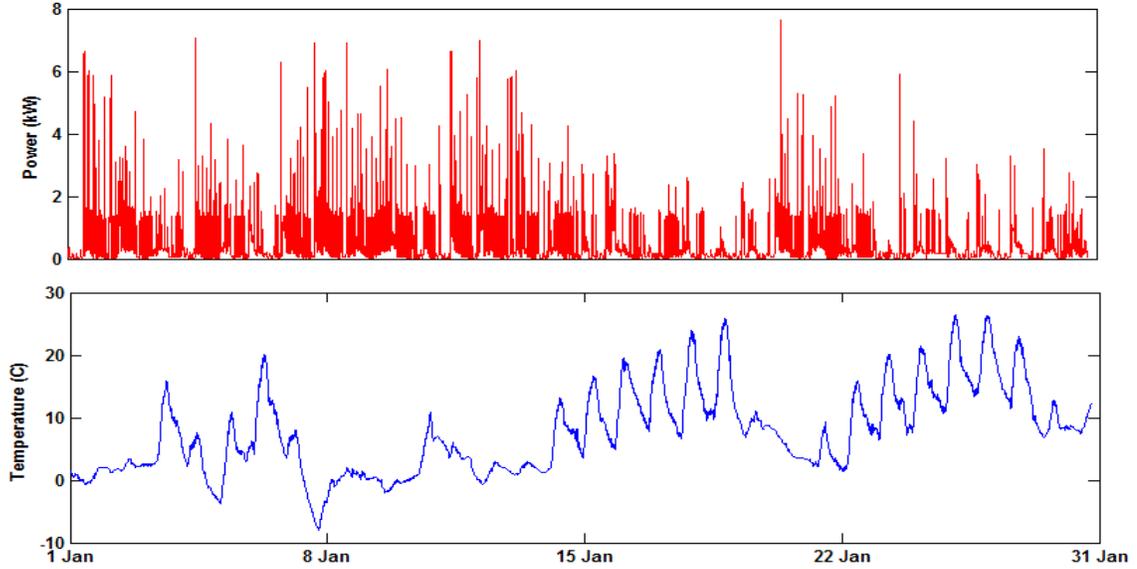


Figure 17. Load and temperature variation during January 2015, Austin TX

A model was sought in the following form:

$$P(t) = f(P(t - 1440), T(t), P(t - 1)) \quad (22)$$

where $P(t)$ is the minutely power consumption at time t . $P(t - 1440)$ is the previous day same minute load, $T(t)$ is the temperature at time t and $P(t - 1)$ is the load consumption one minute ago. The following regression model was used:

$$P = \gamma_0 + \gamma_1 P_L + \gamma_2 T + \gamma_3 P_L' \quad (23)$$

where P is the observed minutely power consumption, T is the minutely air temperature, P_L is the previous day same minute load P_L' is the previous minute load. $\gamma_0, \gamma_1, \gamma_2, \gamma_3$ are the regression coefficients. It is worth mentioning that the sub-hourly (5-minute) outside air temperature data are obtained from U.S. Climate Reference Network (USCRN) dataset [77]. Linear interpolation was then used to increase the original sampling rate of the sequence to a higher rate.

5. Simulation Results and Discussion

Three common model evaluation metrics were used to assess the adequacy of the results:

1. The CC measure between actual and predicted power loads indicates the strength and direction of a linear relationship between the forecasted and actual loads and calculated by:

$$CC_{yy^*} = \sqrt{1 - \frac{\sum_{i=1}^N (y_i - y_i^*)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}} \quad (24)$$

where y_i is the i th actual data, \bar{y} is the average of all actual data, y_i^* is the i th predicted data and N is the number of data points.

2. The Mean Absolute Percentage Error (MAPE) which has been traditionally used to measure accuracy in load forecasting. It captures the proportionality between the forecast error and the actual load. The MAPE is calculated by:

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - y_i^*}{y_i} \right| \times 100\% \quad (25)$$

3. The Root Mean Square Error (RMSE) which is used to evaluate the error (differences) between the forecasted and actual loads. The general form of the RMSE equation for the actual power loads (y_i) and the predicted ones (y_i^*) is given by:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (y_i - y_i^*)^2}{N}} \quad (26)$$

Residual sum of squares (RSS) and RMSE are two common measures of error regression and RMSE is simply the square root of the mean RSS. However, RMSE can be more intuitive than RSS, since its units are the same as that of the target prediction. In our case, the unit is MW or kW and does not grow with the number of data points like the RSS does.

Figures 18 and 19 depict the histogram of regression coefficients for model $\mathcal{M}1$ for both case studies respectively, along with the estimated parameters using Ordinary Least Squares (OLS) shown by the red vertical line. Note that the result of OLR method is simply the mean value of the Bayesian approach.

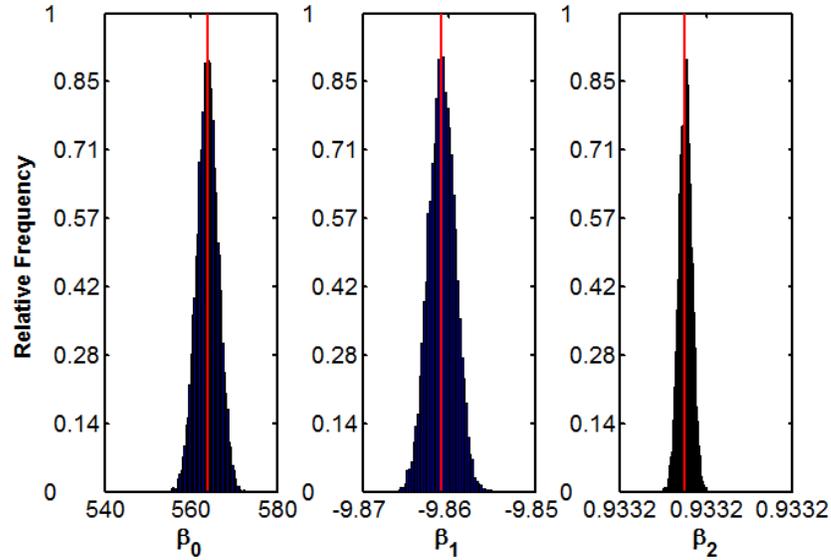


Figure 18. Histograms of BLR coefficients for BC area baseline for model **1** . Vertical line is the ordinary least square results

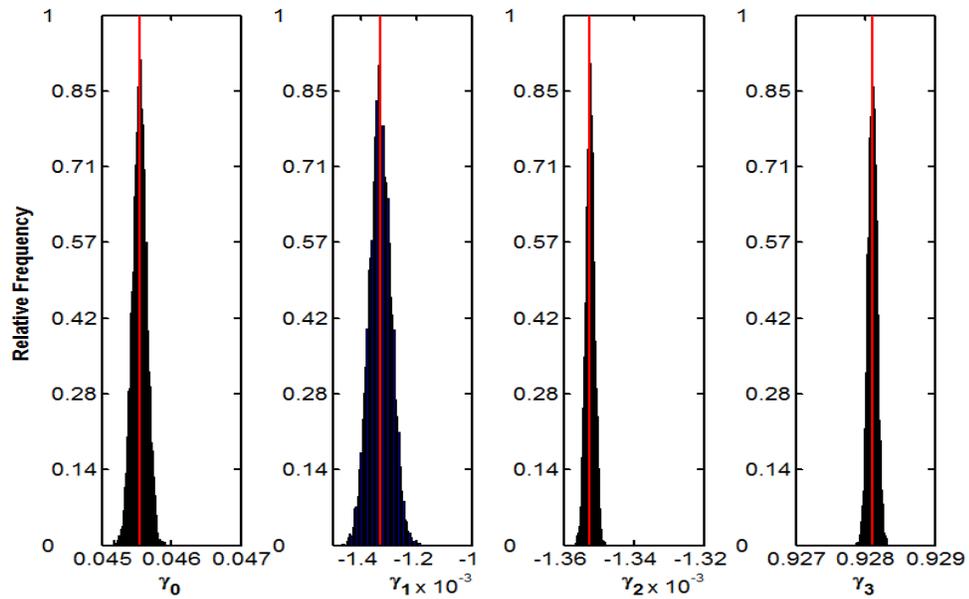


Figure 19. Histograms of BLR coefficients for Pecan St. household baseline for model $\mathcal{M}1$. Vertical line is the ordinary least square results

Figure 20. shows the change in the approximate pdf of the three regression coefficients and the variance for the $\mathcal{M}1, \mathcal{M}2, \mathcal{M}3$ models for BC area baseline.

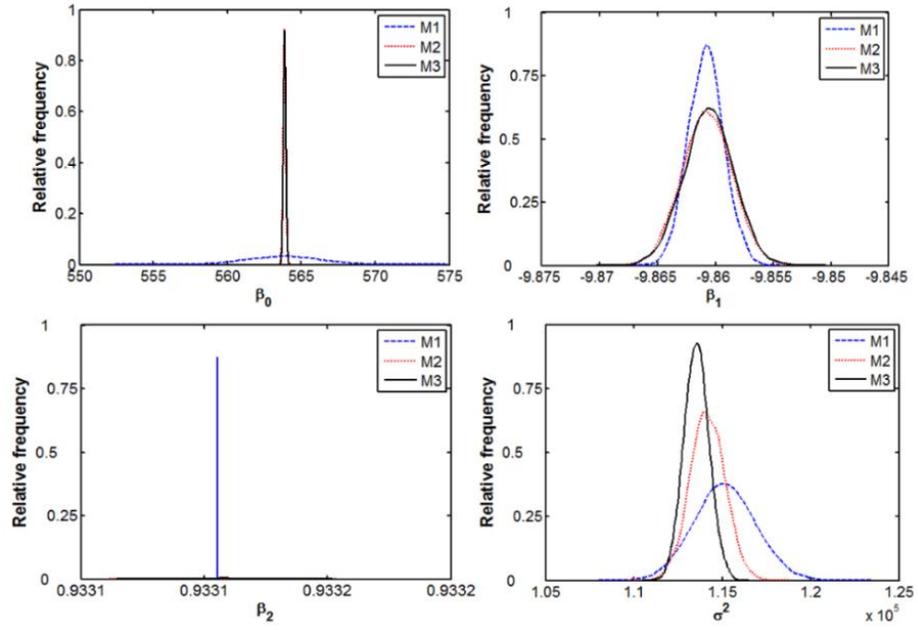


Figure 20. Approximate pdf of the variables of three models for BC area baseline

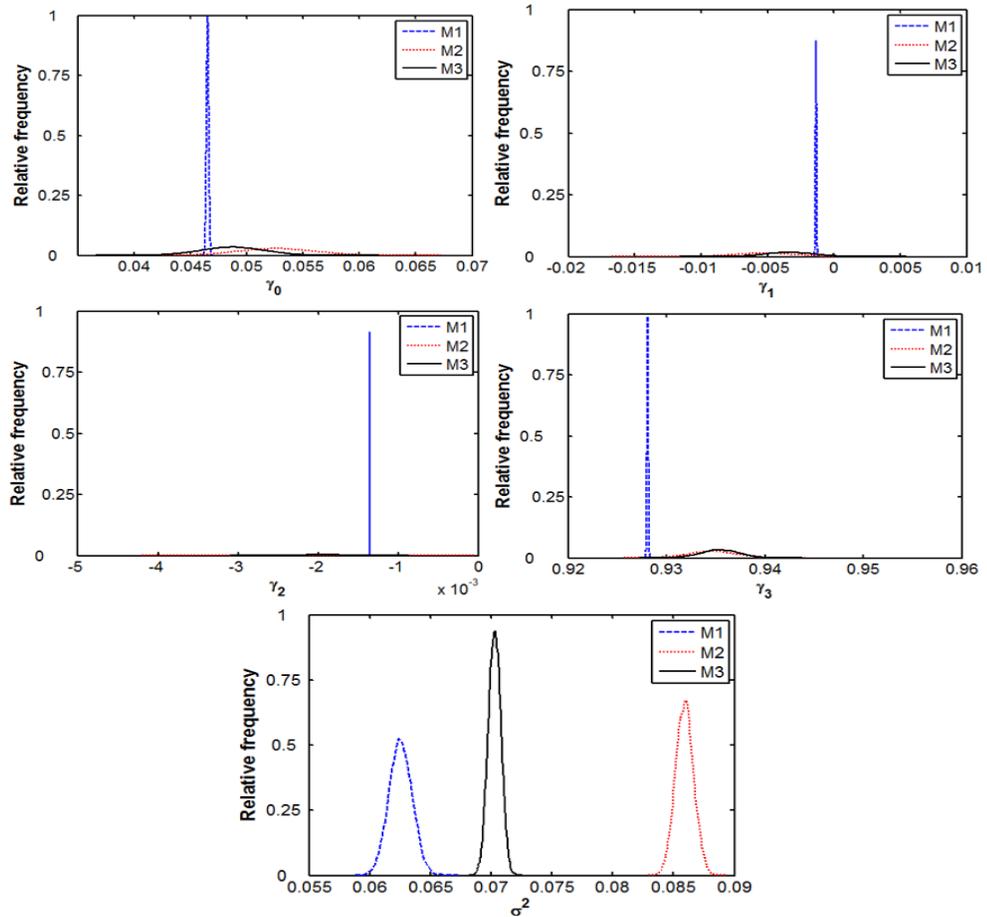


Figure 21. Approximate pdf of the variables of three models for Pecan St. household baseline

The value of σ^2 (i.e. $1/\tau$) decreases with each subsequent year of added data and as more data is added the spread of the variance decreases which clearly demonstrates the functionality of the Bayesian approach for data analysis. Similar results are shown in Figure 21 for Pecan St. Household data.

Figure 22 shows selected results (first week of January) for the validation data for model $\mathcal{M}1$ (2013 data used for validation), $\mathcal{M}2$ (2014 data used for validation) and $\mathcal{M}3$ (2015 data used for validation). Each subplot shows the trend of the observed and means predicted hourly power consumption along with the 10th and 90th percentile of predicted hourly load. For all cases, the observed load generally falls within the predicted interval. Note the advantage of the Bayesian approach in predicting an interval of the load at a given confidence level versus simple standard mean predictions from basic regression models.

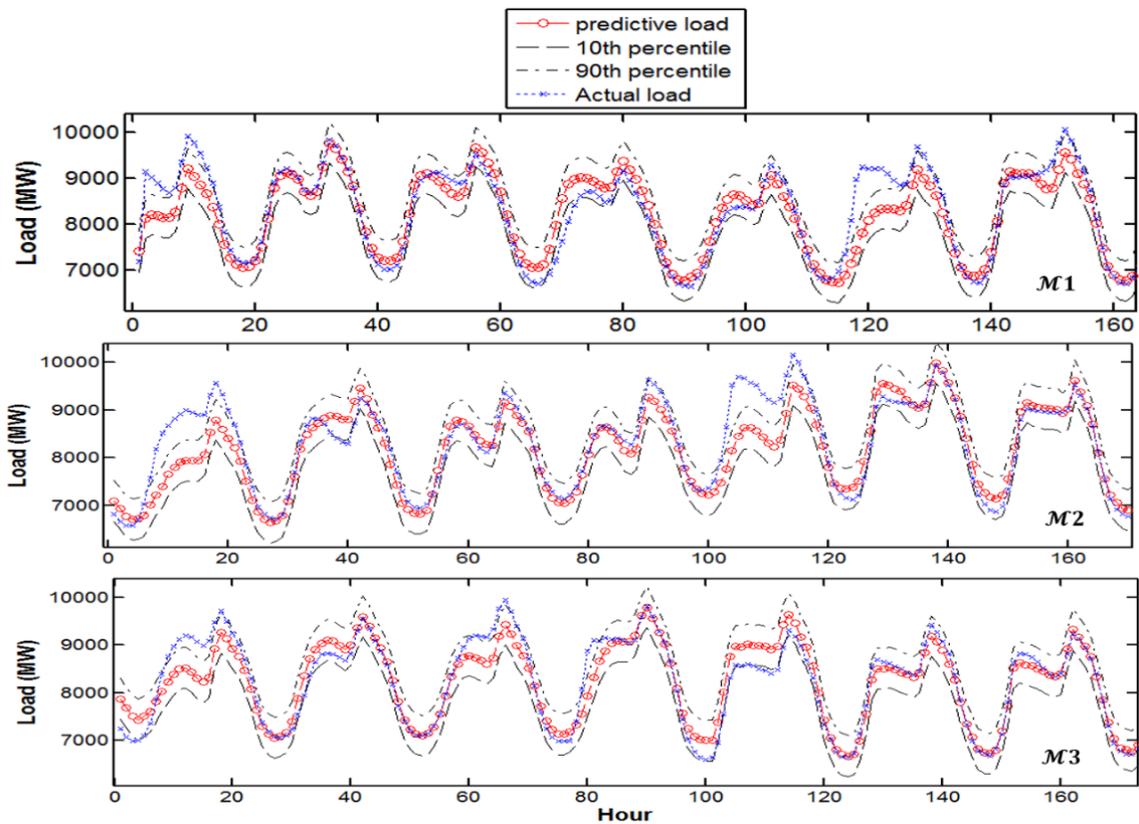


Figure 22. Trend plots for validation results of $\mathcal{M}1$, $\mathcal{M}2$ and $\mathcal{M}3$ for the first week of January 2013, 2014 and 2015 respectively for BC area baseline

Figure 23 shows selected results of the validation data for a single day for model $\mathcal{M}1$ (Jan. 8th used for validation), $\mathcal{M}2$ (Jan. 15th used for validation) and $\mathcal{M}3$ (Jan. 22nd used

for validation). Each subplot shows the trend of the observed and means predicted hourly power consumption along with the 10th and 90th percentile of predicted hourly load. For all cases, the observed load falls within the predicted interval. The figure shows that the model can capture the observed trend with only one week's worth of data.

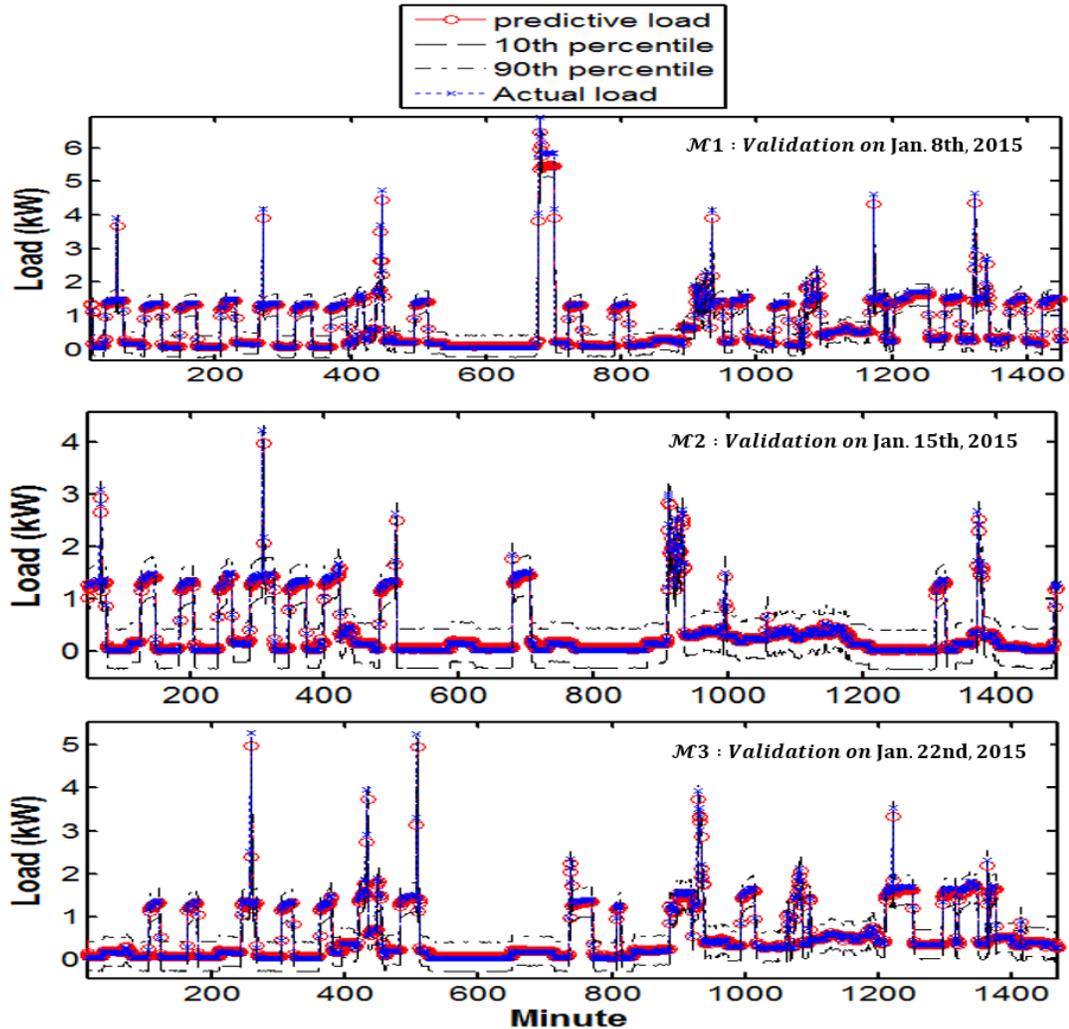


Figure 23. Trend plots for validation results of $\mathcal{M}1$, $\mathcal{M}2$ and $\mathcal{M}3$ for Pecan St. Household

Generally speaking, the size of the BLR interval is larger for $\mathcal{M}1$, reflecting the relative "lack of knowledge" and impact of selecting a non-informative prior for the first model while for $\mathcal{M}2$ and $\mathcal{M}3$ this interval decreases as more data is added.

Recall that the Bayesian approach returns a distribution to describe the unknowns. Hence, we can see the uncertainty in our estimates. The histogram of predicted load consumption is shown in Figure 24 and Figure 25 for the January 1st, 2013 and 2014

respectively along with the actual load. The closer the actual load to the mean value of the histogram the better the forecast which is not the case in some hours. It is worth mentioning that applying the Bayesian approach does not necessarily improve the forecast for a particular time as the regression features are the main drivers for predicted values for each model. Specifically, the forecast accuracy improves for the second hour of January 1st in $\mathcal{M}2$ compared to $\mathcal{M}1$ while it declines for the 16th hour. However, this result can greatly inform and improve a load aggregator's risk averse bidding strategy.

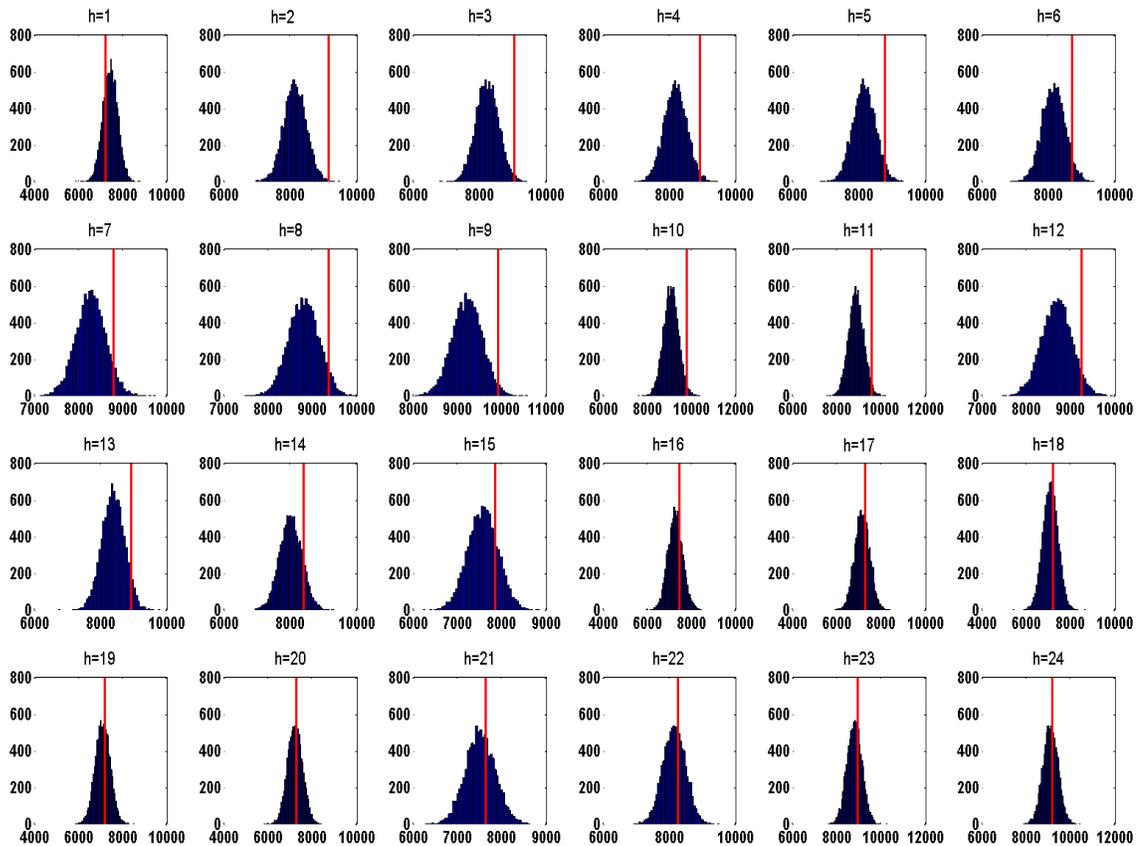


Figure 24. Histogram for validation results of $\mathcal{M}1$ on Jan. 1st 2013 for each hour of the day. Actual load is shown by red line

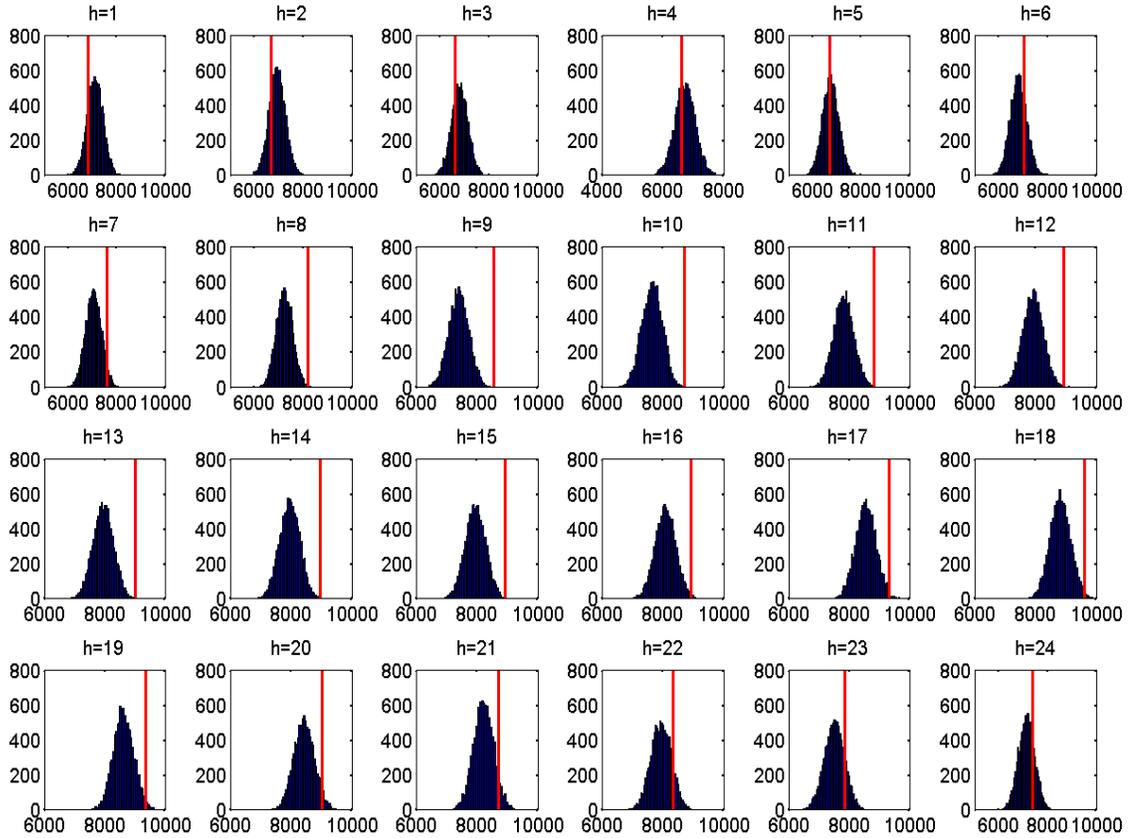


Figure 25. Histogram for validation results of $\mathcal{M}2$ on Jan. 1st 2014 for each hour of the day. Actual load shown by red line

The performance metrics calculated at the mean value of the samples in the BLR interval are summarized for each model in Table 9 and 10 for both case studies. As more data is added, the performance improves in all cases which confirm the benefits of the Bayesian approach in updating the baseline analysis. The MAPE values are consistent with the results previously reported in [78] but it does not mean that accuracy improves with the larger mean load. It depends on which performance metric is picked to discuss the results. For the first case study (total BC Load), the MAPE and RMSE improve 2.7 % and 0.8% respectively from $\mathcal{M}1$ to $\mathcal{M}3$ for the validation dataset.

Note that the results of OLR are simply the same as the result of the non-informative prior model; hence, one can see how recursive BLR outperforms OLR technique.

Table 9. Performance metrics for BC area aggregate baseline

<i>BC Area Aggregate Baseline</i>	<i>CC measure</i>	<i>MAPE</i>	<i>RMSE[MW]</i>
<i>Non-informative prior Training Result (M1)</i>	<i>0.956</i>	<i>3.569%</i>	<i>339.25</i>
<i>Non-informative prior Validation Result (M1)</i>	<i>0.959</i>	<i>3.384%</i>	<i>334.97</i>
<i>Informative prior Training Result (M2)</i>	<i>0.958</i>	<i>3.476%</i>	<i>337.15</i>
<i>Informative prior Validation Result (M2)</i>	<i>0.957</i>	<i>3.304%</i>	<i>333.34</i>
<i>2nd Informative prior Training Result (M3)</i>	<i>0.958</i>	<i>3.42%</i>	<i>335.99</i>
<i>2nd Informative prior Validation Result (M3)</i>	<i>0.938</i>	<i>3.293%</i>	<i>332.1</i>

For the second case study (Pecan St.), the MAPE and RMSE improve 32 % and 53% respectively from $\mathcal{M}1$ to $\mathcal{M}3$ for the validation dataset which means the accuracy improves more on lower aggregation level in a Bayesian framework as it is clear from RMSE values.

Table 10. Performance metrics for Pecan St. Household baseline

<i>Pecan St. Household Baseline</i>	<i>CC measure</i>	<i>MAPE</i>	<i>RMSE[kW]</i>
<i>Non-informative prior Training Result (M1)</i>	<i>0.93</i>	<i>40.24%</i>	<i>0.25</i>
<i>Non-informative prior Validation Result (M1)</i>	<i>0.942</i>	<i>43.11%</i>	<i>0.336</i>
<i>Informative prior Training Result (M2)</i>	<i>0.937</i>	<i>43.22%</i>	<i>0.311</i>
<i>Informative prior Validation Result (M2)</i>	<i>0.94</i>	<i>34.1%</i>	<i>0.209</i>
<i>2nd Informative prior Training Result (M3)</i>	<i>0.939</i>	<i>39.18%</i>	<i>0.269</i>
<i>2nd Informative prior Validation Result (M3)</i>	<i>0.921</i>	<i>29.1%</i>	<i>0.155</i>

6. Case Study III

To further investigate this method, BLR was applied on another case study of forecasting water consumption rate of an output storage tank at a municipal water treatment facility in U.S. For confidentiality reasons geographical location and facility name are not mentioned. An aggregator regulates the operation of two 700 Hp variable frequency drive (VFD) pumps in parallel with three binary pumps at this facility to participate in PJM regulation market.

Following the storage model is shown in Figure 3, for each type of device, there is an associated input or output storage, or both. For example, a pump device may pump water out of a reservoir, into a reservoir, or between two reservoirs. The model is constructed such that when the asset is running it will decrease the storage level of the input reservoir, and increase the storage level of the output reservoir. Therefore, varying the electrical demand of the device will change storage levels relative to the upper and lower limits of storage imposed by the site. For example, upper and lower limits of storage for a pump would be reservoir water levels.

The rate at which the device increases or decreases storage is known as the *Asset Rate (AR)*, and varies with the electrical demand of the device. The asset rate affects both input and output storages the same. Therefore, units of the input and output storage must match e.g. two different sized reservoirs could not have storage levels measured by storage tank height, but could have storage measured by volume of water.

There are often external processes or devices that run in parallel with the controlled device. These processes or devices may also cause input or output storage levels to change. The rate at which all external processes and not- controlled devices (combined) affect storage levels is known as the *Process Rate (PR)*. There is a process rate for input storage, and a separate process rate for output storage. Process rates are often not known directly and have to be inferred from measurements of storage level and asset rates.

Aggregator needs to predict when assets will be available to participate in the network, and also how much regulation they will be able to provide. It calculates the PR and the upper/lower AR for the bid period in four steps:

1. calculate the PR affecting storage over the last hour

2. Average the calculated PR and assumes the averaged PR will remain constant until the end of the bid period. This is the forecast that aggregator currently uses (persistent model).

3. Using the averaged PR and the current storage level, calculate upper and lower AR that, if held constant, would keep storage levels within storage limits until the end of the bid period

4. The upper and lower calculated AR are compared against the upper and lower AR limits for the asset, and the available range of the asset is calculated for the bid period (asset rate are proportional to the upper and lower control limits of the asset)

6.1. Data Cleaning and Filtering

First of all, a reliable historical PR for the output storage tank which basically translates to the rate of water consumption and is independent of flow through the pumps is required. However, there are some limitations of available data that makes the data preprocessing a crucial first step:

- **Uncontrolled Pumps:** Asset rate stored in the site Database, exclude the flow through the uncontrolled binary pumps, (Aggregator control 2 VFD pumps only. Although the information on which Binary pumps operate is obtainable)
- **Changing Storage Limits:** The physical storage levels within which we are allowed to operate change occasionally. This implies that the process rate needs to be calculated in terms of raw values rather than micro %. (Confirm which tank - Input or Output Storage tanks or both)
- **Flatlines:** Storage levels for output storage tank occasionally appears as a flat line, (a constant value appears for a period of time). Attributed to hardware malfunction. The longest period with almost acceptable values in 2015 is found to be from May-01-2015 until the end of August 2015. Data was sampled by average every 5 minutes.
 - •*Training set 2015-5-1 00:05:00 --- 2015-08-28 24:00:00 (120 days, 34560 data points)*
 - •*Test set 2015-08-29 00:05:00 --- 2015-08-29 24:00:00 (1 day, 288 data points)*

Asset rate can be found via appropriate linear transformation from asset electrical demand.

```
OldRange = (AssetUpperControl-AssetLowerControl);
```

```
NewRange = (AssetUpperRate-AssetLowerRate);
```

```
AR = (((AssetDemand-AssetLowerControl) .*NewRange) ./OldRange)+AssetLowerRate;
```

The mean of $(NewRange ./ OldRange)$ array is calculated as 3.5×10^5 [lit/5minMW]. The scatter plot of Asset rate vs. Electrical demand shows separate straight lines corresponds to different operational status as shown in Figure 26.

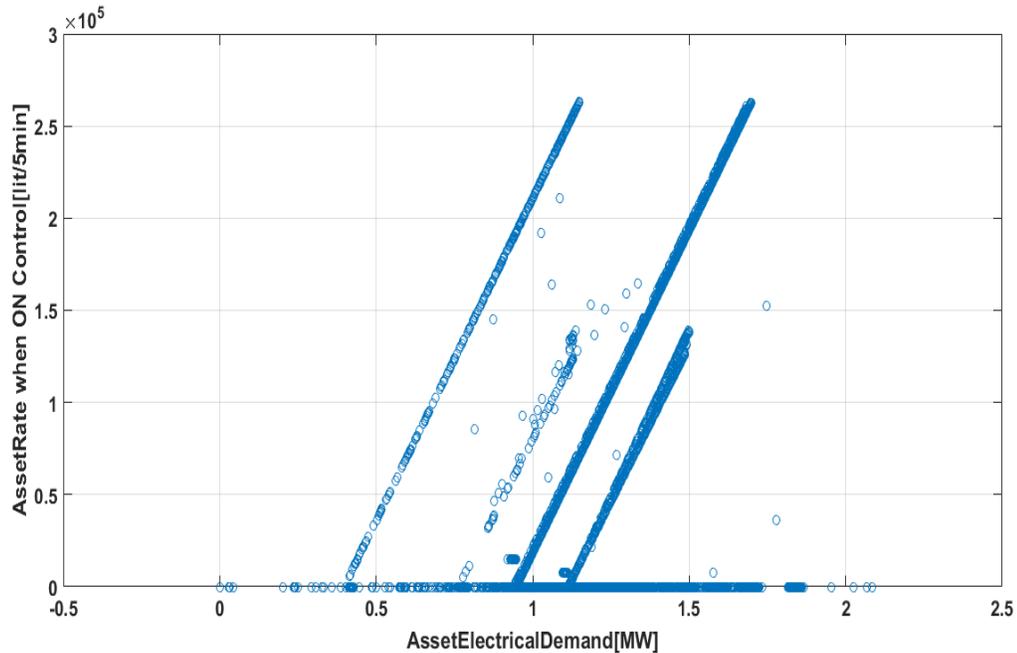


Figure 26. Asset rate vs. asset electrical demand

- **Binary Pumps Switching Transients:** When a binary pump switches ON/OFF, a momentary high variation in output storage rate of change is observed which directly affect the output process rate. For appropriate modeling the data at switching instances should be discarded.
- **Unknown spikes in storage:** Smoothing is how we discover important patterns in our data while leaving out things that are unimportant (i.e. noise). Median filtering

(replaces every point of the signal by the median of that point and a specified number of neighboring points) is used to perform this smoothing. The goal of smoothing is to produce slow changes in value so that it is easier to see trends in our data.

6.2. Output Process Rate Data Modeling

The final output PR (after filtering and interpolation) is calculated as the difference between output storage tank level rate of change and asset rate and shown in Figure 4. A closer look at the dataset reveals a clear diurnal pattern which intuitively makes sense as depicted in Figure 27. Same as electricity usage, water consumption follows a similar trend each day. It starts decreasing after midnight, as people are sleeping and businesses are closed down. After 6-7 AM it starts increasing as people wake up and the business day starts until reaches to the maximum around 6-7 PM when people get back home again.

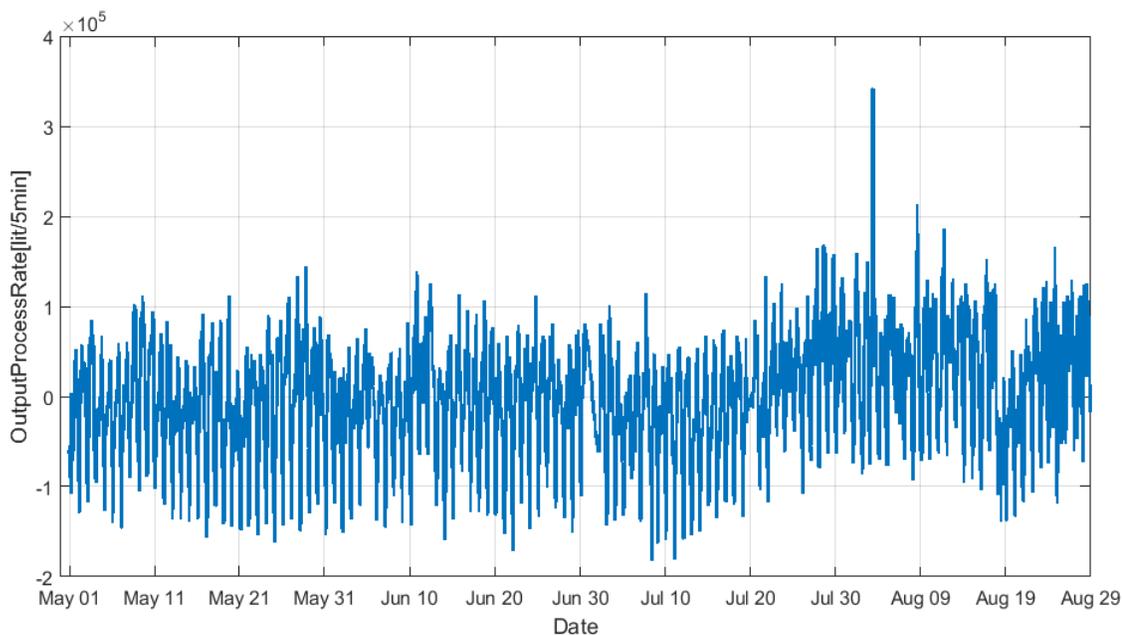


Figure 27. Output process rate at Shire Oaks May-Aug 2015

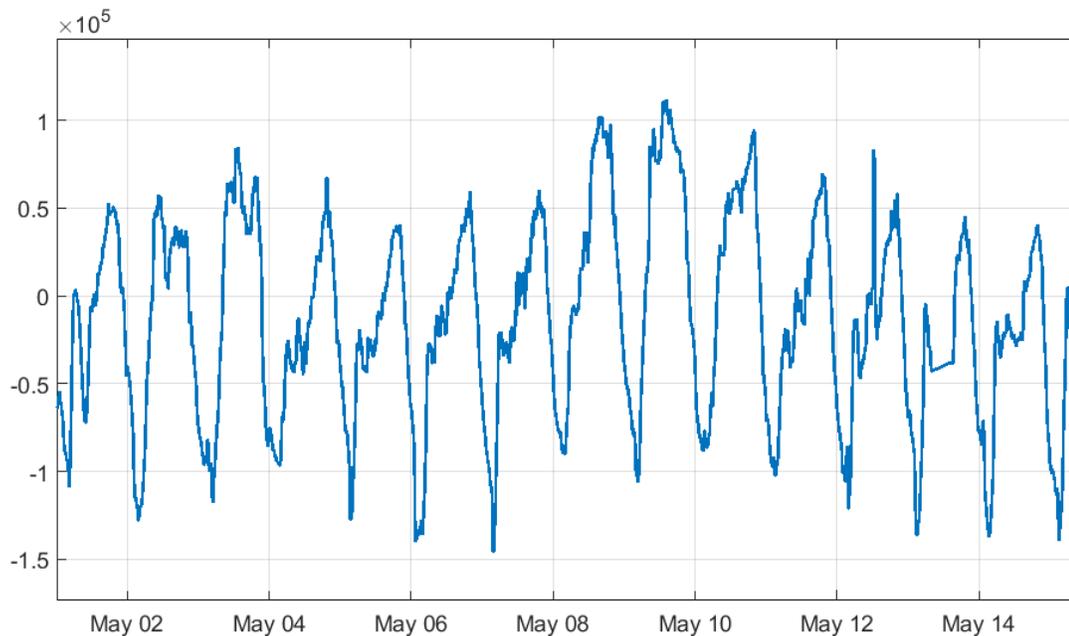


Figure 28. A closer look at the output process rate at Shire Oaks 1-14 May 2015

Note that each day contains 288 data points (every 5-min sample). Moreover, a significant difference between weekdays and weekends was not observed. Using a 288-point moving average filter, an increase in water consumption is observed as outside air gets warmer shown in Figure 29.

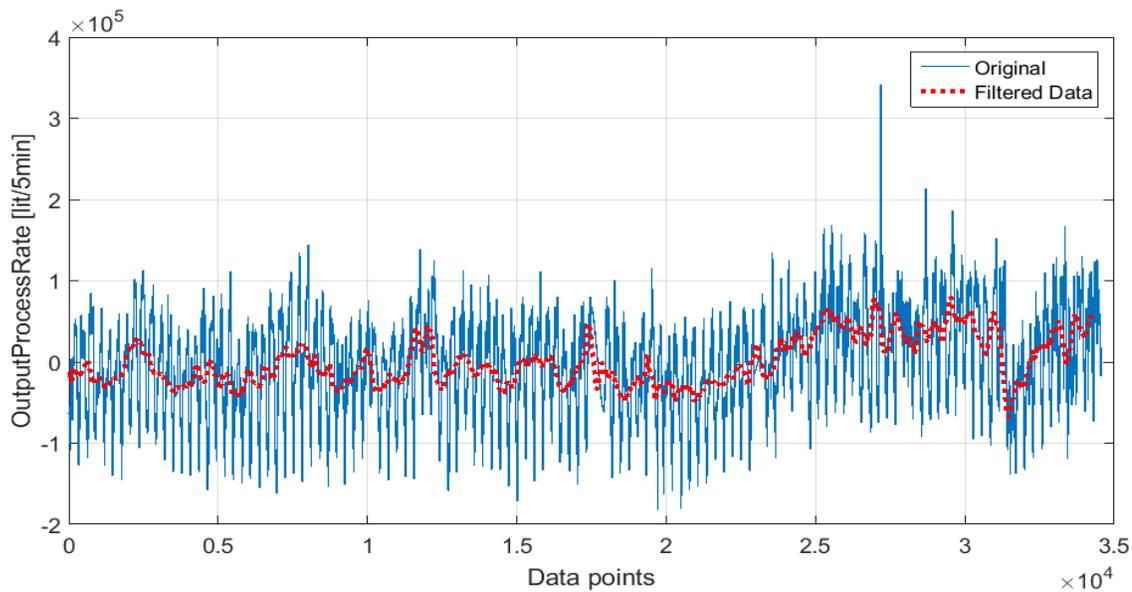


Figure 29. Filtered output process rate

To have a better understanding of the diurnal variation, we tried to fit different Fourier models to the average daily process rate. Figures 30-32 show the average PR along with the Fourier fit with one, two and three terms. Clearly, a Fourier model with two terms (27) shown in Figure 31 serves the purpose without too much complexity.

$$f(x) = a_0 + a_1 * \cos\left(x * \frac{2\pi}{288}\right) + b_1 * \sin\left(x * \frac{2\pi}{288}\right) + a_2 * \cos\left(2 * x * \frac{2\pi}{288}\right) + b_2 * \sin\left(2 * x * \frac{2\pi}{288}\right) \quad (27)$$

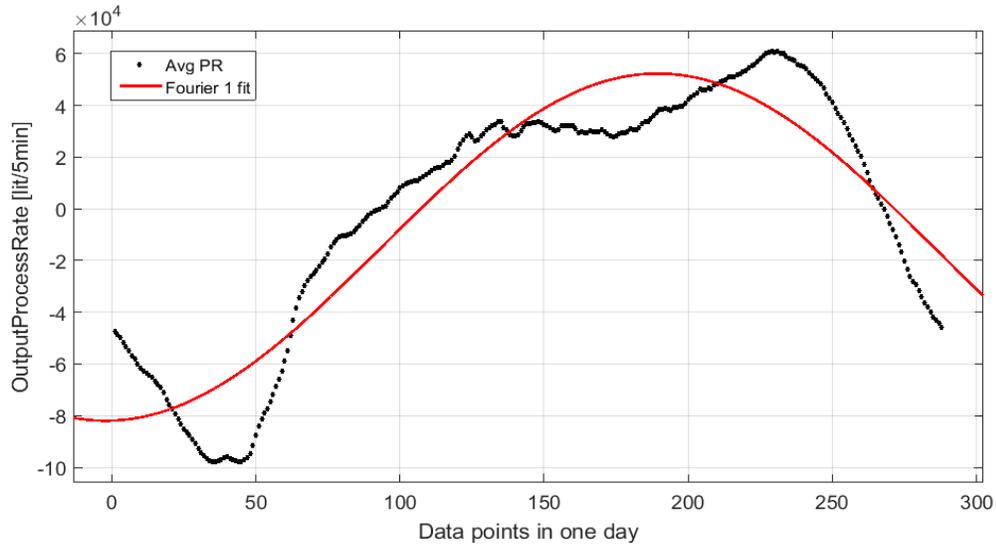


Figure 30. Average daily PR and a Fourier fit with one term

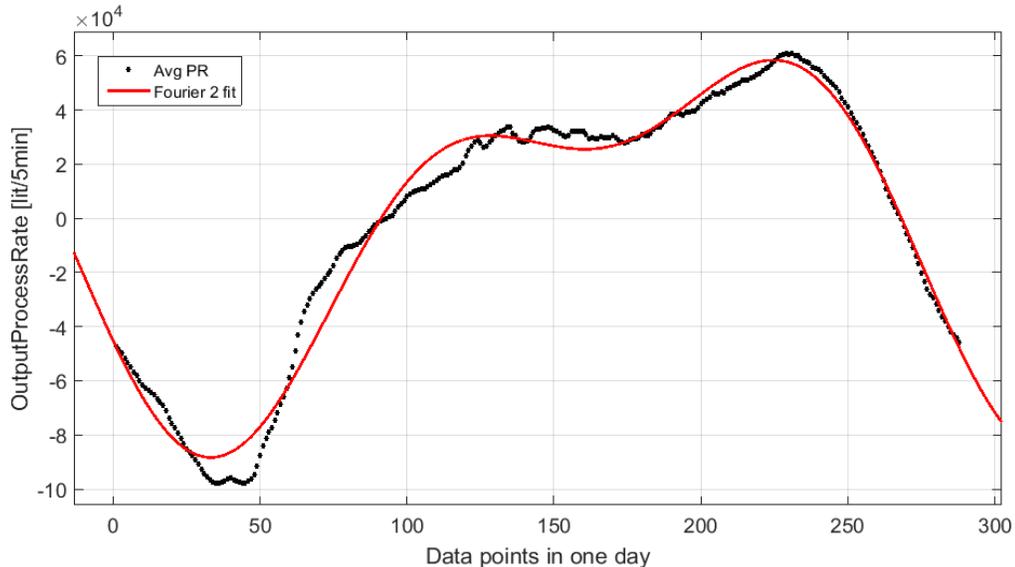


Figure 31. Average daily PR and a Fourier fit with two terms

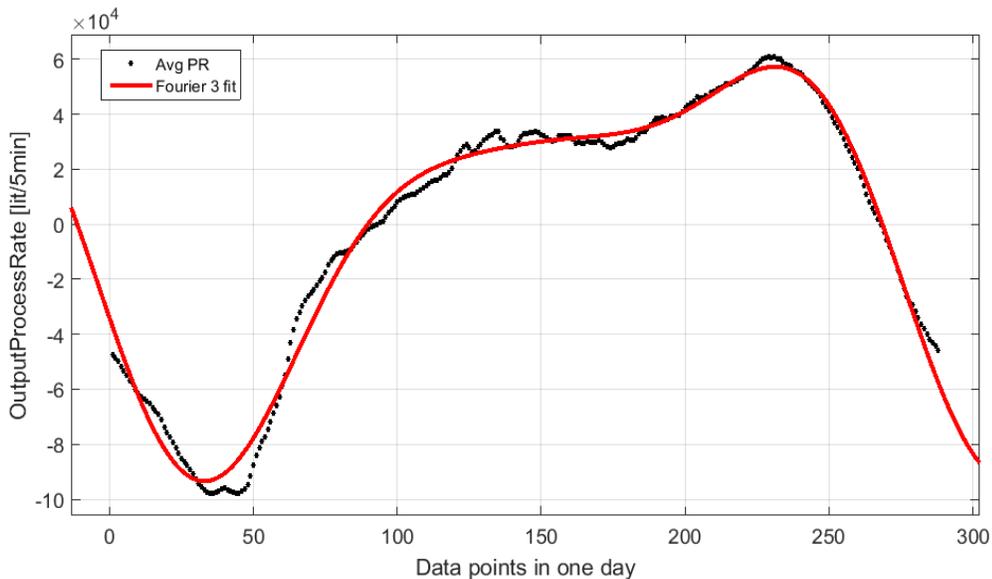


Figure 32. Average daily PR and a Fourier fit with three terms

Diurnal variation modeling is necessary but it is not enough as shown in Figure 33 since it fails at capturing the trend. Therefore, we need to consider other features to have a decent prediction.

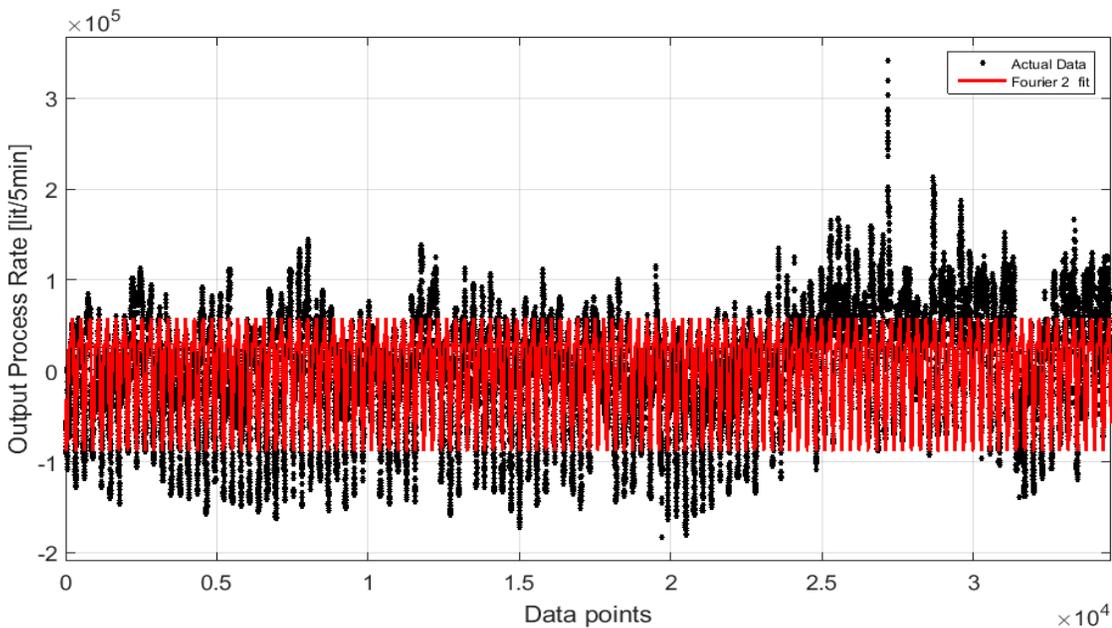


Figure 33. Fourier series curve fitting

6.3. Process Rate Forecast (1-hr ahead)

For the hour-ahead forecast, the following predictor variables were selected:

- Previous hour average process rate
- Previous hour process rate
- Previous day process rate
- Temperature
- Precipitation

Mathematically the model is expressed as:

$$y \sim w_0 + w_1.x_1 + w_2.x_2 + w_3.x_3 + w_4.x_4 + w_5.x_5 + w_6.x_6 + w_7.x_7 + w_8.x_8 + w_9.x_9$$

where

x1: Prev. hour PR

x2: Prev. hour avg PR

x3: Prev. day PR

*x4: $\cos((2 * \pi/288) * t)$*

*x5: $\sin((2 * \pi/288) * t)$*

*x6: $\cos((4 * \pi/288) * t)$*

*x7: $\sin((4 * \pi/288) * t)$*

x8: Temperature

x9: Precipitation

Initial results show that all variables are statistically significant:

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	6759.4	602.55	11.218	3.8006e-29
x1	-0.45672	0.0043328	-105.41	0
x2	1.4077	0.0045314	310.66	0
x3	0.024583	0.0016336	15.048	5.4389e-51
x4	-3037.9	115.99	-26.191	1.5343e-149
x5	1108.4	122.57	9.0425	1.6201e-19
x6	-3326.8	93.179	-35.703	1.6229e-273
x7	435.28	99.477	4.3757	1.2142e-05
x8	73.378	15.767	4.6538	3.2719e-06
x9	-1583.4	720.47	-2.1978	0.027974

However, the Lasso regression result shows that temperature and precipitation do not affect the process rate.

The final forecast of Bayesian Linear Regression model is shown in Figure 34.

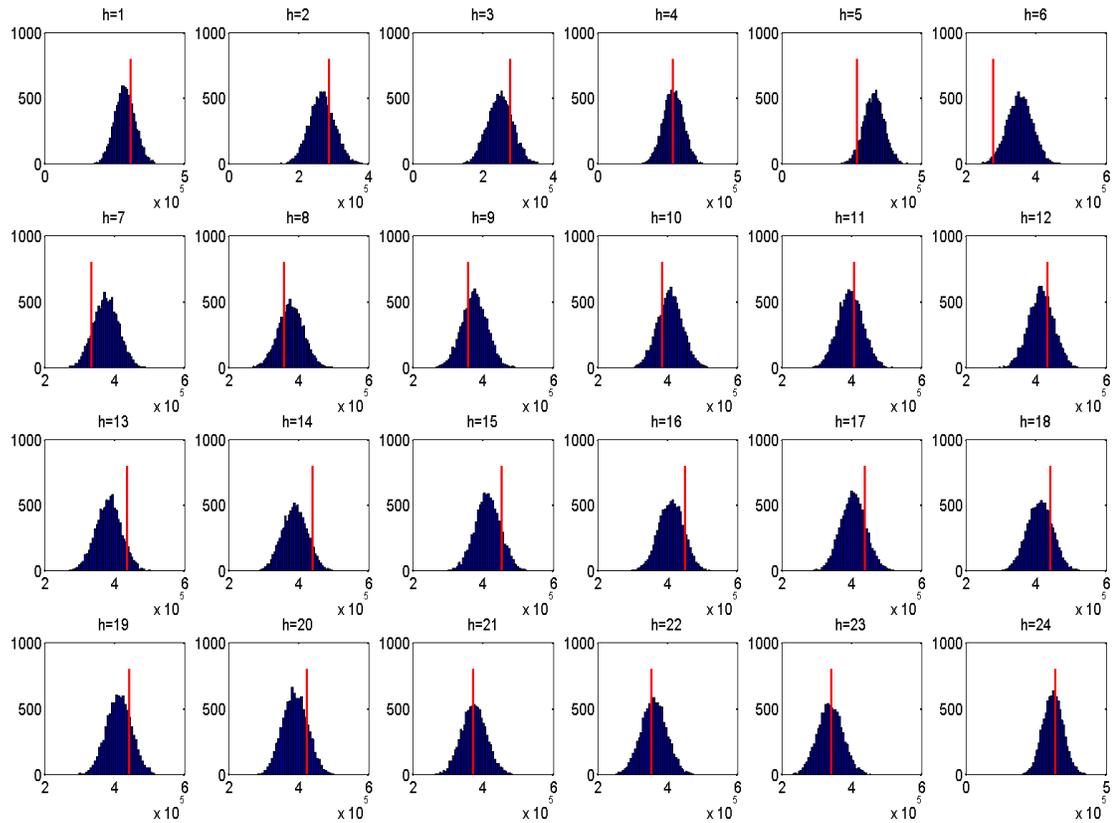


Figure 34. Day-ahead output process rate forecast, red line represents the actual value

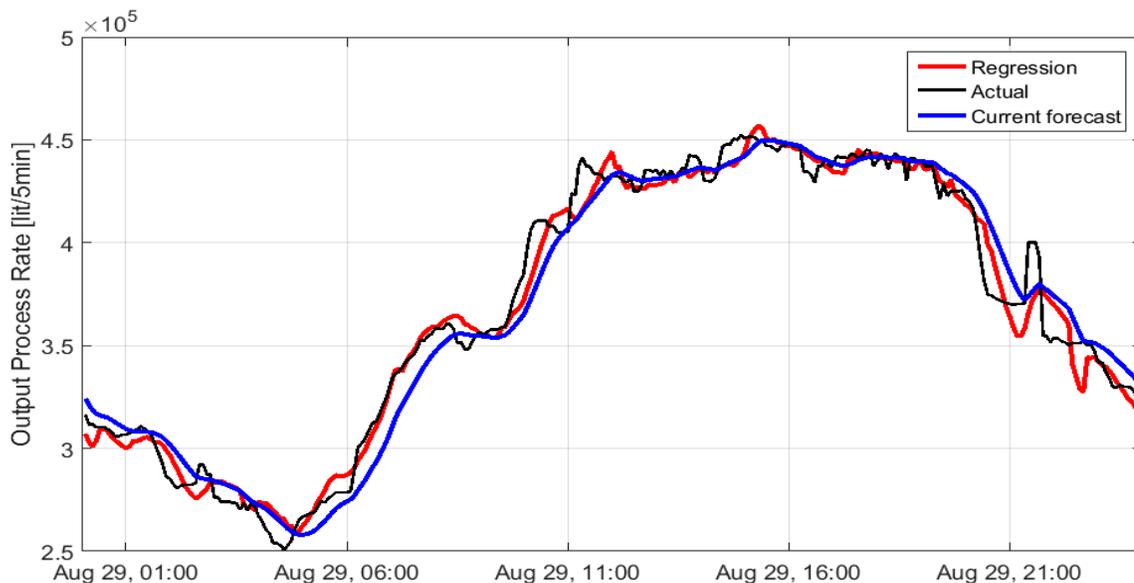


Figure 35. Current process rate calculation vs. Regression (update every 5 min)

Taking the mean value as the output of BLR, the Regression forecast on test set improves 24% compare to the persistent model as shown in Figure 35. Mean Absolute Error (MAE) of regression is 0.019 whereas persistent model MAE is 0.025. (Typically update every 5 min)

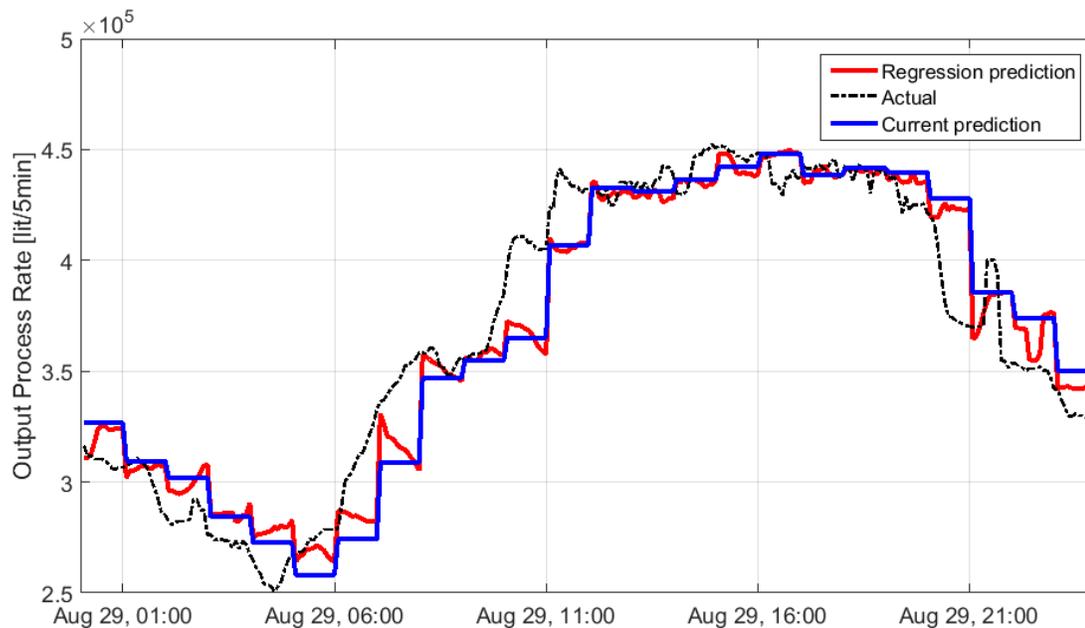


Figure 36. Current process rate calculation vs. Regression (1-hr commitment)

Aggregator calculates the average PR over the previous hour and assumes it will remain constant until the end of the bid period (typically ~ 1 -hr commitment). In this case, the MAE decreases by 15% as depicted in Figure 36.

Residual analysis on training data highlights that regression outperforms the persistent model shown in Figure 37.

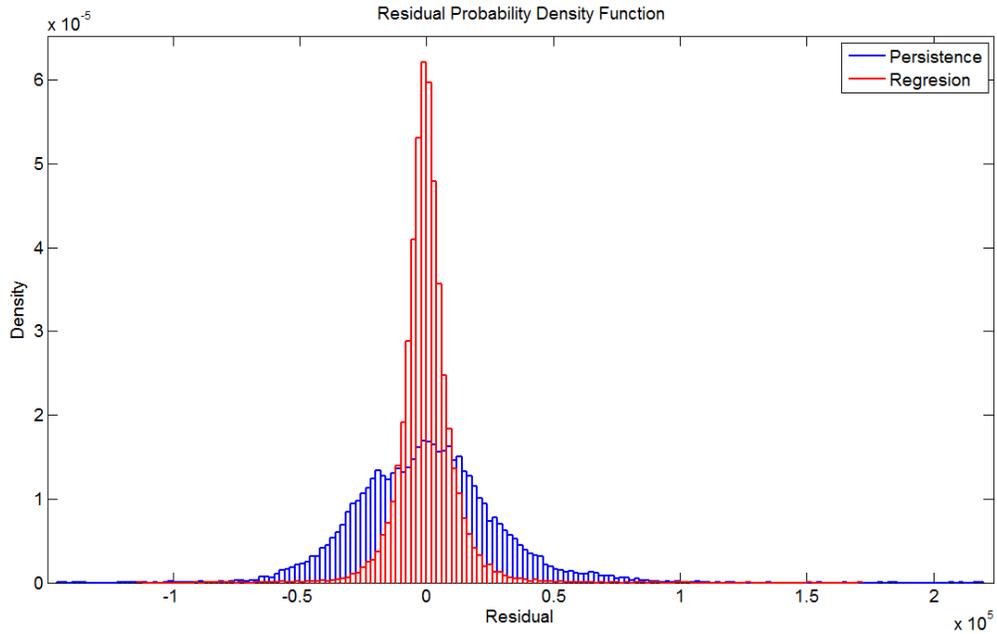


Figure 37. Residual probability density function for regression and persistent model

7. Summary

This paper introduces the idea of using a recursive Bayesian linear regression approach for baseline load analysis in a smart grid framework where real-time model prediction is of importance for operators and market participants.

After defining the principle regression equation; for simulation purposes, the available data was broken into four smaller data sets. A quasi-real-time Bayesian updating algorithm was used to train and validate the model. The first batch of data was used to calibrate the regression model, while data from the second batch was used to validate the model ($\mathcal{M}1$). Then, first and second of data were used to calibrate the model and the third batch for validation ($\mathcal{M}2$). Then, first, three batches of data were used to calibrate the model and the

last batch of data for validation(\mathcal{M}_3). This structure is used to simulate a real-time recursive algorithm, where when more data is available, the model updates its parameters. By doing so, any changes in the system are implicitly captured by the data-driven model as each subsequent year is added to the dataset.

The results show that BLR approach provides a very good forecast while including forecast uncertainty inherently. As more data is added, the performance improves in all cases which illustrate the effectiveness of the Bayesian method and the benefits of implementing an online learning algorithm for baseline analysis.

Chapter 5: On-line Baseline Load Learning Using Kernel Adaptive Filtering

In this chapter, we continue the topic of baseline load forecasting using well-established techniques in signal processing. Specifically, due to the advances in Smart Grid technologies and other reasons mentioned in the previous chapter, we are interested in deploying an online algorithm, i.e., learn from a sequence of signal samples. An adaptive filter is a filter that adapts its transfer function to signal property variation over time by minimizing an error or loss function that represents how far the filter response deviates from ideal behavior and seeks optimal models for the time series. Adaptive filtering is also a regression in functional spaces, where the user controls the size of the space by choosing the model order. In this context, filtering is the process of sequentially estimating the states or parameters of the dynamic system. In contrast with the Bayesian Regression technique, the filtering approach does not require definition of a principle equation to start with or having historical data to train the model first. Furthermore, Bayesian learning becomes complicated due to the sequential Monte Carlo iterations. The materials in the following introductory section summarizing machine learning are informed by the excellent overview presented by Liu et al [79].

1. What is Learning? Why Online Learning?

Our primary interest is to study a method that facilitates our *learning from data*. The term *data* is taken to mean measurements that come from the real world. Suppose we observe a data set \mathcal{D} , which has an explicit and causal input-output structure. We want to utilize the data to learn the “*governing law*” (some authors refer to this as the *underlining process*, *target function*, *target distribution*, *target operator*, or *supervisor’s operator*) [80] that is responsible for how the data was produced. Such a desire to learn is quite natural, because understanding this governing law would greatly facilitate reliable prediction of what will occur next or, stated in more formally, facilitate us to infer something outside \mathcal{D} . As a matter of fact, to infer what will occur outside \mathcal{D} is the central goal of every learning process. At a glance, learning the governing law from a given set of data appears to be a hard, if not impossible mission.

Historically, the basic assumption in the machine learning framework is that we have collected a set of observations which we called *training data*, and the purpose of learning is to uncover the governing law that is responsible for the production of the data observed. It is constructed a priori and learning stops when we fully process the data set. Depending on the structure of the training data, machine learning methods may be categorized as *supervised* and *unsupervised* learning. Despite its wide applicability, there are two issues associated with it:

- Application conditions may be non-stationary, i.e. the model must be continuously adapting to track changes in the underlying system.
- In many important applications, data arrives in real time, one sample at a time, so on-line learning methods are necessary. This is called *sequential learning*.

One famous form of sequential learning is *Bayesian learning* as introduced in the previous chapter. It has a close connection to the Kalman Filter in adaptive filtering theory but is complicated for arbitrary data distributions due to the sequential Monte Carlo methods required.

2. Linear Adaptive Filters

Filtering is a regression in a functional space (time series). The adaptive capability relies on *error-correction learning*. Consider the filtering structure depicted in Figure 38.

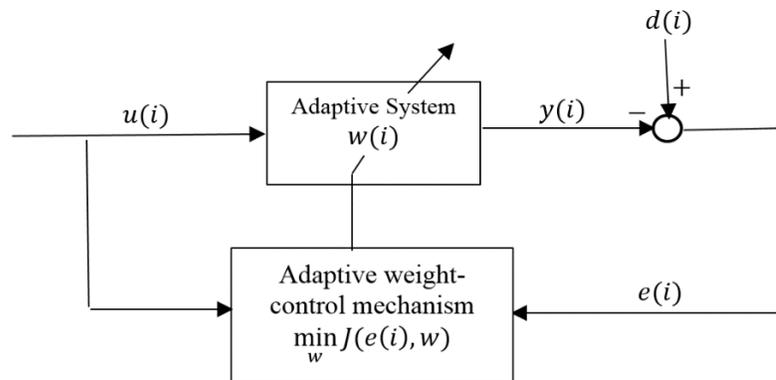


Figure 38. Basic structure of a linear adaptive filter (adapted from [79])

An input signal vector $u(i)$ applied to the filter at time i , producing the response $y(i)$. The response is subtracted from the desired response $d(i)$ to obtain the estimation error vector $e(i)$. The error signal is used to produce an adjustment to the weight vector $w(i)$ of

the filter denoted by $\Delta w(i) = G(i)e(i)$. In another words, the current estimate of $w(i)$ is computed in terms of the previous estimate and an adjustment. The adjustment is the result of a minimization operation shown in the bottom block of Figure 38. $e(i)$ is the model prediction error arising from the use of $w(i - 1)$ and $G(i)$ is the gain term.

$$w(i) = w(i - 1) + \Delta w(i) \quad (1)$$

Starting with an initial weight $w(0)$, one possible weight control mechanism can be defined as a cost function $J(i)$:

$$J(i) = E[e^2(i)], \quad i = 1, 2, 3, \dots \quad (2)$$

which also traces the learning curve of the adaptive filtering process. If the learning curve (i.e. error progression over time) is convergent, $J(i)$ reaches to a relatively steady-state value with an increasing number of iterations. In this case, the adaptive filter is convergent in the *mean-square-error sense*. The learning curve is a plot of root mean square error (RMSE) versus the number of iterations i .

3. Least-Mean-Square (LMS) Algorithm

The LMS algorithm operates by minimizing the instantaneous error term as:

$$J(i) = \frac{1}{2} e^2(i) \quad (3)$$

In this case and following the definitions in the last section we can calculate the optimal weight at iteration i . Prediction Error $e(i)$ is defined by:

$$e(i) = d(i) - y(i) = d(i) - w(i - 1)^T u(i) \quad (4)$$

Following a version of the *gradient descent method*, the instantaneous gradient vector is calculated as:

$$\frac{\partial J(i)}{\partial w(i-1)} = -e(i)u(i) \quad (5)$$

Thus, the adjustment $\Delta w(i)$ is equal to $\eta e(i)u(i)$ where η is the *step-size parameter*.

$$w(i) = w(i - 1) + \eta e(i)u(i) \quad (6)$$

At iteration i , given point \mathbf{u}^* , the output of the system is $f(\mathbf{u}^*) = \mathbf{u}^{*T} \mathbf{w}(i)$

The LMS algorithm is model independent and robust. For best performance, the step-size parameter η should be relatively small to ensure convergence, but not too small such that the algorithm converges very slowly. The overall algorithm is given below:

Algorithm 1. The least-mean-square algorithm

Initialization

$\mathbf{w}(0) = 0$, choose η

Computation

while $\{\mathbf{u}(i), d(i)\}$ available **do**

$e(i) = d(i) - \mathbf{w}(i-1)^T \mathbf{u}(i)$

$\mathbf{w}(i) = \mathbf{w}(i-1) + \eta e(i) \mathbf{u}(i)$

end while

4. Recursive Least-Square (RLS) Algorithm

The RLS algorithm tries to minimize the sum of the squared prediction errors up to the current time, whereas the LMS algorithm only minimizes the instantaneous prediction error at every iteration. Mathematically, the cost function of the RLS algorithm can be expressed as:

$$J(i) = \sum_{j=1}^i [d(j) - \mathbf{w}^T \mathbf{u}(j)] \quad (7)$$

To derive the RLS algorithm, recall that the prediction error (4) is the new information supplied to the algorithm at iteration i and the adjustment to the weight vector is denoted by $\Delta \mathbf{w}(i) = \mathbf{G}(i)e(i)$ where $\mathbf{G}(i)$ is the *gain vector*. In RLS, the gain vector is defined by

$$\mathbf{G}(i) = \mathbf{P}(i)\mathbf{u}(i) \quad (8)$$

where $\mathbf{P}(i)$ is the *state - error correlation matrix*. In fact, the matrix $\mathbf{P}(i)$ is the inverse of the time-averaged correlation matrix $\mathbf{R}(i)$ of the input vector $\mathbf{u}(i)$, as given by:

$$\mathbf{P}(i) = \mathbf{R}^{-1}(i) \quad (9)$$

$$\mathbf{R}(i) = \sum_{j=1}^i \mathbf{u}(j)\mathbf{u}^T(j) \quad (10)$$

The convergence rate of the RLS algorithm is typically an order of magnitude faster than LMS but is not model independent, i.e., the convergence rate of RLS depends on the order of the model. However, LMS is a simpler algorithm compared to RLS, as it propagates the prediction error one iteration at the time, while RLS propagates the error covariance matrix.

The computational complexity of LMS scales linearly with the dimension of the weight vector, whereas RLS scales quadratically.

5. Nonlinear Adaptive Filters

To overcome the limited computational power of linear filters, we need to build a nonlinear adaptive filter that can model any continuous input-output mapping $y = f(u)$ such that:

$$f_i = f_{i-1} + G(i)e(i) \quad (11)$$

Nonlinear solutions either append nonlinearities to the linear filters or require the availability of all data (Volterra series, Neural Networks), typically rendering them impractical.

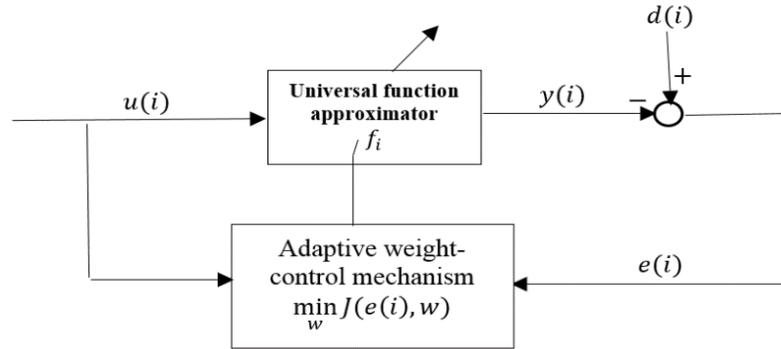


Figure 39. Basic structure of a nonlinear adaptive filter (adapted from [79])

Provided that the adaption algorithm is written as an inner product, we can take advantage of the “kernel trick” detailed below. Kernel adaptive filters also offer a very interesting alternative to neural networks, as they are universal approximators with no local minima. They also have moderate complexity in terms of computation and memory. The general structure of a nonlinear adaptive filter is depicted in Figure 27.

5.1. The Kernel Trick

$K(\mathbf{x}, \mathbf{x}')$ is said to be a valid *kernel* function if it is symmetric and positive semidefinite, i.e. $K(\mathbf{x}, \mathbf{x}') = K(\mathbf{x}', \mathbf{x})$, and the matrix:

$$\begin{bmatrix} K(x_1, x_1) & \cdots & K(x_1, x_N) \\ \vdots & \ddots & \vdots \\ K(x_N, x_1) & \cdots & K(x_N, x_N) \end{bmatrix}$$

is positive semidefinite for any x_1, \dots, x_N , which is known as *Mercer's condition*. Commonly used kernels include ‘polynomial’ kernels of the form $K_Q(\mathbf{x}, \mathbf{x}') = (\mathbf{x}^T \mathbf{x}' + c)^Q$ with $c > 0$ and Q a positive integer, and ‘Gaussian’ kernels of the form:

$$K_G(\mathbf{x}, \mathbf{x}') = e^{-\|\mathbf{x} - \mathbf{x}'\|^2 / 2\sigma^2} \quad (12)$$

Given a basic feature space χ , a kernel function induces a dilated feature space \mathcal{Z} such that computing an inner product in space \mathcal{Z} becomes straightforward.

For example, consider a kernel $K(\mathbf{x}, \mathbf{x}') = (\mathbf{x}^T \mathbf{x}' + 1)^2$ for a two-dimensional feature space $\chi = \mathbb{R}^2$ with $\mathbf{x} = [x_1 \ x_2]^T$ and $\mathbf{x}' = [x_1' \ x_2']^T$. We can write:

$$K_2(\mathbf{x}, \mathbf{x}') = (\mathbf{x}^T \mathbf{x}' + 1)^2 = 1 + 2x_1x_1' + 2x_2x_2' + x_1^2x_1'^2 + 2x_1x_1'x_2x_2' + x_2^2x_2'^2 \quad (13)$$

If we consider a dilated feature space \mathcal{Z} with:

$$\mathbf{z} = \Phi_2(\mathbf{x}) = [1 \ \sqrt{2}x_1 \ \sqrt{2}x_2 \ x_1^2 \ \sqrt{2}x_1x_2 \ x_1^2] \quad (14)$$

then we have $\mathbf{z}^T \mathbf{z}' = (\mathbf{x}^T \mathbf{x}' + 1)^2 = K_2(\mathbf{x}, \mathbf{x}')$. Similarly, a Q^{th} -order polynomial kernel $K_Q(\mathbf{x}, \mathbf{x}') = (\mathbf{x}^T \mathbf{x}' + 1)^Q$ induces a Q^{th} -order feature space \mathcal{Z} such that $\mathbf{z}^T \mathbf{z}' = (\mathbf{x}^T \mathbf{x}' + 1)^Q = K_Q(\mathbf{x}, \mathbf{x}')$. Furthermore, note that for a scalar feature space χ we can write the Gaussian kernel with $\sigma = 1/\sqrt{2}$ as:

$$K_G(x, x') = e^{-\|x - x'\|^2} = e^{-x^2} e^{-x'^2} \sum_{k=0}^{\infty} \frac{2^k x^k x'^k}{k!} \quad (15)$$

Hence a Gaussian kernel induces an infinite-dimensional feature space \mathcal{Z} such that:

$$\mathbf{z}^T \mathbf{z}' = e^{-\|x - x'\|^2 / 2\sigma^2} = K_G(\mathbf{x}, \mathbf{x}') \quad (16)$$

where σ specifies the width of the Gaussian. In the context of Support Vector Machines (SVM) [92], a large σ encompasses many neighbors and leads to a smooth boundary. As σ decreases, the boundary becomes more curved. A small value of σ usually leads to overfitting. An SVM is a classifier defined by a separating hyper-plane. In other words, given labeled training data (supervised learning), the algorithm outputs an optimal hyper-plane which categorizes new examples.

In summary, a valid kernel $K(\mathbf{x}, \mathbf{x}')$ induces a dilated feature space \mathcal{Z} in which

$$\mathbf{z}^T \mathbf{z}' = K(\mathbf{x}, \mathbf{x}') \quad (17)$$

Note that using the kernel trick to compute the inner product in space \mathcal{Z} is quite economical, especially when both the dimension of the basic feature space χ and order of the nonlinear transformation Q are not small.

6. Kernel Adaptive Filtering (KAF)

When the mapping between input signal vector and desired response is highly nonlinear, as in the case of baseline load forecasting, an algorithm capable of learning arbitrary nonlinear mapping is of interest. Kernel-based mapping can be used to transform the input vector $\mathbf{u}(i)$ into a high dimensional feature $\boldsymbol{\varphi}(\mathbf{u}(i))$. Using the “Kernel Trick” as a basic building block, many kernel-based versions of popular LMS and RLS algorithms have been proposed [81-87]. If we define a kernel expansion as:

$$f(x) = \sum_{i=1}^N \alpha_i K(x_i, x) \quad (18)$$

Kernel recursive least-squares (KRLS) algorithms calculate the coefficients α_i by solving a least-squares problem which involves the inversion of a *kernel matrix* whose dimensions depend on the amount of stored data \mathbf{M} . Therefore, they have quadratic computational and memory complexity in terms of \mathbf{M} . On the other hand, Kernel least-mean squares (KLMS) algorithms use stochastic gradient descent to obtain α_i , and they have linear complexity. For these reasons, we focus mainly on this latter type of filter in this chapter.

KAF methods suffer from increasing complexity as the number of kernels grows with the amount of data. To handle this issue, several different fixed-budget approaches can be applied to limit the amount of data that will be stored and curtail the growth of the filter structure. One can find the following algorithms in the literature:

- Kernel Least-Mean-Square [83]
- Fixed-Budget Kernel Recursive Least-Squares [88]
- Kernel Recursive Least-Squares Tracker [86]
- Quantized Kernel Least Mean Squares [85]
- Approximate Linear Dependency Kernel Recursive Least-Squares [82]
- Sliding-Window Kernel Recursive Least-Squares [89]
- Naive Online Regularized Risk Minimization Algorithm [81]
- Random Fourier Feature Kernel Least Mean Square algorithm [90]

- Extended Kernel Recursive Least Squares [91]
- Kernel Affine Projection algorithm with Coherence Criterion [84]
- Kernel Normalized Least-Mean-Square algorithm with Coherence Criterion [84]
- Kernel Recursive Least-Squares algorithm with exponential weighting [92]
- Multi-kernel Normalized Least Mean Square algorithm with Coherence based Sparsification [87]
- Parallel HYperslab Projection along Affine Sub-Space algorithm [93]
- Fixed-budget kernel least mean squares algorithm [94]
- Leaky Kernel Affine Projection Algorithm [95]
- Normalized Leaky Kernel Affine Projection Algorithm [95]
- Kernel Affine Projection Subgradient Method [96]
- Kernel Least Mean Squares algorithm with Coherence-Sparsification criterion and L1-norm regularization and with active L1-norm regularization [97]
- Mixture Kernel Least Mean Square algorithm [98]

It is worth mentioning that due to the nature of adaptive filter implementation in the sense of having an embedded time delay (filter order selection), this approach is useful for very short-term baseline forecasting.

Moreover, Support Vector Machines, KAF and other models employing the kernel trick do not scale well due to the number of training samples and large magnitudes or a large number of features in the input space. Hence, it is often helpful to normalize the training data.

7. Kernel Least Mean Square Algorithm

Following the steps mentioned in the LMS algorithm and considering the fact that we have transformed the input vector into a high dimensional feature space $\boldsymbol{\varphi}(\mathbf{u}(i))$, one can write:

$$\boldsymbol{\omega}(0) = \mathbf{0} \quad (19)$$

$$\mathbf{e}(i) = d(i) - \boldsymbol{\omega}(i-1)^T \boldsymbol{\varphi}(\mathbf{u}(i)) \quad (20)$$

$$\boldsymbol{\omega}(i) = \boldsymbol{\omega}(i-1) + \eta \mathbf{e}(i) \boldsymbol{\varphi}(\mathbf{u}(i)) \quad (21)$$

$$\boldsymbol{\omega}(i) = [\boldsymbol{\omega}(i-2) + \eta \mathbf{e}(i-1) \boldsymbol{\varphi}(\mathbf{u}(i-1))] + \eta \mathbf{e}(i) \boldsymbol{\varphi}(\mathbf{u}(i)) \quad (22)$$

$$\boldsymbol{\omega}(i) = \boldsymbol{\omega}(i-2) + [\eta \mathbf{e}(i-1) \boldsymbol{\varphi}(\mathbf{u}(i-1)) + \eta \mathbf{e}(i) \boldsymbol{\varphi}(\mathbf{u}(i))] \quad (23)$$

$$\omega(i) = \eta \sum_{j=1}^i e(j) \varphi(u(j)) \quad (\text{assuming } \omega(0) = 0) \quad (24)$$

The weight estimate after i -step of training is expressed as a linear combination of all previous and currently transformed input vectors, weighted by the prediction error and scaled by step-size. Moreover, given point u^* at iteration i , the output of the system can be expressed in terms of inner products between transformed inputs.

$$\omega(i)^T \varphi(u^*) = \left[\eta \sum_{j=1}^i e(j) \varphi(u(j)) \right]^T \varphi(u^*) = \eta \sum_{j=1}^i e(j) [\varphi(u(j))]^T \varphi(u^*) \quad (25)$$

Using the Kernel trick we can write the filter output as:

$$\omega(i)^T \varphi(u^*) = \eta \sum_{j=1}^i e(j) K(u(j), u^*) \quad (26)$$

If we define f_i as the estimate of the non-linear input-output mapping at iteration i , we can write the following sequential learning rule:

$$f_{i-1}(u(i)) = \eta \sum_{j=1}^{i-1} e(j) K(u(j), u(i)) \quad (27)$$

$$e(i) = d(i) - f_{i-1}(u(i)) \quad (28)$$

$$f_i = f_{i-1} + \eta e(i) K(u(i), \cdot) \quad (29)$$

which means the KLMS defines a new Kernel for the new training data with input $u(i)$ as the center and $\eta e(i)$ as the coefficient.

Algorithm 2. The Kernel least-mean-square algorithm

Initialization

Choose step-size and Kernel

$$a_1(1) = \eta d(1), \mathcal{C}(1) = \{u(1)\}, f_1 = a_1(1) K(u(1), \cdot)$$

Computation

while $\{u(i), d(i)\}$ available **do**

$$f_{i-1}(u(i)) = \eta \sum_{j=1}^{i-1} a_j(i-1) K(u(i), u(j)) \quad \% \text{compute the output}$$

$$e(i) = d(i) - f_{i-1}(u(i)) \quad \% \text{compute the error}$$

$$\mathcal{C}(i) = \{\mathcal{C}(i-1), u(i)\} \quad \% \text{store the new centre}$$

$$a_i(i) = \eta e(i) \quad \% \text{compute and store the new coefficient}$$

end while

It has been proven mathematically that KLMS does not need explicit regularization [83]. However, Kivinen et al. [81] proposed a similar algorithm called Naive Online Regularized Risk Minimization Algorithm (NORMA) by differentiating the regularized function:

$$\min_f J(f) = \sum_{i=1}^n |d(i) - f(u(i))|^2 + \lambda \|f\|^2 \quad (30)$$

with λ as the regularization parameter. The result leads to the following update rule:

$$f_i = (1 - \eta\lambda)f_{i-1} + \eta e(i)K(u(i), \cdot) \quad (31)$$

The scaling factor $(1 - \eta\lambda)$ compared to the KLMS estimate scales down the training data with small coefficients exponentially. Since this algorithm imposes a memory loss mechanism; it is also termed a Leaky KLMS. The regularization however introduces a bias in the solution which degrades its performance compared to KLMS [99].

It is also straightforward to obtain a Normalized KLMS algorithm based on its counterpart NLMS. The only difference stems from weight update equation as:

$$\omega(i) = \omega(i - 1) + \frac{\eta}{\varepsilon + \|\varphi(u(i))\|^2} e(i)\varphi(u(i)) \quad (32)$$

where $\varepsilon + \|\varphi(u(i))\|^2$ is the normalizing term and ε is the small positive number to prevent division by zero. Note that using the definition of the norm in the feature space, we have:

$$\|\varphi(u(i))\|^2 = \langle \varphi(u(i)), \varphi(u(i)) \rangle = K(u(i), u(i)) \quad (33)$$

8. Computer Simulation

In the following, the aforementioned algorithms were selected to evaluate and compare kernel adaptive filtering algorithms in the context of baseline load forecasting. The Kernel Adaptive Filtering Toolbox (KAFBOX) [100] was used as a benchmarking tool. The same dataset (household aggregate load consumption signal from the Pecan St. project and BC aggregate load in 2012) mentioned in previous chapters is used for the case studies.

First of all, the convergence behavior of the LMS and KLMS algorithms is examined. A learning curve is used to compare the performance of these two filters. The 2012 BC aggregate load dataset contains 8760 data points. The first 8000 points are the training data and the last 760 points are the testing data. The time delay is set to be 10 and the kernel parameter is fixed at one. To produce the learning curves, a segment of 800 samples is used for training and another 200 samples as the test data. At each iteration, the RMSE is computed on the test set as a result of the filter obtained in the training set. As shown in Figure 28, the KLMS outperforms the LMS by converging to a smaller RMSE because of

its nonlinear structure. Note the declining trend of RMSE for KLMS as it progresses over time.

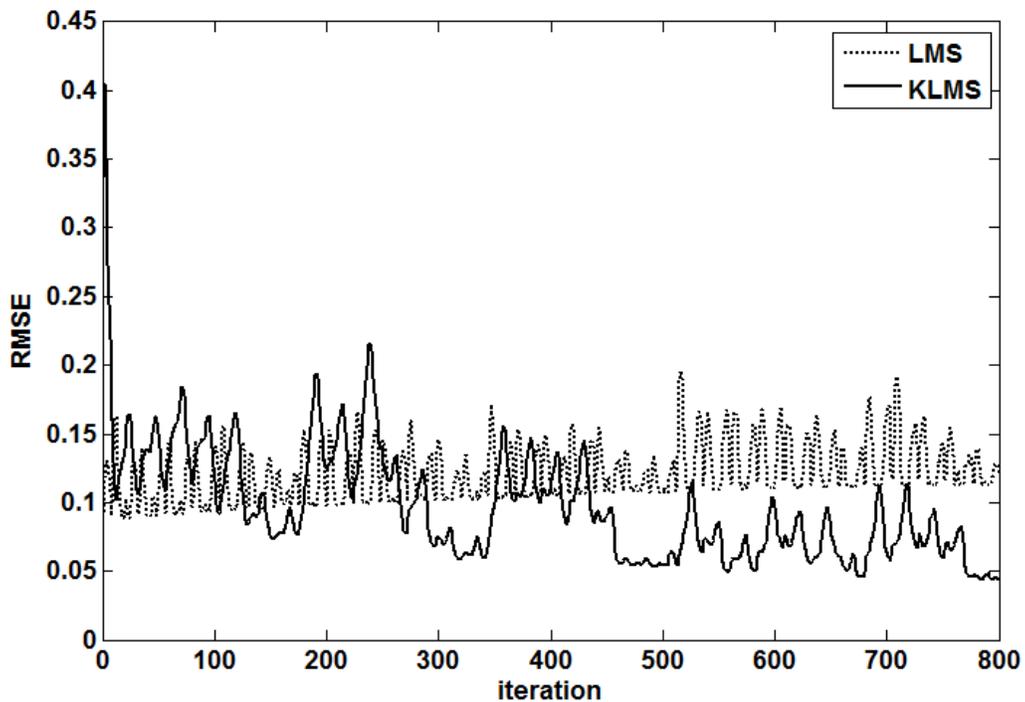


Figure 40. LMS and KLMS learning curves

Furthermore, different training data sizes are used to investigate how the size of the training dataset affects the KLMS performance. As shown in Table. 11, KLMS performs slightly better with a larger number of training data points. However, it negatively affects the computational time. Theoretical analysis shows that training data set does not have an effect on the regularization of KLMS [1].

To study the effect of the regularization parameter on the performance of NORMA, one needs to plot the RMSE vs. different values of the regularization parameter. Figure 29 shows that mean and standard deviation of RMSE tends to increase for higher values of the regularization parameter.

Table 11. RMSE for different testing data sizes for LMS and KLMS

Algorithm	Testing RMSE
LMS (N=500)	0.062622+/-0.0007284
LMS (N=800)	0.054538+/-0.00066083
LMS (N=1000)	0.048098+/-0.00061448
KLMS (N=500)	0.054555+/-0.0011731
KLMS (N=800)	0.046518+/-0.00059118
KLMS (N=1000)	0.041922+/-0.001003

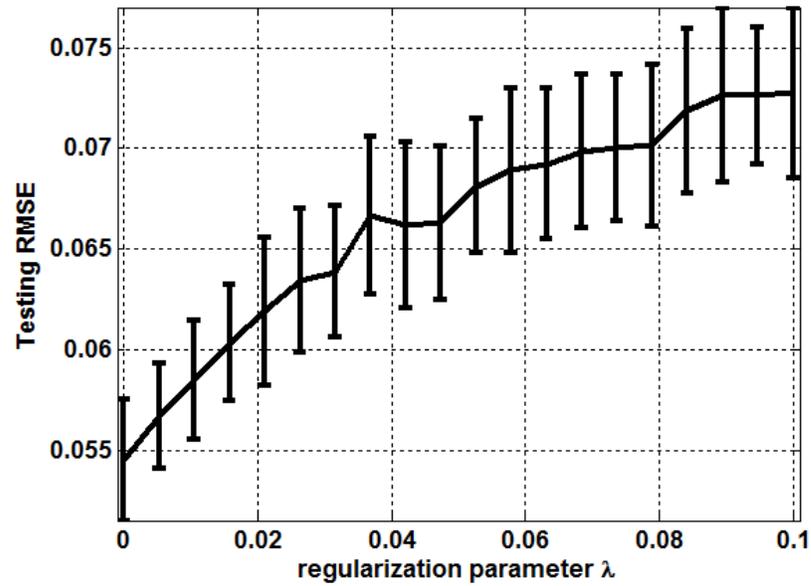


Figure 41. Effect of regularization parameter on RMSE mean and standard deviation

KLMS are implemented for different values of step size, kernel parameter, and regularization parameter. The RMSE calculated after 1000 data point for both data sets. The results are shown in Table 12 along with the Figure 28 for the BC aggregate load data ($\eta = 0.5$, $kernelpar = 10$) and Table 13 along with Figure 29 for Pecan St. data set.

Table 12. KLMS comparison for different value of step size and kernel parameter (BC Load)

Step size (η)	Kernel parameter	RMSE
0.1	1	0.06
0.5	1	0.05
1	1	0.04
0.5	5	0.04
0.5	10	0.04

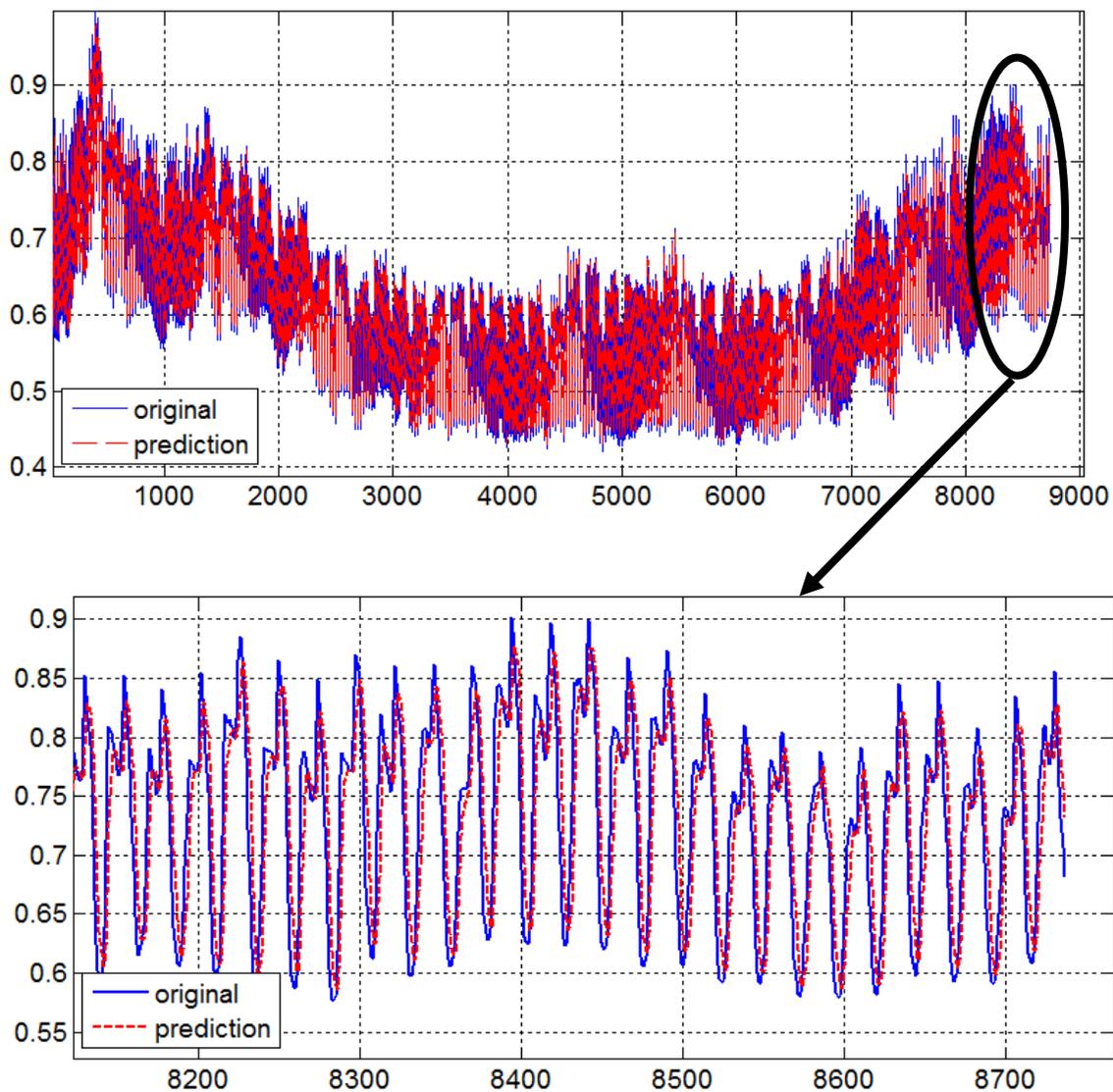


Figure 42. Prediction vs. actual time series data for BC aggregate load 2012

Table 13. KLMS comparison for different value of step size and kernel parameter (Pecan St)

Step size (η)	Kernel parameter	RMSE
0.1	1	0.18
0.5	1	0.13
1	1	0.14
0.5	5	0.13
0.5	10	0.13

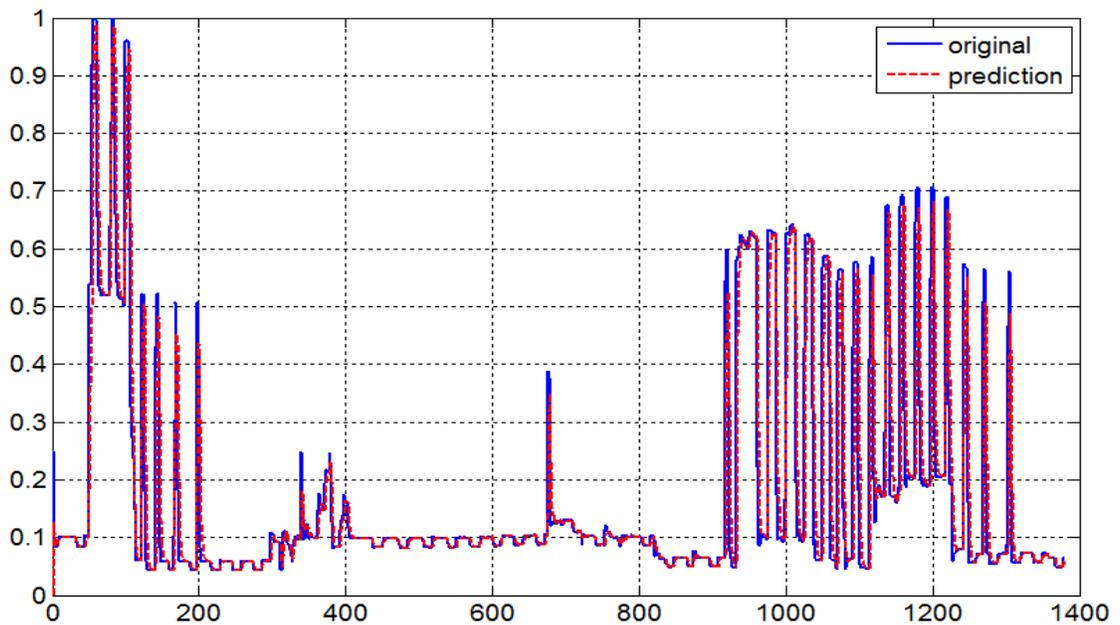


Figure 43. Prediction vs. actual time series data for an aggregate household in Pecan St.

9. Summary

When the mapping between input signal vector and desired response is highly nonlinear, as in the case of baseline load forecasting, an algorithm capable of learning arbitrary nonlinear mapping is of interest. Kernel-based mapping can be used to transform the input vector into a high dimensional feature. Using the “Kernel Trick” as a basic building block, one can reduce the filtering process to an efficient vector multiplication. In this chapter, we investigate the potential of Kernel Adaptive Filtering for customer baseline load

forecasting using two case studies employed in previous chapters. The computer simulations show promising results however some issues need to be considered. Due to the nature of adaptive filter implementation in the sense of having an embedded time delay (filter order selection), this approach is useful for short term baseline forecasting.

Moreover, kernel trick does not scale well due to the number of training samples and large magnitudes or many features in the input space. Hence, it is often helpful to normalize the data first. Tuning of filter parameters is required to achieve the best results.

Chapter 6: Conclusions and Future Work

1. Summary & Discussion

The research presented in this thesis provides novel ways to address demand response scheduling and assessment. In chapter 3, a stochastic optimization model has been used to schedule energy and reserve provided by generation and demand-side resources. It is necessary to adopt stochastic approaches for decision making under uncertainty, such as in the case of clearing electricity markets with high levels of wind generation. This permits the scheduling of loads and services in coordination with the system operator, thus enhancing the power system's efficiency and security while reducing its environmental impact. The proposed stochastic model is formulated as a two-stage SMIP problem. Several case studies are conducted and the results presented demonstrate the benefits of customers' response for ISO. Utilizing a higher share of DR reserve will increase the overall system welfare and decrease the wind power spillage. This facilitates a higher share of renewable energy integration in the power system. For example, when demand side reserve costs are 30 \$/MWh, on average 2.3 MW of wind power production is spilled compared to 1.46 MW wind power spillage when the LA reserve cost reduced to \$5/MWh. Moreover, DR improves commitment schedules of generating units and decreases marginal cost variations. DR can potentially benefit consumers by offering cheaper electricity while facilitating a higher share of renewable energy sources in the power system.

Chapter 4 introduces the idea of using a recursive Bayesian linear regression approach for baseline load analysis in a smart grid framework where real-time model prediction is of importance for operators and market participants. After defining the principle regression equation; for simulation purposes, the available data was broken into four smaller data sets. A quasi-real-time model updating algorithm was used to train and validate the model. The first batch of data was used to calibrate the regression model, while data from the second batch was used to validate the model ($\mathcal{M}1$). Then, first and second of data were used to calibrate the model and the third batch for validation ($\mathcal{M}2$). Then, first, three batches of data were used to calibrate the model and the last batch of data for validation ($\mathcal{M}3$). This structure is used to simulate a real-time recursive algorithm, where when more data is available, the model updates its parameters. By doing so, any changes in the system are implicitly captured by the data-driven model as each subsequent year is added to the dataset.

The results show that BLR approach provides a very good forecast of the hourly/minute load. Forecasting accuracy as measured by RMSE improves with lower mean load while relative error (MAPE) decreases higher levels of aggregation. Additionally, the Bayesian approach offers an integrated solution for estimation with uncertainty. As more data is added, the performance improves in all cases which illustrate the effectiveness of the Bayesian method and the benefits of implementing an online learning algorithm for baseline analysis.

In chapter 5, Kernel adaptive filtering was proposed as an on-line learning algorithm for short time baseline load forecasting which can be used by utilities or aggregators. KAF is useful because the model can continuously adapt to track changes in the underlying system as new data arrives. To evaluate and compare Kernel Least mean Squares vs. Ordinary Least Mean Square, different simulation results were performed on two test cases introduced in Chapter 4: BC aggregate load and a single house from Pecan St. project. KLMS outperforms the LMS by converging to a smaller RMSE because of its nonlinear structure. The results are promising, but one needs to consider fixed-budget approaches to limit the amount of data that will be stored and curtail the growth of the filter structure. Furthermore, for each dataset careful investigation is required for tuning the filter parameters to achieve the best result.

2. Future possibilities for current research work

Despite various advantages offered by the hourly DR providing reserve shown in chapter 3, there are several open issues that can be addressed. Stochastic optimization models such that Load Aggregator provides frequency regulation can be included. Uncertainty in load was not considered. Thus, the total electricity load may not be the same as expected. Another interesting venue to investigate is studying optimization platform for DR aggregator in terms of various contracts like load curtailment, load shifting, peak demand management, etc. It is worth mentioning that due to the power flow constraints in the formulation, the computational complexity does not increase linearly as we increase the number of nodes (depends on how many cross connections added). Future studies on the complexity analysis of the formulation can be pursued.

The Bayesian method introduced in chapter 4 can be applied to other models, like support vector regression and Neural Networks. It would be interesting to investigate the performance of those methods compared to Bayesian Linear Regression. A comparison between our proposed approach and day-matching CBL method being employed by utilities needs to be done in the presence of real demand response events.

This topic could also be expanded for peak demand forecasting. For aggregators that provide services such as peak demand management for their customers, peak demand value and temporal forecasting is of great importance. Moreover, in most markets like Ontario's IESO, large industrial consumers (customers with an average peak demand of greater than 5 MW annually) pay a \$/kWh charge based on a percentage of their contribution to the top five Ontario peaks (i.e. peak demand factor) over a 12-month base period. There is a big value add if an aggregator is able to predict when peak days occur and manage demand during events.

The purpose of chapter 5 was to lay the groundwork for using online learning methods in the context of consumer baseline analysis. Our goal was to introduce the concept of employing Kernel Adaptive Filtering in short-term forecasting. In context of DR, one can practically deploy this method to predict baseline in services like frequency regulation due to short duration of this market. However, Further studies and complete implementation of other sequential learning algorithms like Kernel recursive least squares are required to be able to conduct a thorough comparison in chapter 5.

In most cases for baseline estimation utilities are interested to know the baseline for the last month to set the financial settlements. An important note is that for baselining case studies in this thesis; a data set where DR was dispatched was not available. In practice, dispatching DR will change the load profile and cannot be used for future baseline calculation. A simple solution is to discard those time stamps where a DR event was dispatched.

As we are seeing larger penetration level of demand dispatch (not only as a contingency measure only), new methods are needed to compensate end-use customers. Historical baseline estimation to calculate the demand reduction is not the answer since the load profile is completely going to change. This thesis was an attempt to tackle this problem

using (pseudo) real time methods for baseline calculation. This also addresses one of the biggest hesitation of utilities around DR in terms of dispatchability.

Bibliography

- [1] Renewables Account for Two-thirds of New US Generating Capacity in 2015; 3,500 Times More Than Coal, Accessed on 15 Feb 2016
URL: <http://www.renewableenergyworld.com/articles/2016/02/renewables-provides-two-thirds-of-new-us-generating-capacity-in-2015-3-500-times-more-than-coal.html>
- [2] Demand Response Research Center Annual Report 2013, Available: <http://drrc.lbl.gov/news/drrc-annual-report-2013>
- [3] Benefits of Demand Response in Electricity Markets and Recommendations for Achieving Them U.S. Department of Energy, 2006.
- [4] JL. Mathieu, “Modeling, Analysis, and Control of Demand Response Resources” Ph.D. dissertation, Dept. Mechanical Eng., University of California, Berkeley, 2012.
- [5] J. Page, S. Kiliccote, J.H. Dudley, M.A. Piette, A.K. Chiu, B. Kellow, E. Koch, and P. Lipkin. Automated demand response technology demonstration project for small and medium commercial buildings. Technical report, Lawrence Berkeley National Laboratory, July 2011.
- [6] Energy Storage and Demand Response for improved reliability in an outage-prone community; Nov2015. <<http://tinyurl.com/odbag2p>>
- [7] C. Guille and G. Gross, "A conceptual framework for the vehicle-to-grid (V2G) implementation," *Energy Policy*, vol. 37, pp. 4379-4390, 2009.
- [8] M. Parvania and M. Fotuhi-Firuzabad, “Demand response scheduling by stochastic SCUC,” *IEEE Trans. Smart Grid*, vol. 1, no. 1, pp. 89–98, Jun. 2010.
- [9] Load Participation in the ERCOT Market Electric Reliability Council of Texas (ERCOT) [Online]. Available: <http://www.ercot.com/services/programs/load/laar/>
- [10] Installed Capacity Manual New York Independent System Operator (NYISO) [Online]. Available: http://www.nyiso.com/public/markets_operations/market_data/demand_response/index.jsp
- [11] Scheduling Operations Manual Pennsylvania–New Jersey–Maryland (PJM) Interconnection [Online]. Available: <http://www.pjm.com/markets-and-operations/demand-response.aspx>
- [12] P. L. Langbein, “Demand response participation in PJM wholesale markets,” in *Proceedings of the 2012 IEEE PES Innovative Smart Grid Technologies (ISGT) Conference*, Washington, U.S.A., January 2012.
- [13] ISO New England Load Response Program Manual. New England, Independent System Operator (ISO) New England [Online]. Available: <http://www.iso-ne.com/>
- [14] J. Torriti, M. G. Hassan, and M. Leach, “Demand response experience in Europe: policies, programmes and implementation,” *Energy*, vol. 35, no. 4, pp. 1575–1583, Apr. 2010.

- [15] V. Giordano, F. Gangale, G. Fulli, M. S. Jiménez, I. Onyeji, A. Colta, and et al., “Smart Grid projects in Europe: lessons learned and current developments,” European Commission - Joint Research Centre, Tech. Rep. EUR 24856, 2011. [Online]. Available: http://ses.jrc.ec.europa.eu/sites/ses/files/documents/smart_grid_projects_in_europe_lessons_learned_and_current_developments.pdf
- [16] G. Heffner, C. Goldman, M. Kintner-Meyer, and B. Kirby, “Loads providing ancillary services: review of international experience. Technical appendix: Market descriptions,” Lawrence Berkeley National Laboratory, Tech. Rep. LBNL-62701, May 2007. [Online]. Available: <http://certs.lbl.gov/pdf/62701-app.pdf>
- [17] C. Søndergren, N. C. Bang, C. Hay, M. Tøgeby, and J. Østergaard, “Electric vehicles in future market models,” EDISON Project, Tech. Rep. D2.3, June 2011. [Online]. Available: <http://www.edison-net.dk/Dissemination/Reports.aspx>
- [18] J. Cochran, L. Bird, J. Heeter, and D. J. Arent, “Integrating variable renewable energy in electric power markets: best practices from international experience,” National Renewable Energy Laboratory, Tech. Rep. NREL/TP-6A00-53732, April 2012. [Online]. Available: <http://www.nrel.gov/docs/fy12osti/53732.pdf>
- [19] R. Bessa. "Methodologies for the Participation of an Electric Vehicles Aggregator in the Electricity Markets", Repositório Aberto da Universidade do Porto, 2014.
- [20] <https://www.terna.it/en-gb/homepage.aspx>
- [21] T. Woolf, E. Malone, L. Schwartz & J. Shenot, “A Framework for Evaluating the Cost Effectiveness of Demand Response”, Tech. Rep., Feb. 2013 Available: <http://emp.lbl.gov/publications/framework-evaluating-cost-effectiveness-demand-response>
- [22] Wilson, R.; B. Biewald. 2013. *Best Practices in Utility Resource Planning: Examples of State Regulations and Recent Utility Plans*. Synapse Energy Economics.)
- [23] FERC. Order No. 719, Wholesale competition in regions with organized electric markets. Federal Energy Regulatory Commission. 2008. URL: <https://www.ferc.gov/whats-new/comm-meet/2008/101608/E-1.pdf>
- [24] FERC. Order No. 745, Demand response compensation in organized wholesale energy markets. Federal Energy Regulatory Commission. 2011. URL: <https://www.ferc.gov/EventCalendar/Files/20110315105757-RM10-17-000.pdf>
- [25] Weinschenk, C. (January 26, 2016) Supreme Court Overturns DC Circuit, Oks FERC Order 745. URL: <http://www.energymanagertoday.com/supreme-court-overturns-dc-circuit-oks-ferc-order-745-0121502/>
- [26] DP. Chassin, “New Residential Thermostat for Transactive Systems” Master. dissertation, Dept. Mechanical Eng., University of Victoria, 2014.
- [27] CAISO demand response initiative, [online]. Available:

<https://www.caiso.com/informed/Pages/StakeholderProcesses/DemandResponseInitiative.aspx>

[28] Hammerstrom DJ, Ambrosio R, Brous J, Carlon TA, Chassin DP, DeSteele JG, et al. Pacific northwest gridwise testbed demonstration projects. Part I. Olympic Peninsula Project; 2007.

[29] Widergren SE, Subbarao K, Fuller JC, Chassin DP, Somani A, Marinovici MC, et al. AEP Ohio gridSMART demonstration project real-time pricing demonstration.

[30] S. Behboodi, DP. Chassin, C. Crawford, N. Djilali. Renewable Resources portfolio optimization in the presence of demand response. *Appl Energy* 2016; 162:139-148.

[31] Accommodating High Levels of Variable Generation, North American Electric Reliability Corporation (NERC), Apr. 2009.

[32] R. N. Boisvert, P. A. Cappers, and B. Neenan, "The benefits of customer participation in wholesale electricity markets," *Electricity J.*, vol. 15, no. 3, pp. 41–51, Apr. 2002.

[33] Federal Energy Regulatory Commission, "National Assessment of Demand Response Potential", The Brattle Group, Freeman, Sullivan & Co, Global Energy Partners, LLC, Staff Report, June 2009, p72.

[34] M. Parvania and M. Fotuhi-Firuzabad, "Integrating load reduction into wholesale energy market with application to wind power integration," *IEEE Syst. J.*, vol.6, no. 1, pp. 35–45, Mar. 2012.

[35] M. Parvania, M. Fotuhi-Firuzabad, and M. Shahidehpour, "ISO's Optimal Strategies for scheduling the Hourly Demand Response in Day-Ahead Markets," *IEEE Trans. Power System*, vol. 29, no. 6, pp. 2636–2645, Nov. 2014.

[36] A. Khodaei, M. Shahidehpour, and S. Bahramirad, "SCUC with hourly demand response considering intertemporal load characteristics," *IEEE Trans. Smart Grid*, vol. 2, no. 3, pp. 564–571, Sep. 2011.

[37] H. Wu, M. Shahidehpour, and M. Khodayar, "Hourly demand response in day-ahead scheduling considering generating unit ramping cost," *IEEE Trans. Power System*, vol. 28, no. 3, pp. 2446–2454, Aug. 2013.

[38] F. C. Schweppe, M. C. Caramanis, R. D. Tabors, and R. E. Bohn. *Spot Pricing of Electricity*. Springer, 1988.

[39] R. Weron, *Modeling and forecasting electricity loads and prices, A Statistical Approach*: John Wiley & Sons, Ltd, 2006.

[40] J. M. Morales, A. J. Conejo, K. Liu, J. Zhong, "Pricing Electricity in Pools With Wind Producers," *IEEE Trans. Power System*, vol. 27, no. 3, pp. 1366–1376, Aug. 2014.

[41] J. R. Birge and F. Louveaux. *Introduction to stochastic programming*. Springer-Verlag, New York, 1997.

- [42] A. J. Conejo, M. Carrion, J. M. Morales, *Decision Making Under Uncertainty in Electricity Markets*, Springer, New York, NY, 2010.
- [43] F. Bouffard, and F. Galiana, “Stochastic Security for Operations Planning With Significant Wind Power Generation,” *IEEE Trans. Power System*, vol. 23, no. 2, pp. 306–316, May. 2008.
- [44] Reliability Test System Task Force, “The IEEE reliability test system—1996,” *IEEE Trans. Power System*, vol. 14, no.3, pp. 1010–1020, Aug. 1999.
- [45] CPLEX 12.2 Manual, IBM Corporation. Armonk, NY, USA, 2011.
- [46] [Online]. Available: <http://www.nordpoolspot.com/>
- [47] A. Al-Awami and M. El-Sharkawi, “Coordinated Trading of Wind and Thermal Energy,” *IEEE Trans. Sustainable Energy*, vol. 2, no. 3, pp. 277–287, Jul. 2011.
- [48] Coughlin, K., Piette, M., Goldman, C., and Kiliccote, S., 2009, “Statistical Analysis of Baseline Load Models for Non-Residential Buildings,” *Energy Build.*, 41(4), pp. 374–381.
- [49] Goldberg, M., and Agnew, G., 2003, “Protocol Development for Demand Response Calculation—Findings and Recommendations,” Technical Report No. CEC 400-02-017F, California Energy Commission (KEMA-XENERGY).
- [50] “*Demand Response – Measurement and Verification Applications for Load Research*”, March 2009, AEIC Load Research Committee Publication
- [51] Katipamula, S., Reddy, T., and Claridge, D., 1998, “Multivariate Regression Modeling,” *ASME J. Sol. Energy Eng.*, 120(3), pp. 177–184.
- [52] Kissock, J., Reddy, T., and Claridge, D., 1998, “Ambient-Temperature Regression Analysis for Estimating Retrofit Savings in Commercial Buildings,” *ASME J. Sol. Energy Eng.*, 120(3), pp. 168–176.
- [53] Kozikowski, D., Breidenbaugh, A., and Potter, M., 2006, “The Demand Response Baseline, v.1.75,” EnerNOC OPS Publication.
- [54] N. Addy, S. Kiliccote, D. Callaway and JL. Mathieu, “How Baseline Model Implementation Choices Affect Demand Response Assessments,” *ASME J. Sol. Energy Eng.*, 2014, 137(2), pp. 021008.
- [55] Mathieu, J., Price, P., Kiliccote, S., and Piette, M., 2011, “Quantifying Changes in Building Electricity Use, With Application to Demand Response,” *IEEE Trans. Smart Grid*, 2(3), pp. 507–518.
- [56] Wi, Y.-M., Kim, J.-H., Joo, S.-K., Park, J.-B., and Oh, J.-C., 2009, “Customer Baseline Load (CBL) Calculation Using Exponential Smoothing Model With Weather Adjustment,” *Transmission Distribution Conference Exposition: Asia and Pacific*, Seoul, Korea, Oct. 26–30, pp. 1–4.

[57] “*Demand Response – Measurement and Verification Applications for Load Research*”, March 2009, AEIC Load Research Committee Publication

[58] Khan, U. T. & Valeo, C. Comparing a Bayesian and fuzzy number approach to uncertainty quantification in short-term dissolved oxygen prediction. (In Press) *Journal of Environmental Informatics*, 2015.

[59] Khan, U. T. & Valeo, C. Short-term peak flow rate prediction and flood risk assessment using fuzzy linear regression. (In Press) *Journal of Environmental Informatics*, 2015.

[60] Khan, U. T., & Valeo, C. Predicting Dissolved Oxygen Concentration in Urban Watersheds: A Comparison of Fuzzy Number Based and Bayesian Data-Driven Approaches. In *The International Conference on Marine and Freshwater Environments (iMFE 2014)* (Vol. 1, No. 1), 2014.

[61] Solomantine, D. P., & Ostfeld, A. (2008). Data-driven modelling: some past experiences and new approaches. *Journal of Hydroinformatics*, 10(1), 3–22. doi:10.2166/hydro.2008.015.

[62] Kutz, J. Nathan. *Data-driven modeling & scientific computation: methods for complex systems & big data*. Oxford University Press, 2013.

[63] Vrugt, J. A., ter Braak, C. J. F., Gupta, H. V., & Robinson, B. A. (2009). Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modelling? *Stochastic Environmental Research and Risk Assessment*, 23(7), 1011–1026. doi: 10.1007/s00477-008-0274-y.

[64] Freni, G., & Mannina, G. (2010). Bayesian approach for uncertainty quantification in water quality modelling: The influence of prior distribution. *Journal of Hydrology*, 392(1), 31–39. doi: 10.1016/j.jhydrol.2010.07.043.

[65] Tyralis, H., & Koutsoyiannis, D. (2014). A Bayesian statistical model for deriving the predictive distribution of hydroclimatic variables. *Climate dynamics*, 42(11–12), 2867–2883. doi: 10.1007/s00382-013-1804-y.

[66] Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian Data Analysis*. (3rd ed.). CRC Press, Boca Raton, FL, USA.

[67] Thiemann, M., Trosset, M., Gupta, H., & Sorooshian, S. (2001). Bayesian recursive parameter estimation for hydrologic models. *Water Resources Research*, 37(10), 2521–2535. doi: 10.1029/2000WR900405.

[68] Kingston, G. B., M. F. Lambert, & H. R. Maier (2005), Bayesian training of artificial neural networks used for water resources modeling, *Water Resources Research*, 41, W12409. doi: 10.1029/2005WR004152.

[69] Koop, G. (2003). *Bayesian Econometrics*. John Wiley & Sons. West Sussex, England, UK.

[70] Greenberg, E. (2008). *Introduction to Bayesian econometrics*. Cambridge University Press. New York, NY, USA. doi: 10.1017/CBO9781139058414.

- [71] Birkes, D., & Dodge, Y. (1993). *Alternative Methods of Regression*. John Wiley & Sons. New York, NY, USA. doi: 10.1002/9781118150238.
- [72] Shrestha, D. L., & Solomatine, D. P. (2008). Data-driven approaches for estimating uncertainty in rainfall-runoff modelling. *International Journal of River Basin Management*, 6(2), 109–122. doi: 10.1080/15715124.2008.9635341.
- [73] Vicens, G. J., Rodriguez-Iturbe, I., & Schaake, J. C. (1975). A Bayesian framework for the use of regional information in hydrology. *Water Resources Research*, 11(3), 405–414. doi: 10.1029/WR011i003p00405.
- [74] URL: <http://transmission.bchydro.com/home>
- [75] URL: http://climate.weather.gc.ca/data_index_e.html
- [76] (2015) Pecan street database. [Online]. Available: <http://www.pecanstreet.org/>
- [77] Diamond, H. J., T. R. Karl, M. A. Palecki, C. B. Baker, J. E. Bell, R. D. Leeper, D. R. Easterling, J. H. Lawrimore, T. P. Meyers, M. R. Helfert, G. Goodge, and P. W. Thorne, 2013: U.S. Climate Reference Network after one decade of operations: status and assessment. *Bull. Amer. Meteor. Soc.*, **94**, 489-498. doi: [10.1175/BAMS-D-12-00170.1](https://doi.org/10.1175/BAMS-D-12-00170.1)
- [78] Sevlian R. and Rajagopal R. Short term electricity load forecasting on varying levels of aggregation. *arXiv preprint arXiv: 1404.0058,2014*.
- [79] Weifeng Liu, José C. Principe, Simon Haykin (March 2010). *Kernel Adaptive Filtering: A Comprehensive Introduction*. Wiley. pp. 12–20. ISBN 978-0-470-44753-6.
- [80] V. N. Vapnik, *Statistical Learning Theory*, Wiley, 1998.
- [81] J. Kivinen, A. J. Smola, and R. C. Williamson, “Online learning with kernels,” *IEEE Trans. on Sig. Proc.*, vol. 52, no. 8, pp. 2165–2176, Aug. 2004.
- [82] Y. Engel, S. Mannor, and R. Meir, “The kernel recursive least squares algorithm,” *IEEE Trans. on Sig. Proc.*, vol. 52, no. 8, pp. 2275–2285, Aug. 2004.
- [83] W. Liu, P. P. Pokharel, and J. C. Principe, “The kernel least-mean-square algorithm,” *IEEE Trans. on Sig. Proc.*, vol. 56, no. 2, pp. 543–554, 2008.
- [84] C. Richard, J. C.M. Bermudez, and P. Honeine, “Online prediction of time series data with kernels,” *IEEE Trans. on Sig. Proc.*, vol. 57, no. 3, pp. 1058–1067, Mar. 2009.
- [85] B. Chen, S. Zhao, P. Zhu, and J. C. Principe, “Quantized kernel least mean square algorithm,” *IEEE Trans. On Neural Networks and Learning Systems*, vol. 23, no. 1, pp. 22–32, Jan. 2012.
- [86] S. Van Vaerenbergh, M. Lazaro-Gredilla, and I. Santamaria, “Kernel recursive least-squares tracker for time varying regression,” *IEEE Trans. on Neural Networks and Learning Systems*, vol. 23, no. 8, pp. 1313–1326, Aug. 2012.

- [87] M. Yukawa, "Multikernel adaptive filtering," *IEEE Trans. on Sig. Proc.*, vol. 60, no. 9, pp. 4672–4682, 2012.
- [88] S. Van Vaerenbergh, I. Santamaria, W. Liu and J. C. Principe, "Fixed Budget Kernel Recursive Least Squares", 2010 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2010), Dallas, Texas, U.S.A., March 2010.
- [89] S. Van Vaerenbergh, J. Via, and I. Santamaria. "A sliding window kernel RLS algorithm and its application to nonlinear channel identification", 2006 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Toulouse, France, 2006.
- [90] Abhishek Singh, Narendra Ahuja and Pierre Moulin, "Online learning with kernels: Overcoming the growing sum problem," 2012 IEEE International Workshop on Machine Learning for Signal Processing (MLSP), Sept. 2012.
- [91] W. Liu and I. Park and Y. Wang and J.C. Principe, "Extended kernel recursive least squares algorithm", *IEEE Transactions on Signal Processing*, volume 57, number 10, pp. 3801-3814, Oct. 2009.
- [92] S. Haykin, "Adaptive Filtering Theory (3rd Ed.)", Prentice Hall, Chapter 13.
- [93] M. Takizawa and M. Yukawa, "An Efficient Data Reusing Kernel Adaptive Filtering Algorithm Based on Parallel Hyperslab Projection along Affine Subspace," 2013 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), pp.3557-3561, May 2013.
- [94] D. Rzepka, "Fixed budget kernel least mean squares," 2012 IEEE 17th Conference on Emerging Technologies & Factory Automation (ETFA), Krakow, Poland, Sept. 2012.
- [95] W. Liu and J.C. Principe, "Kernel Affine Projection Algorithms", *EURASIP Journal on Advances in Signal Processing*, Volume 2008, Article ID 784292, 12 pages.
- [96] K. Slavakis, S. Theodoridis, and I. Yamada, "Online kernel based classification using adaptive projection algorithms," *IEEE Transactions on Signal Processing*, Vol. 56, No. 7, pp. 2781-2796, 2008.
- [97] Wei Gao, Jie Chen, Cédric Richard, Jianguo Huang, and Rémi Flamary, "Kernel LMS algorithm with forward backward splitting for dictionary learning," 2013 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2013), Vancouver, Canada, March 2013.
- [98] R. Pokharel, S. Seth, and J.C. Principe, "Mixture kernel least mean square," *The 2013 International Joint Conference on Neural Networks (IJCNN)*, pp.17, 49 Aug. 2013.
- [99] P. Pokharel, W. Liu, and J. C. Principe. Kernel LMS. In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing 2007*, volume 3, pages 1421 – 1424, 2007.

- [100] S. Van Vaerenbergh and I. Santamana, "A comparative study of kernel adaptive filtering algorithms", *Digital Signal Processing and Signal Processing Education Meeting (DSPISPE), 2013 IEEE. IEEE*, pp. 181-186

Appendix A: Customer Baseline Load (CBL) Calculation in NYISO

Following are the steps for calculating the CBL:

Average day CBL for weekdays:

Step 1: Establish the CBL window: Beginning the day (This day not included) prior to the event date, select hourly peak data for last 30 days only for those hours that cover the event for which the CBL is being calculated. Multiply the highest peak load value by 0.25 to calculate a seed value. For example, if the event has been called on 9 October from 4 PM to 7 PM then we should consider the peak load of each hour between 4 PM to 7 PM for last 30 days starting from 7 October.

Now within the selected 30 days exclude the data for days when there was an NYISO defined holiday or any day and also the day prior when NYISO called an SCR/EDRP event or any day and also the day prior when resources DADRP curtailment bid was accepted in DAM.

For the remaining days, calculate the simple average of the resource's actual usage for the hours for which the CBL is being calculated. Now if this average value is less than the value of step 1 (Highest Peak * 0.25) than exclude that day.

If the number of days remaining is less than 10 but not less than 5 than continue similar calculation starting from step 2 and take beginning the day as the day before the current beginning day (so 7-1=6 October as the beginning day) to select last 30 days. The process continues till we have 10 days for the CBL window. Please be noted that if in the first round of calculation the remaining number of days are less than 5 than call NYISO for assistance.

Step 2: Establish the CBL basis for weekdays: Identify the 5 days from the CBL window to be used to develop CBL values for each hour of the event. To select 5 days from the CBL window rank the days according to their average daily event period usage level and select top 5 days with the highest average.

Step 3: Hourly average day CBL values for the event: For the selected 5 days from the CBL window, compute the CBL as the simple average of the hourly loads for those hours that cover the event for which the CBL is being calculated.

Average day CBL for weekend days:

Calculation procedures are very much same like of weekdays except for following changes:

Step1: Establish the CBL window: CBL window comprises of most recent 3 like weekend days and, there are no exclusions for holidays or event days.

Step 2: Establish the CBL basis for weekends: For each of the selected 3 days, calculate daily average of the values of the hours that corresponds to the event hours. Now rank the days according to their Average Daily Event Period Usage level. Eliminate the day with lowest average value. Thus Weekend CBL basis has only 2 days.

Step 3: Hourly average day CBL values for event: For each hour of the event, the CBL value is average of Load in that hour in the two days that comprise the CBL basis

Adjusted Customer Base Line Load

Step 1: Calculate the Average Day CBL values: Calculate the Average Day CBL values for each hour of the event period as described in the previous 'Customer Base Line Load' section.

Step 2: Calculation of Final Event Adjustment Factor: The hourly average day CBL values are multiplied by this factor. The calculation involves following steps:

Calculate Adjustment Basis Average CBL

It is simply the average of the usages of the two hour period beginning with the start of the hour that is four hours prior to the commencement of the event through the end of the hour three hours prior to the event or in other words the two hours (also termed as *adjustment period*) prior to the notification, over a period of 5 days of selected from the CBL window. For example, if the event notification is sent at 10 am than Adjustment Basis Average CBL is the average of usages of hours ending 8 and 9, for the 5 days selected out of CBL window.

Calculate Adjustment Basis Average Load

The Adjustment Basis Average Load is the simple average of the resource's load over the two-hour adjustment period on the event day

Calculate Gross Adjustment Factor

The Gross Adjustment Factor is equal to the Adjustment Basis Average Load divided by the Adjustment Basis Average CBL

Calculate Final Adjustment Factor

The final adjustment factor will be defined according to following rules:

If the Gross Adjustment Factor is greater than 1.00, then the Final Adjustment Factor is the lesser of the Gross Adjustment Factor or 1.20

If the Gross Adjustment Factor is less than 1.00, the Final Adjustment Factors is the greater of the Gross Adjustment Factor or .80

If the Gross Adjustment Factor is equal to 1.00, the Final Adjustment Factor is equal to the Gross Adjustment Factor

Calculate the Adjusted CBL values

The Event Adjusted CBL value for each hour of an event is the product of the Final Adjustment Factor and the Average CBL value for that hour