Exploiting Emerging Data for Enhanced Load Modeling

Final Project Report

Power Systems Engineering Research Center

Empowering Minds to Engineer
the Future Electric Energy System

Exploiting Emerging Data for Enhanced Load Modeling

Final Project Report

Project Team

Santiago Grijalva, Project Leader
Xiaochen Zhang and Elias Khalil, Graduate Students
Georgia Institute of Technology

Ali Abur, Faculty Investigator
Alireza Rouhani, Graduate Student
Northeastern University

Colin Christy, Adjunct Assistant Professor
Ashok Rajan, Graduate Student
Iowa State University

PSERC Publication 14-10

September 2014

For information about this project, contact

Santiago Grijalva
Van Leer E284
Georgia Tech
777 Atlantic Dr. N.W.
E-Mail: sgrijalva@ece.gatech.edu
Phone: (404)-894-2974

Power Systems Engineering Research Center

The Power Systems Engineering Research Center (PSERC) is a multi-university Center conducting research on challenges facing the electric power industry and educating the next generation of power engineers. More information about PSERC can be found at the Center's website: http://www.pserc.org.

For additional information, contact:

Power Systems Engineering Research Center
Arizona State University
527 Engineering Research Center
Tempe, Arizona 85287-5706
Phone: 480-965-1643
Fax: 480-965-0745

Notice Concerning Copyright Material

Acknowledgements

This is the final report for the Power Systems Engineering Research Center (PSERC) research project titled “Exploiting Emerging Data for Enhanced Load Modeling” (project S-49). We express our appreciation for the support provided by PSERC’s industry members and by the National Science Foundation under the Industry / University Cooperative Research Center program.

We thank our industry advisors:

- Kwok Cheung, ALSTOM Grid
- Olivier Despouys, RTE
- Alan Engelman, ComEd
- Jay Giri, ALSTOM Grid
- Sebastien Henry, RTE
- Avnaesh Jayantilal, ALSTOM Grid
- James Kleitsch, American Transmission Company
- Naim Logic, Salt River Project
- Art Mander, Tri-State G&T
- Mirrasoul Mousavi, ABB
- Paul Myrda, EPRI
- Matthew Olearczyk, EPRI
- Patrick Panciatici, RTE
- Michael Swider, New York ISO
- Bill Timmons, Western Area Power Administration
- Robert Uluski, EPRI.

Executive Summary

Massive amounts of novel data are currently being acquired and stored as part of ongoing electricity grid transformation efforts. This data will enable discovery of complex system behavior and enable new decision-making processes to realize higher grid reliability, economy, and sustainability objectives. This project presents various mechanisms to systematically exploit the new data, advancing the current understanding of the load.

Part I: Real-Time Dynamic Parameter Estimation for an Exponential Dynamic Load Model (Abur, Northeastern University)

Load modeling has been an important area of investigation due to the importance of loads as part of the network model used in various different power system studies. Dynamic behavior of loads in particular is of special interest in studies involving power system dynamics. This report is also concerned about real-time modeling and identification of dynamically changing loads in power systems. The motivation for the study is the availability of synchronized measurements which can be used to identify the composite behavior of loads behind a measured bus.

An exponential dynamic load model was proposed earlier and was well accepted by several investigators who worked on this topic. This work considers this model and identifies its parameters in real-time by using measurements. An Unscented Kalman Filter (UKF) is used to track the unknown parameters of the exponential dynamic load model.

The report first implements and tests the proposed method using simulated measurements. The method is then applied to actual recorded utility measurements to identify and track the bus load of a utility feeder.

The results suggest that the proposed approach can provide reasonably accurate dynamic loads for on-line applications requiring detailed load models.

Part II: Exploiting Smart Meter Data for Enhanced Load Modeling (Grijalva, Georgia Institute of Technology)

As part of the ongoing smart grid transformation, smart meters have been widely installed producing massive amounts of data and information. One of the critical needs for distribution system operations and planning applications is modeling of the load, in particular, its dependence on the voltage. This study is aimed at using smart meter measurements for enhanced load modeling by using data-mining methods.

The two major barriers for a data-mining-based load model are the load's time-variant properties and the low resolution sampling rate of the current advanced metering infrastructure (AMI). We address the first barrier through data aggregation and hour partitioning processes. We address the second barrier by introducing the load condition assumption, which justifies the data-mining-based modeling method intuitively from the statistical point of view. Meanwhile, various data-mining and machine learning algorithms are evaluated such as K-subspace method, Davies-Bouldin Index (DBI) and Silhouette Coefficients.

In the first section, we introduce the smart meter deployment on the Georgia Tech main campus as the testbed for the study. We developed an interactive visualization tool, "Smart Grid Plotter", for easier visualization of the cumulative smart meter database. The visualization tool allows researchers to navigate historical data collected by smart meters for all buildings on campus. Users can further configure and save the desired plot through various parameters on the menu.

In the second section, we propose a novel enhanced load modeling method based on data-mining and machine learning algorithms. The enhanced load model is a time-variant model that writes the load's active and reactive power usage as a function of both time and voltage. The detailed steps for the new load modeling method are further discussed in details through three aspects: data aggregation, hour partition and the load condition assumption.

In the third section, we further explore the smart meter data for both off-line and real-time utility functions. In the report, we show that as a very important information source, smart meter data (both real time data and historical data) can be the core of other 18 potential applications when combined with other data, such as weather and GIS information. Two sample applications, refined power flow analysis and dynamic distribution network reconfiguration, are studied to show how smart meter data and the proposed enhanced load model improve power system analysis results and facilitate advanced energy efficiency operations.

In the future, the smart meter data will be more tightly integrated into the vast majority of utility applications for both energy efficiency and reliability improvements. One of the immediate integrations includes a next generation customer information system (CIS) based on smart meter database, weather information and GIS data.

Part III: Exploiting Weather and Load Recording Data to Enhance Load Modeling (Christy, Iowa State University)

In Part III, we show a method for calculating the AC motor load using historical load and temperature data. Historical temperature data is now readily available from internet sources, and historical load data is more and more available through various recording means. The concepts have been illustrated using the load for various companies in the PJM interconnection, but the same concepts can be applied at the feeder level or even at the individual customer level.

Estimation of the AC fraction of load was performed using change-point curves and it was shown that separate change-point curves should be constructed for each different system loading state: weekday daytime, weekday nighttime, weekend daytime, and weekend nighttime. Of course, construction of these curves has been programmed so that the process is automated. Furthermore, a straightforward program can be written to estimate the AC fraction of load at any particular time of interest. The steps for this application were shown.

Naturally, the amount of motor load will depend on the outdoor ambient temperature at the time of the event. The higher the temperature, the higher the AC motor load will be. These results use historical load and temperature data to take the guess work out of estimating the AC motor fraction of load.

Project Publications:

[1] A. Rouhani and A. Abur, "Improving performance of dynamic state estimators under unknown load changes," in *Proc. IEEE Power & Energy Society General Meeting,* July 21-25, 2013*, Vancouver, CA.*

[2] A. Rouhani and A. Abur, "Distributed implementation of an augmented state dynamic estimator," in *Proc. IEEE North American Power Symposium* Sep. 22-24, 2013, *KS, USA.*

[3] Xiaochen Zhang, Grijalva, S., Mathew Reno, "A Time-Variant Load Model Based on Smart Meter Data Mining", accepted by IEEE PES 2014 General Meeting.

Table of Contents

Part 1. Real-Time Dynamic Parameter Estimation for an Exponential Dynamic Load Model 1

1.1 Load Modeling in Power Networks 1

1.1.1 Introduction 1

1.1.2 Project Objectives and Description 3

1.2 Proposed Algorithms and Dynamic Models 3

1.2.1 Exponential Dynamic Load Model 4

1.2.2 Unscented Transformation 6

1.2.3 Unscented Kalman Filter (UKF) 8

1.3 Proposed Approach 10

1.3.1 Implementation of the Proposed Approach Using Simulated Measurements 11

1.3.2 Implementation of the Proposed Approach Using Actual Recorded Measurements 15

1.4 Conclusions and Future Work 23

Part 2. Exploiting Smart Meter Data for Enhanced Load Modeling 25

2.1 Introduction 25

2.1.1 Background 25

2.1.2 Overview of the Problem 25

2.1.3 Report Organization 28

2.2 Data Background and the Visualization Tool 28

2.2.1 Data Background and Analysis 28

2.2.2 Interactive Visualization Tool 30

2.3 Enhanced Load Modeling 32

2.3.1 Data Aggregation 33

2.3.2 Hour Partition 35

2.3.3 Load Condition Assumption 39

2.4 Exploiting Smart Meter Data for Utility Functions 44

2.4.1 Smart Meter Data Application for Online and Offline Operations 45

2.4.2 Application Examples 48

2.5 Conclusions 53

Part 3. Exploiting Weather and Load Recording Data to Enhance Load Modeling 55
3.1 Research Objective 55
3.2 Load and Temperature Relationships 56
3.3 Segmented Regression to Create the Change-Point Curve 60
3.4 Calculation of Air Conditioning (AC) Load Fraction 62
3.5 Temperature-Humidity Index as a Potential Weather Variable 63
3.6 What Causes Separation in the Data Cloud? 64
3.7 Change-point Curves for Four Different System States 66
3.8 Application to System Studies 69
3.9 Conclusions 70
3.10 Future Work 71
References 72

List of Figures

Figure 1.1 Load response under a voltage step change 5
Figure 1.2 A comparison of the randomly generated points. 8
Figure 1.3 Actual and estimated plot of P_r -Scenario 1. 12
Figure 1.4 Actual and estimated plot of α_s -Scenario1. 12
Figure 1.5 Actual and estimated plot of α_t -Scenario 1. 13
Figure 1.6 Actual and estimated plot of T_p -Scenario 1. 13
Figure 1.7 Actual and estimated plot of P_r -Scenario 2. 14
Figure 1.8 Actual and estimated plot of α_s -Scenario 2. 14
Figure 1.9 Actual and estimated plot of α_t -Scenario 2. 14
Figure 1.10 Actual and estimated plot of T_p -Scenario 2. 15
Figure 1.11 Voltage of the distribution feeder (V). 16
Figure 1.12 Plot of estimated values for P_r. 17
Figure 1.13 Plot of estimated values for α_s. 17
Figure 1.14 Plot of estimated values for α_t. 18
Figure 1.15 Plot of estimated values for T_p. 18
Figure 1.16 Plot of actual and estimated values for P_l. 19
Figure 1.17 Plot of MSE when tracking total real power. 20
Figure 1.18 Plot of estimated values for Q_r. 20
Figure 1.19 Plot of estimated values for β_s. 21
Figure 1.20 Plot of estimated values for β_s. 21
Figure 1.21 Plot of estimated values for T_q. 22
Figure 1.22 Plot of actual and estimated values for Q_l. 22
Figure 1.23 Plot of MSE when tracking total reactive power. 23
Figure 2.1 Enhanced load modeling method 27
Figure 2.2 ION webreach main menu , list of buildings, and a building menu 29
Figure 2.3 ION webreach meter interface 30
Figure 2.4 Structure of “Smart Grid Plotter” 30
Figure 2.5 Smart grid plotter 31

Figure 2.6 Zoom in/out 31
Figure 2.7 Q-V plot of a student residence hall 32
Figure 2.8 The enhanced load modeling method 33
Figure 2.9 Tree-structured data labels 35
Figure 2.10 Time index adds to the complexity of the model 36
Figure 2.11 Hour partitioning serves to reduce the model complexity 36
Figure 2.12 Hour grouping method (clustering vector formulation) 37
Figure 2.13 Hour grouping results for an apartment building (winter,weekdays) 38
Figure 2.14 Voltage deviations Vs. customer behaviors 39
Figure 2.15 Load condition assumption 40
Figure 2.16 P-V and Q-V plot for a commercial building in summer night 41
Figure 2.17 Comparisons between K-subspace method and K-means method 42
Figure 2.18 The load condition clustering based on hour partition results 42
Figure 2.19 Silhouette coefficients for K-subspace method 44
Figure 2.20 Comparison between K-subspace method and K-means method 45
Figure 2.21 Baran & Wu 33-Bus test system 48
Figure 2.22 Refined power flow analysis flow chart 50
Figure 2.23 Refined power flow results comparison 51
Figure 2.24 System loads during working hours 52
Figure 2.25 System loads during off-working hours 52
Figure 2.26 Working hours configuration 53
Figure 2.27 Off-working hour configuration 53
Figure 3.1 Load model for system simulation.. 55
Figure 3.2 Normalized load and temperature for a single company for a full year.. 57
Figure 3.3 Real power versus the outdoor ambient temperature.. 57
Figure 3.4 Normalized load and normalized temperature plotted against time. 58
Figure 3.5 Cross-correlation of (real power) load and temperature 59
Figure 3.6 Load versus temperature before and after a shift of the temperature data. 59
Figure 3.7 Segmented regression withand without shifting the temperature data.. 60
Figure 3.8 The high temperature portion of the data. 61
Figure 3.9 RMSE of linear fit before and after shifting temperature data 61
Figure 3.10 Temperature-independent and temperature-dependent fractions of load 62
Figure 3.11 Two different weather variables for estimating the AC fraction. 63

Figure 3.12 Comparison of weather variables. 64
Figure 3.13 Separation of the data cloud at medium to low temperatures. 65
Figure 3.14 Separation of the data points into different day types. 65
Figure 3.15 Average daily load profile (time in in ½ hour increments). 66
Figure 3.16 Data points separated into daytime load (blue) and nighttime load (red). ... 66
Figure 3.17 Load divided into four different system states. 67
Figure 3.18 A change-point curve created for each different loading state. 68
Figure 3.19 Change point curves for all data 68
Figure 3.20 Application for system studies.. 70

List of Tables

Table 2.1 Comparisons of measurement-based and component-based method 26
Table 2.2 Time-variant load model and data structure 34
Table 2.3 Time label identification results (weekdays, fall)..... 39
Table 2.4 Test system bus data 49
Table 2.5 Test system line data..... 49
Table 3.1 AC fraction of load calculated at 40°C for each system loading state 69

Part 1. Real-Time Dynamic Parameter Estimation for an Exponential Dynamic Load Model

1.1 Load Modeling in Power Networks

1.1.1 Introduction

Obtaining more accurate load models which properly reflect dynamic behavior of loads under various disturbances is one of the challenges in today's energy management systems. Various on-line applications that rely on dynamic simulation studies require detailed and accurate load models. Use of models that fail to accurately capture the dynamic behavior of loads may lead to inconsistent results for dynamic stability and voltage collapse studies [1-5]. A load model is a mathematical representation related to the measured voltage and/or frequency at a bus, and the real and reactive power consumed by the load [6]. Hence, load modeling is considered as a system identification problem.

As expected the topic of load modeling occupies a large volume in power systems literature. Proposed load modeling approaches can be broadly classified into two categories: Component-based [7] and measurement-based [8, 9] approaches. The drawback of the first category is that it requires full knowledge of the load inventory of typical loads in order to synthesize composite load models. Thus, successful implementation of this approach strongly depends on the true inventory of the loads connected to the feeders which is regretfully not always available [10]. The second category estimates load parameters using measurements. This gives a more precise picture of real-time loads and their dynamic characteristics [11]. This project's approach falls in this second category.

Load models can be broadly classified as either static or dynamic [6]. A static load model does not depend on time [12], and therefore it relates the active and reactive power at a given time to the voltage and /or frequency at the same instant of time. Static load model is suitable to represent static load components such as resistive loads and light bulbs. While they are also used to approximate the dynamic load components, their accuracy is

usually not sufficiently high. On the other hand, a dynamic load model describes the load behavior as a function of time and therefore provides a much more accurate tool for dynamic simulations.

This work considers one of the most widely accepted dynamic load models and aims to identify and track its parameters on-line. This is the exponential dynamic load model proposed and described in [8]. Aggregate load model (ZIP augmented with induction motor) is also considered by several researchers [13-16]. Perhaps one of the main shortcomings associated with this model is that, it has considerably more unknown parameters and state variables to be identified and estimated compared to the exponential dynamic load model. Computational burden associated with the estimation of the large number of parameters significantly prohibits the real-time implementation of this model. In order to overcome this limitation, an Extended Kalman Filter (EKF) based technique is used to estimate the dominant parameters of the aggregate load model assuming that the other parameters can be approximated for different types of loads [15]. This approximation however leads to reduced accuracy of the load parameter identification. Other alternatives such as the hybrid learning algorithm which combines the genetic algorithm and nonlinear Levenberg-Marquardt algorithm [16] have also been proposed for parameter identification of the aggregate load model. All of these approaches share the same limitation of high computational burden as a real-time application. In [17] two different approaches for PSS/E CLOD complex load model parameter estimation are investigated. The first approach, compare and re-simulate, solves for the load model using a generic nonlinear minimization routine. This approach suffers from long run times and it is vulnerable to measurement error and noise. The second approach is called simulate then calculate which emulates the time-consuming simulation process using a simple matrix manipulation, which reduces the computation time significantly, but also decreases the accuracy of the solution and it is sensitive to the distribution which is considered for the parameter set.

1.1.2 Project Objectives and Description

This work is motivated by the decisive effect that load representation has on voltage stability studies. One of the main objectives of this work is to find the most updated load model in real-time which can represent the latest characteristics of the corresponding actual load. Consequently, two important factors should be taken into the account. First, the chosen load model should be capable enough to represent the dynamic behavior of the actual load. Second, simultaneously it should not be too complex. Since as the load model becomes more complex (more unknown parameters) the computational effort for load identification will also increase.

Considering different load models proposed by researchers in the literature, the exponential dynamic load model is found to satisfy the conditions mentioned above. The mathematical representation of this load model is described in this work and the corresponding parameters which are needed to be identified are indicated.

Next, it is necessary to implement a dynamic state estimator to track the parameters of the exponential dynamic load model. In this project UKF is implemented as a dynamic state estimator where an Unscented Transformation is used for obtaining the propagated mean and covariance of the state vector, leading to a better performance than EKF which is widely used for dynamic state estimation in the literature [18]. The algorithms associated with Unscented Transformation and UKF are also explained in this work.

This report is organized in four sections. Section 2 includes the details of the algorithms and dynamic models which are used in this work. Section 3 presents the proposed approach and the simulations results where first based on the simulated measurements, the performance of the proposed approach is evaluated considering two different scenarios, then the results are shown based on real measurements coming from utility. The final section will conclude the report and also mention some of the future work.

1.2 Proposed Algorithms and Dynamic Models

In this work, an Unscented Kalman Filter (UKF) is used as a dynamic state/parameter estimator to track the unknown parameters/state variables associated with exponential

dynamic load model. Detailed application of the UKF algorithm to the considered exponential dynamic load model is described below.

1.2.1 Exponential Dynamic Load Model

The assumed load model is expressed as a set of non-linear equations, where real (active) and reactive powers consumed by the load are assumed to be related to the voltage in the following non-linear manner [8], [12]:

$$T_p \frac{dP_r}{dt} + P_r = P_0\left(\frac{V}{V_0}\right)^{\alpha_s} - P_0\left(\frac{V}{V_0}\right)^{\alpha_t}$$

$$P_l = P_r + P_0\left(\frac{V}{V_0}\right)^{\alpha_t} \qquad (1.2.1)$$

where:

V_0 and P_0 are the voltage and power consumption before a voltage change.

P_r is the active power recovery,

P_l is the total active power response,

T_p is the active load recovery time constant,

α_t is the transient active load-voltage dependence, and

α_s is the steady state active load-voltage dependence.

Similar equations are also valid for reactive power. The equations related to reactive power part are given below:

$$T_q \frac{dQ_r}{dt} + Q_r = Q_0\left(\frac{V}{V_0}\right)^{\beta_s} - Q_0\left(\frac{V}{V_0}\right)^{\beta_t}$$

$$Q_l = Q_r + Q_0\left(\frac{V}{V_0}\right)^{\beta_t} \qquad (1.2.2)$$

Similarly,

V_0 and Q_0 are the voltage and reactive power consumption before a voltage change.

Q_r is the reactive power recovery,

Q_l is the total reactive power response,

T_q is the reactive load recovery time constant,

β_t is the transient reactive load-voltage dependence, and β_s is the steady state reactive load-voltage dependence.

As an example Figure 1.1 shows the response of the exponential dynamic load when a disturbance is affecting the system. In this case an ideal voltage step has been applied and as a consequence, the total active power response of the load will reach to a new steady state after recovery [8].

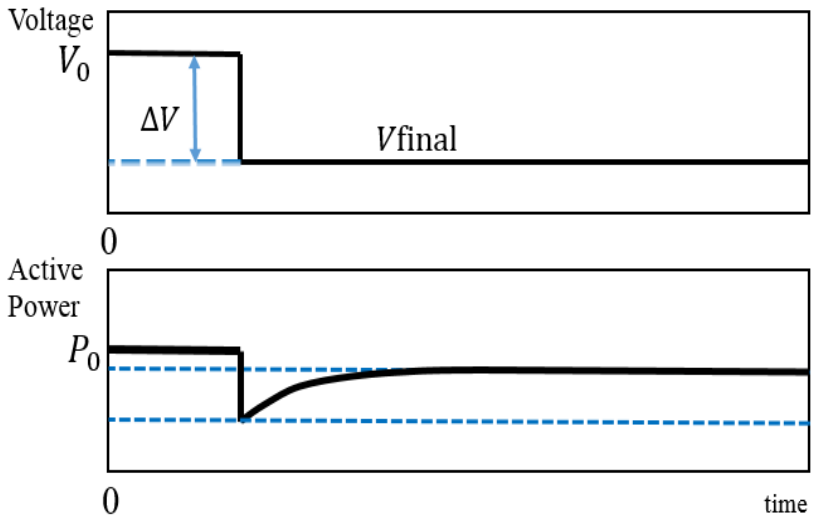

Figure 1.1 Load response of the exponential dynamic load model under a voltage step change

This work is mainly focused on the real-time estimation of the unknown parameters/state variables associated with this load model based on measurements and development of a real-time dynamic load model which represents the behavior of the monitored load with an acceptable accuracy. This is accomplished by using an UKF which is implemented as

a dynamic parameter estimator for the unknown parameters and state variables of the assumed load model. Here the unknown parameters are α_s, α_t *and* T_p (for the real power model of (2.1)) and the state variable is P_r. It is needless to say that an equivalent statement is true for the reactive power model of (2.2).

1.2.2 Unscented Transformation

Given the nonlinearity of the measurement and generator dynamic equations, it is quite cumbersome if not impossible to transform the entire probability density function (pdf) for the modeling and measurement errors. Use of first order approximation as done in EKF will introduce errors which may at times cause significant biases and even divergence [18].

As an alternative, an approximation to the true pdf can be generated by transforming individual points in state space and calculating a sample pdf based on them. This is essentially what an unscented transformation (UT) does. To illustrate the idea, consider a random vector x with mean $\overline{x}$ and covariance P. One can find a set of deterministic vectors called sigma points whose ensemble mean and covariance are equal to $\overline{x}$ and P. Let us also consider the nonlinear measurement function $z = h(x)$ which can be used to obtain measurement points by substituting the sigma points for x in the measurement equations. The ensemble mean and covariance of the transformed vectors will give a good estimate of the true mean and covariance of z.

Suppose that x is an $n \times 1$ vector that is transformed by a nonlinear function $z = h(x)$,The unscented transformation base on Choosing $2n$ sigma points is mentioned by following steps:

$$
\begin{aligned}
x^{(i)} &= \overline{x} + x_*^{(i)} \qquad i = 1,\ldots,2n \\
x_*^{(i)} &= (\sqrt{nP})_i^T \qquad i = 1,\ldots,n \\
x_*^{(n+i)} &= -(\sqrt{nP})_i^T \qquad i = 1,\ldots,n
\end{aligned}
\tag{1.2.3}
$$

Please note that $(\sqrt{nP})_i$ is the i th row of $\sqrt{nP}$. Transform the sigma points as follows:

$$z^{(i)} = h(x^{(i)}) \; i = 1,\ldots,2n \tag{1.2.4}$$

The approximated mean and covariance of z can then be obtained as follows:

$$\bar{z}_u = \frac{1}{2n} \sum_{i=1}^{2n} z^{(i)} \tag{1.2.5}$$

$$P_u = \frac{1}{2n} \sum_{i=1}^{2n} (z^{(i)} - z_u)(z^{(i)} - z_u)^T \tag{1.2.6}$$

The unscented transformation mentioned above (which is based on choosing $2n$ sigma points) is not the only one that exists. As an example, it can be shown that general unscented transformation which uses $2n+1$ sigma points to obtain the propagated mean and covariance, gives the same order of mean and covariance estimation accuracy as given by unscented transformation base on $2n$ sigma points [18]. For computational saving purposes, we can use other type of transformations which use less number of sigma points. It can be shown [18] that minimum number of sigma points which makes it possible to obtain the propagated mean and covariance is equal to $n+1$.

As an example the superior performance of the unscented transformation based on $2n$ sigma points in comparison with linear approximation (which is used in EKF) for finding the propagated mean and covariance of a state vector is shown in the Figure 1.2. In this figure 300 points are randomly generated for the following set of equations:

$$\begin{aligned} y_1 &= r\cos\theta \\ y_2 &= r\sin\theta \end{aligned} \tag{1.2.7}$$

where r is uniformly distributed between ±0.1 and θ is uniformly distributed between ±0.35 rad.

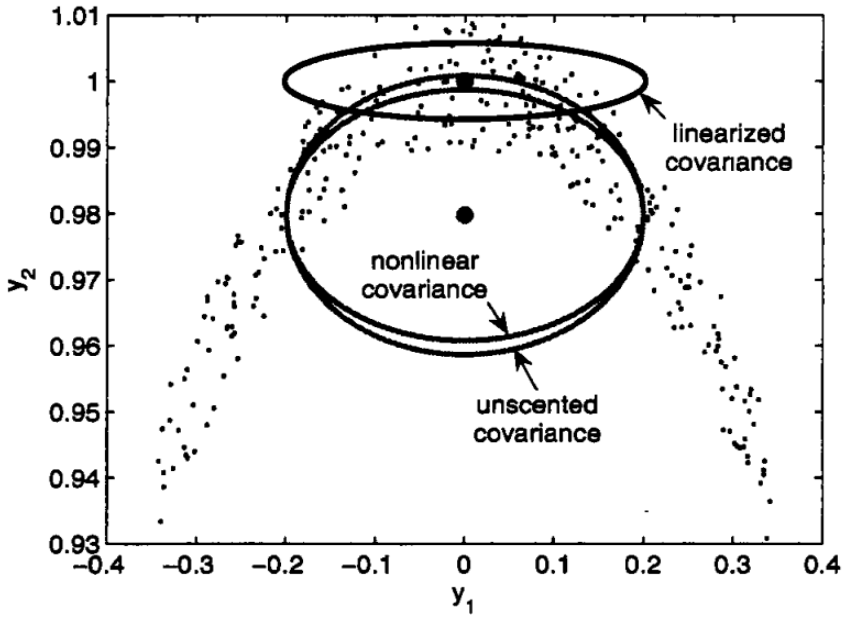

Figure 1.2 A comparison of the exact, linearized, and unscented mean and covariance of 300 randomly generated points based on (2.7).

The small points in Figure 1.2 represent the exact points generated by (2.7) and the bold point at (0,1) shows the linearized mean. The true mean and the approximate unscented mean are very close that they are plotted on top of each other and are both equal to (0, 0.97). Hence, this figure validates the highly accurate estimates of the mean and the covariance when unscented transformations are used instead of linear approximations [18].

In this work we will consider the standard Unscented Kalman Filter or (UKF) which uses $2n$ sigma points. In the next part we summarize the UKF algorithm.

1.2.3 Unscented Kalman Filter (UKF)

The UKF uses the unscented transformation [18] for solving nonlinear problems by considering system dynamics and measurement equations as follows:

$$x_{k+1} = f(x_k, k) + w_k \tag{1.2.8}$$

$$z_k = h(x_k, k) + v_k \tag{1.2.9}$$

where

$x \in \Re^n$ is a discrete state vector.

$z \in \Re^m$ is a discrete measurement vector.

$w_k \sim N(0, Q_k)$ Gaussian process noise at time step k

$v_k \sim N(0, R_k)$ Gaussian measurement noise at time step k

Q_k and R_k are covariance matrices of w_k and v_k respectively.

The UKF is initialized as follows:

$$\hat{x}_0^+ = E(x_0), \quad P_0^+ = E[(x_0 - \hat{x}_0^+)(x_0 - \hat{x}_0^+)^T] \tag{1.2.10}$$

Time update equations are:

(a) Calculation of sigma points:

$$\hat{x}_{k-1}^{(i)} = \hat{x}_{k-1}^+ + x_*^{(i)} \qquad i = 1,\ldots,2n$$

$$x_*^{(i)} = (\sqrt{nP_{k-1}^+})_i^T \qquad i = 1,\ldots,n$$

$$x_*^{(n+i)} = -(\sqrt{nP_{k-1}^+})_i^T \qquad i = 1,\ldots,n \tag{1.2.11}$$

(b) $\hat{x}_k^{(i)} = f(\hat{x}_{k-1}^{(i)}, k-1)$, $\hat{x}_k^- = \dfrac{1}{2n}\sum_{i=1}^{2n} \hat{x}_k^{(i)}$

$$P_k^- = \frac{1}{2n}\sum_{i=1}^{2n} (\hat{x}_k^{(i)} - \hat{x}_k^-)(\hat{x}_k^{(i)} - \hat{x}_k^-)^T + Q_{k-1} \tag{1.2.12}$$

The measurement update equations are as follows:

(a) Calculation of sigma points:

$$\hat{x}_k^{(i)} = \hat{x}_k^- + x_*^{(i)} \qquad i = 1,\ldots,2n$$

$$x_*^{(i)} = (\sqrt{nP_k^-})_i^T \qquad i = 1,\ldots,n$$

$$x_*^{(n+i)} = -(\sqrt{nP_k^-})_i^T \qquad i = 1,\ldots,n \tag{1.2.13}$$

In order to save computational effort step (2.13) can be omitted [18] with a slight degradation in filter performance.

(b) $\hat{z}_k^{(i)} = h(\hat{x}_k^{(i)}, k)$, $\hat{z}_k = \frac{1}{2n}\sum_{i=1}^{2n} \hat{z}_k^{(i)}$

$$P_z = \frac{1}{2n}\sum_{i=1}^{2n}(\hat{z}_k^{(i)} - \hat{z}_k)(\hat{z}_k^{(i)} - \hat{z}_k)^T + R_k$$

$$P_{xz} = \frac{1}{2n}\sum_{i=1}^{2n}(\hat{x}_k^{(i)} - \hat{x}_k^-)(\hat{z}_k^{(i)} - \hat{z}_k)^T$$

$$K_k = P_{xz}P_z^{-1}$$

$$\hat{x}_k^+ = \hat{x}_k^- + K_k(z_k - \hat{z}_k)$$

$$P_k^+ = P_k^- - K_k P_z K_k^T \tag{1.2.14}$$

1.3 Proposed Approach

State dynamics given by (2.8) and the measurement equations given by (2.9) can be developed for the considered exponential dynamic load model given in (2.1) by discretizing the equations using the second order Runge-Kutta method which is a numerically stable discretization method [5] as follows:

$$P_r(k+1) = P_r(k) + \frac{A+B}{2}$$

$$A = \frac{\Delta t}{T_p(k)} \times (-P_r(k) + P_0(\frac{V(k)}{V_0})^{\alpha_s(k)} - P_0(\frac{V(k)}{V_0})^{\alpha_t(k)})$$

$$B = \frac{\Delta t}{T_p(k)} \times (-(P_r(k) + A) + P_0(\frac{V(k)}{V_0})^{\alpha_s(k)} - P_0(\frac{V(k)}{V_0})^{\alpha_t(k)})$$

$$\alpha_s(k+1) = \alpha_s(k)$$
$$\alpha_t(k+1) = \alpha_t(k)$$
$$T_p(k+1) = T_p(k)$$

$$P_l(k) = P_r(k) + P_0(\frac{V(k)}{V_0})^{\alpha_s(k)} \quad (1.3.1)$$

Please note that the discretized equations associated with the reactive part (2.2) can be obtained in a similar way.

In order to validate the dynamic estimator performance, a set of measurements is created by dynamic simulations on a system with known load model parameters. UKF is then implemented and used to estimate the augmented state vector which includes the state variable and unknown parameters associated with the exponential dynamic load model:

$$x_k = [P_r(k), \alpha_s(k), \alpha_t(k), T_p(k)]^T \quad (1.3.2)$$

where the measurement is assumed to be:

$$z_k = P_l(k) + v_k \quad (1.3.3)$$

1.3.1 Implementation of the Proposed Approach Using Simulated Measurements

Two scenarios are simulated in this section. Both scenarios involve changes in the load model. Measurements are created by using dynamic simulations and subsequently adding Gaussian noise according to the assumed additive noise model of (3.3). These two scenarios are described below.

Scenario 1:

In this scenario, the true parameters associated with the load model are assumed to be (the values are selected from [12]):

$$P_0 = 0.867, \alpha_s = -0.32, \alpha_t = 1.65 \,\&\, T_p = 70[\text{sec}].$$

Here are the detailed steps of the simulated events and assumptions made:

- A voltage drop of $\Delta V / V_0 = -5.3\%$ occurs at the load bus at t = 0.
- Load parameter α_s is increased from -0.32 to -0.28 at t = 5 min.
- Total simulation time is 10 min and time-step=0.025 sec.
- The UKF is initialized using arbitrary values in (2.10).
- $R_k = 1e^{-4}$ and $Q_k = 1e^{-6} \times I_{4\times 4}$.

The results are shown in the following figures. Please note that in all plots, dashed and solid lines correspond to the estimated and true trajectories respectively.

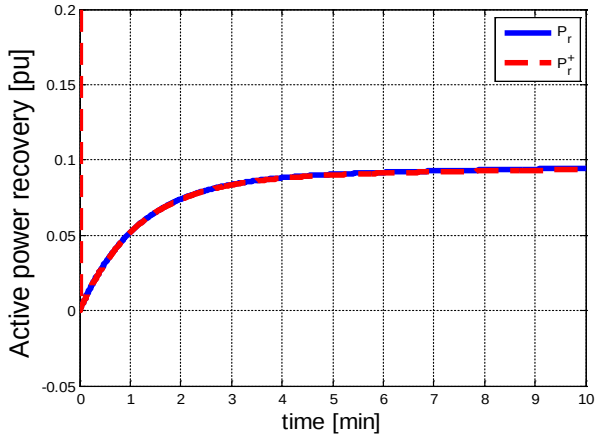

Figure 1.3 Actual and estimated plot of P_r -Scenario 1.

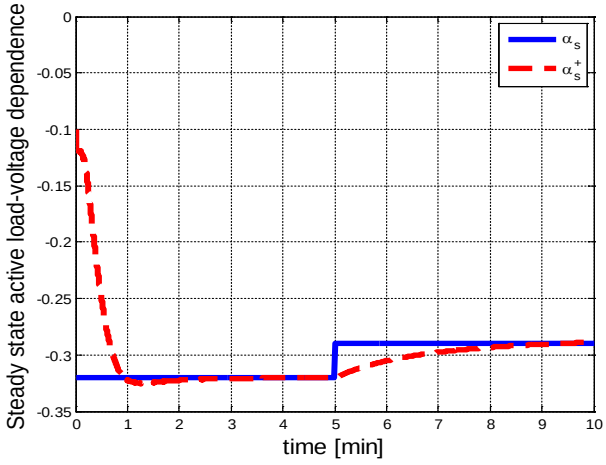

Figure 1.4 Actual and estimated plot of α_s -Scenario1.

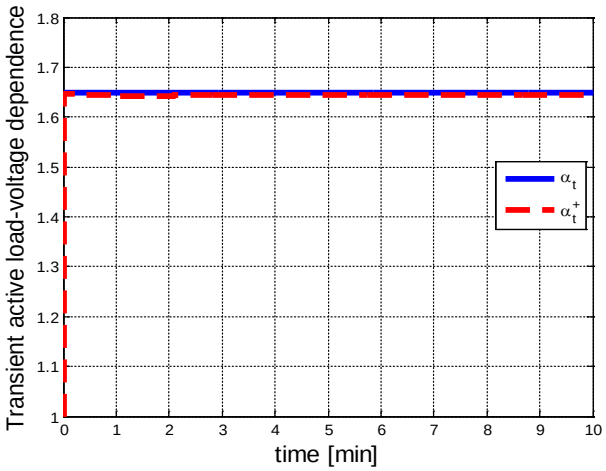

Figure 1.5 Actual and estimated plot of α_t -Scenario 1.

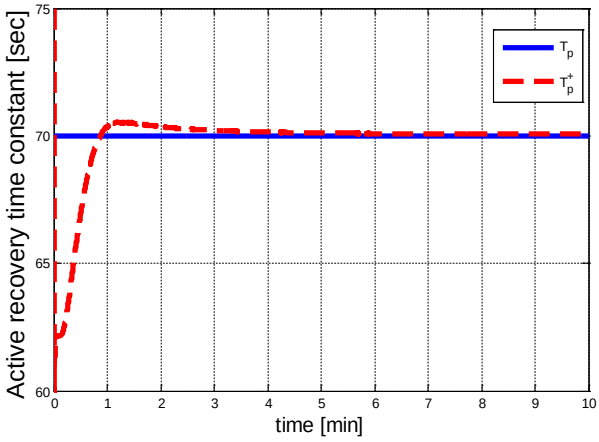

Figure 1.6 Actual and estimated plot of T_p -Scenario 1.

Scenario 2:

This scenario is similar to scenario 1 except for the fact that all load parameters are assumed to change linearly during the simulation. Their ranges of change are assumed as follows:

- α_s is reduced from -0.32 to -0.37.
- α_t is increased from 1.65 to 1.7.
- T_p is increased from 70 sec. to 80 sec.

The results are shown in the following figures.

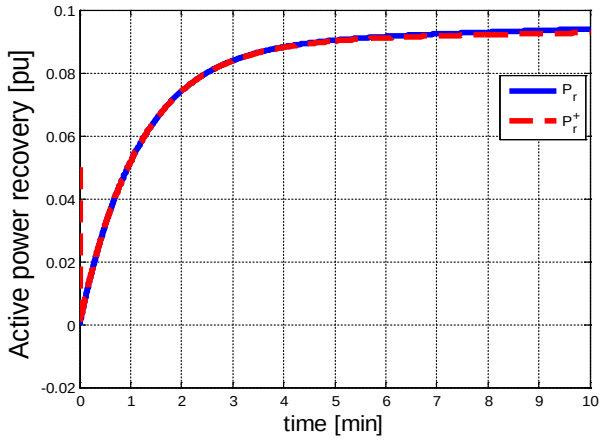

Figure 1.7 Actual and estimated plot of P_r -Scenario 2.

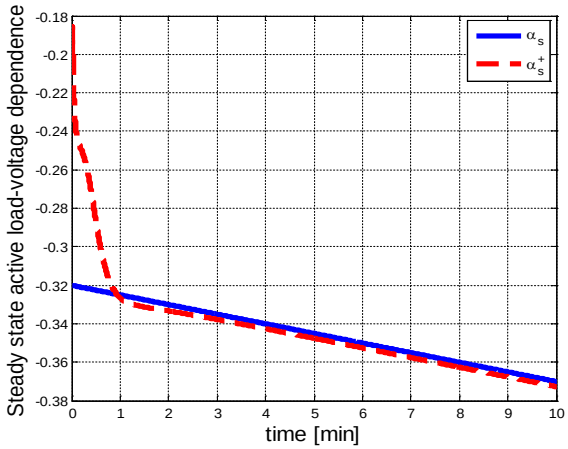

Figure 1.8 Actual and estimated plot of α_s -Scenario 2.

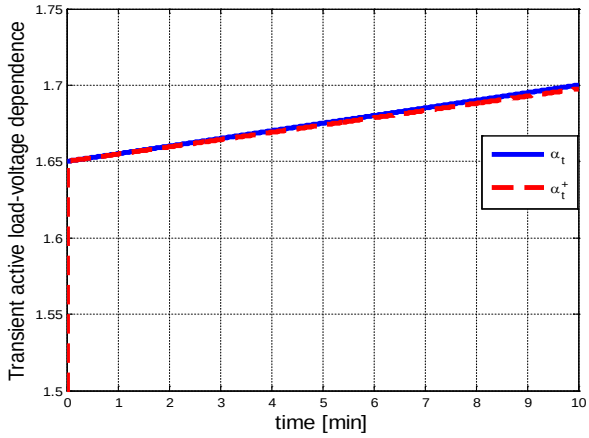

Figure 1.9 Actual and estimated plot of α_t -Scenario 2.

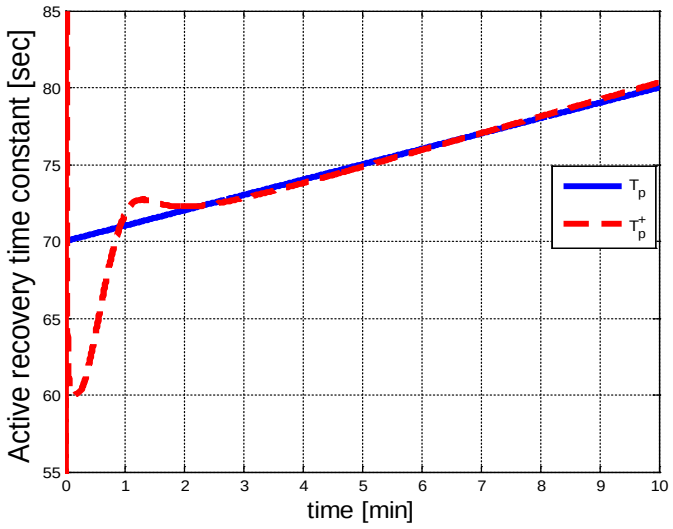

Figure 1.10 Actual and estimated plot of T_p -Scenario 2.

UKF is observed to successfully track changes in model parameters for both scenarios. It is capable of tracking the unknown and time-varying parameters of the exponential dynamic load model accurately when using simulated measurements and an exponential load model. However, in actual system operation, the load model is simply not known and therefore, UKF's performance needs to be tested using actual recorded measurements in order to evaluate its performance as an on-line function. This is done in the next section.

1.3.2 Implementation of the Proposed Approach Using Actual Recorded Measurements

In this section, the proposed approach is evaluated based on actual recorded measurement data where synchronized voltage and power measurements are acquired every 6 seconds for a utility distribution feeder. The total duration of the recordings is 1440 minutes or 24 hours. Figure 1.11 shows that the voltage of the feeder (v) increases during this period (from almost 115 kV to 116.5 kV).

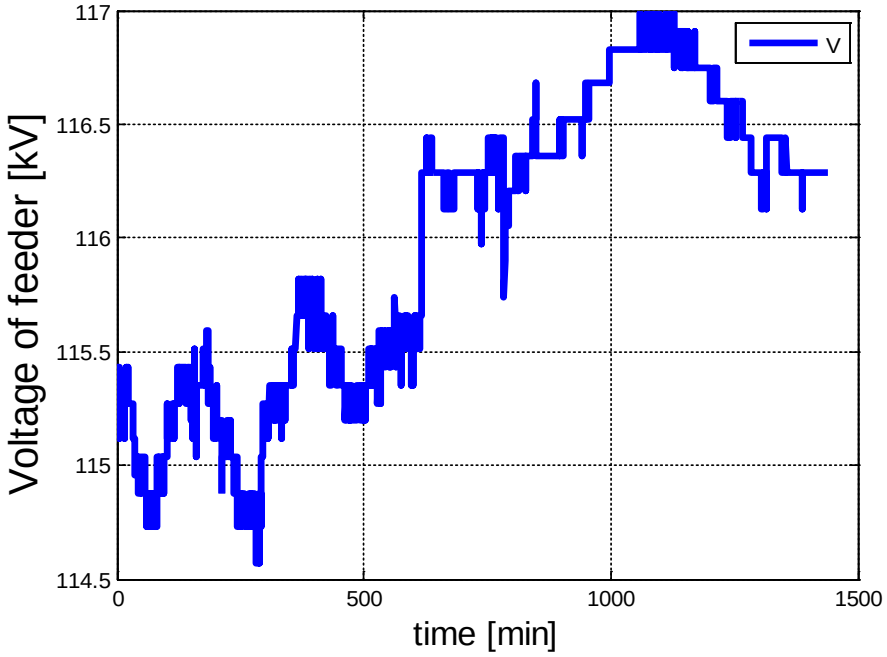

Figure 1.11 Voltage of the distribution feeder (V).

As in section A, UKF is used to estimate the unknown parameters of the assumed exponential dynamic load model. The following assumptions and data are used in implementing the UKF to track parameters of the unknown load based on the recorded measurements:

- A time-step of 6 sec. is used by the filter.
- Total duration of the tracking study is 1440 min.
- UKF is initialized using arbitrary values in (2.10).
- $R_k = 1e^{-4}$ and $Q_k = 1e^{-6} \times I_{4\times 4}$.

Figures 1.12 to 1.15 show the estimated values of the assumed exponential load model for the study duration.

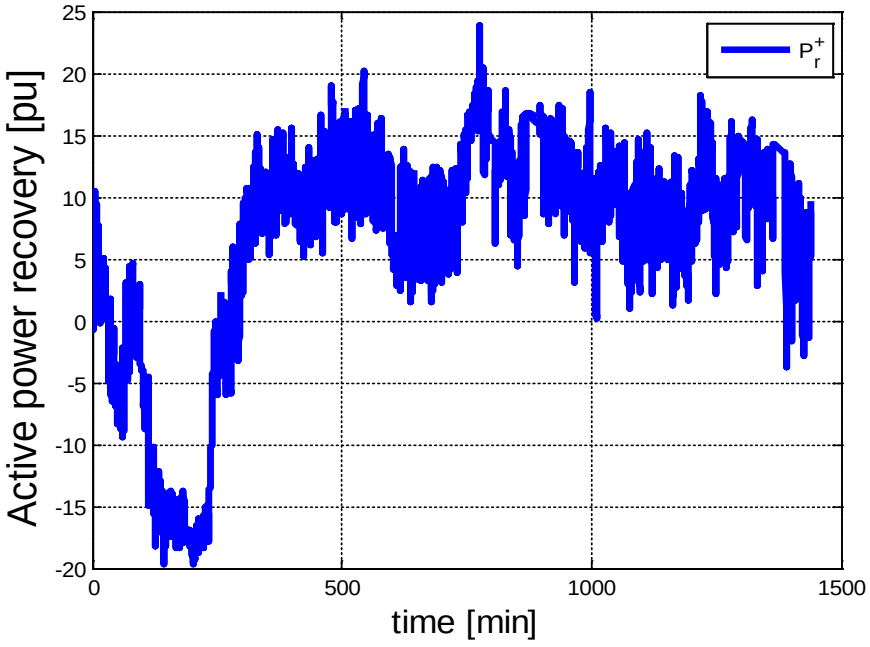

Figure 1.12 Plot of estimated values for P_r .

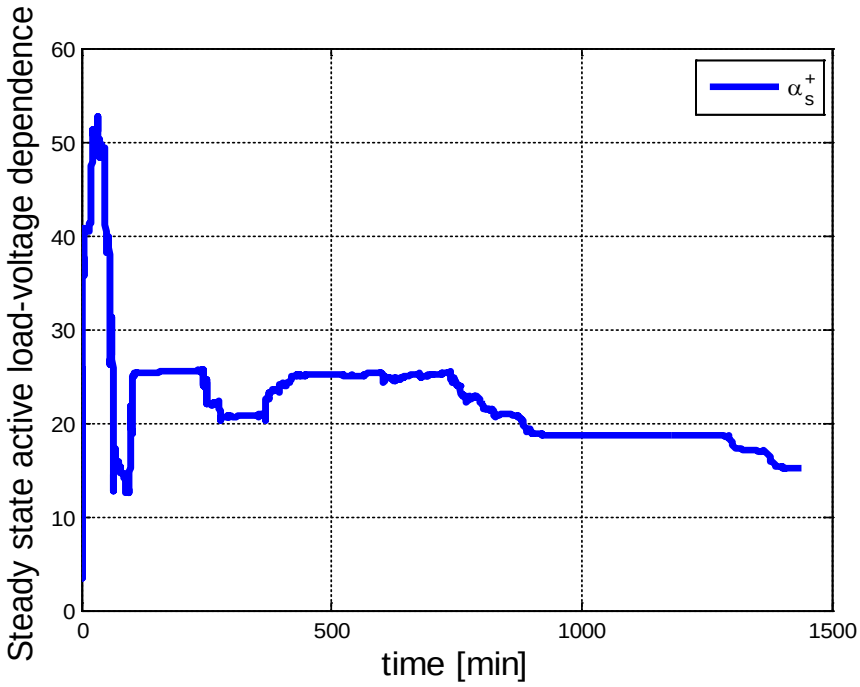

Figure 1.13 Plot of estimated values for α_s .

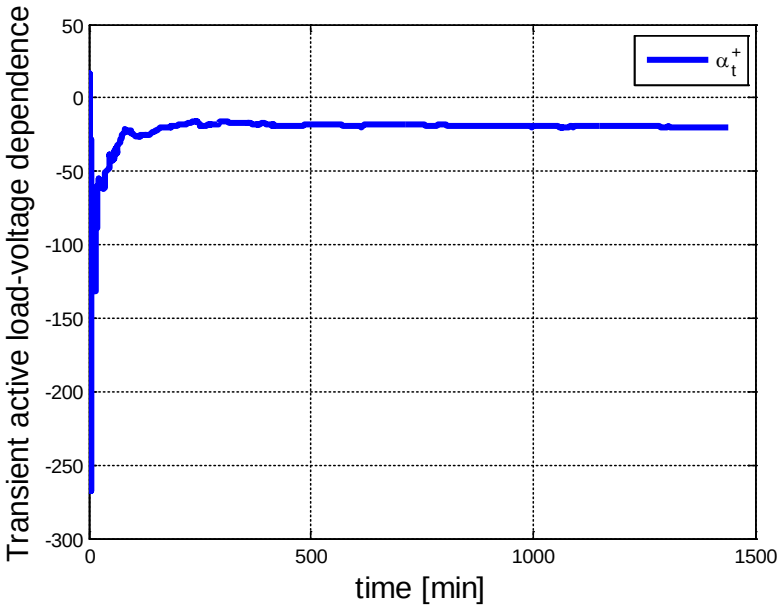

Figure 1.14 Plot of estimated values for α_t.

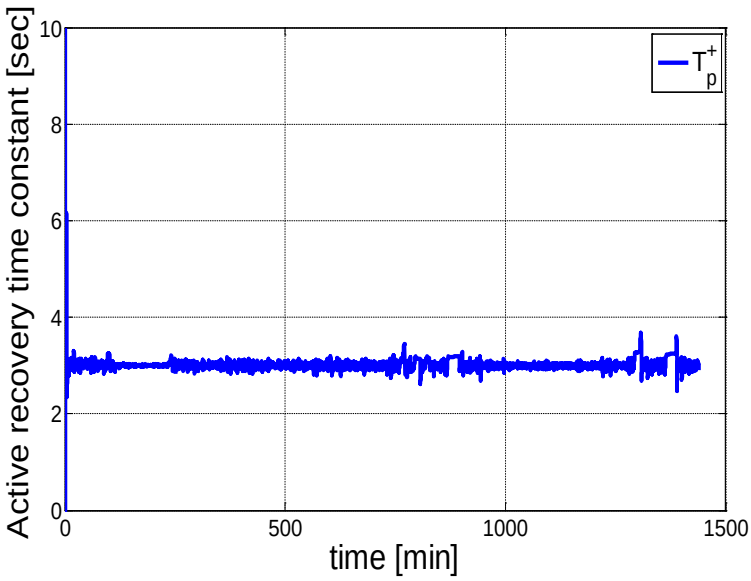

Figure 1.15 Plot of estimated values for T_p.

Note that the true values are not known, so it is not possible to comment on the accuracy of the parameters, however given the long duration of the study time (1440 minutes or 24 hours) one can observe UKF's tracking of parameters as they gradually change in time.

While the actual parameters of the model are unknown and cannot be directly measured, total active and reactive power are available as measured values. Thus, performance of the UKF in tracking the total active power can be evaluated by observing the measured and predicted total active power as shown in Figure 1.16.

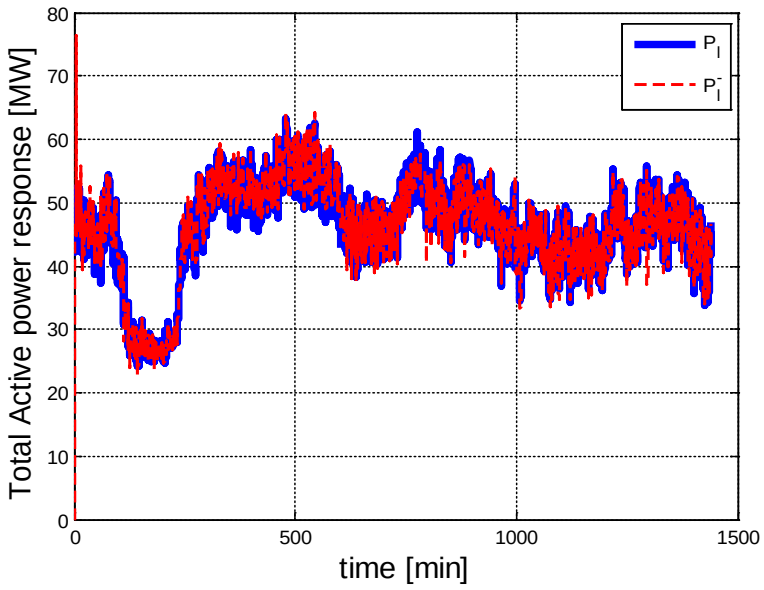

Figure 1.16 Plot of actual and estimated values for P_l.

In order to quantify the accuracy, the following error metric is used to evaluate the performance of the UKF in tracking the real-time model of the active power demand:

$$\sigma_k = \sqrt{\frac{1}{k}\sum_{i=1}^{k}(z_i - \bar{z})^2}$$

$$MSE = \sqrt{\frac{1}{k}\sum_{i=1}^{k}\left(\frac{z_i - \hat{z}_i}{\sigma_k}\right)^2} \qquad (1.3.4)$$

where

z_i and $\hat{z}_i$ are the measured and estimated values at time i.

MSE is the mean squared error.

Figure 1.17 illustrates how the MSE value is gradually reduced finally settling below an acceptable level of 0.25, validating the satisfactory performance of UKF.

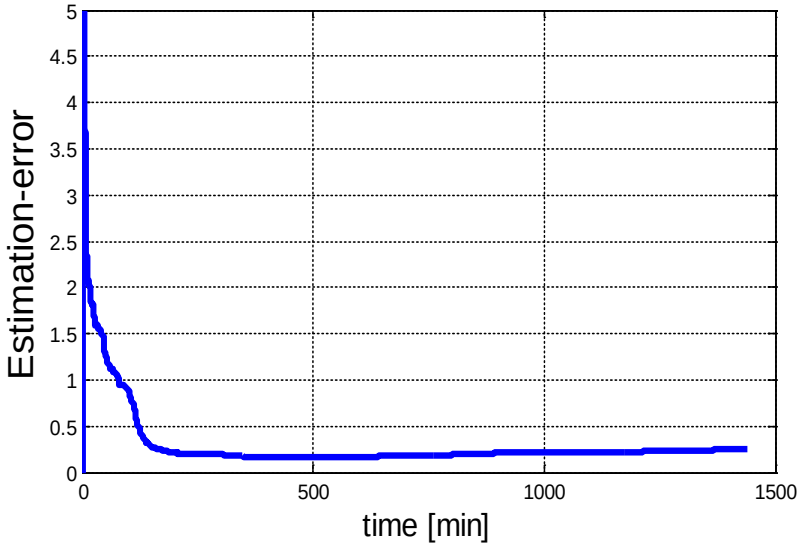

Figure 1.17 Plot of MSE when tracking total real power.

A similar study is repeated for the reactive power model of (2.2), where the model parameters are estimated and plotted as shown in Figures 1.18 to 1.21. Figures 1.19 to 1.21 illustrate how the exponential dynamic model parameters are tracked by the UKF.

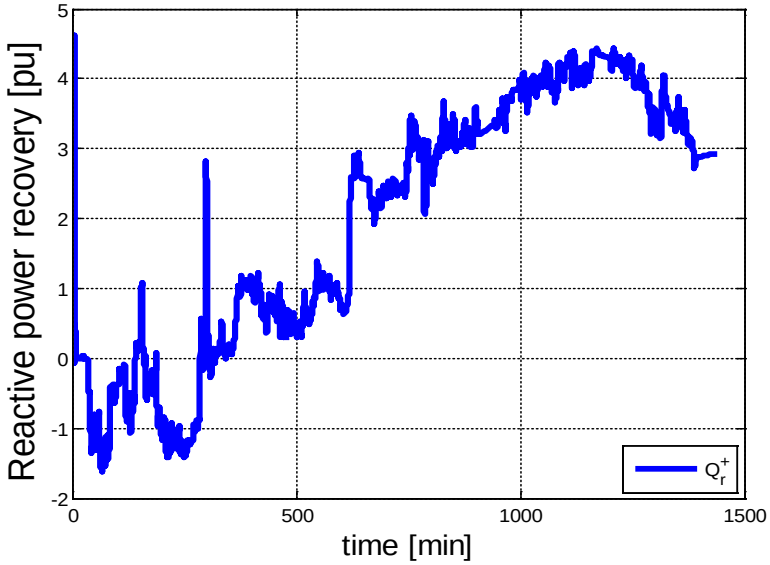

Figure 1.18 Plot of estimated values for Q_r.

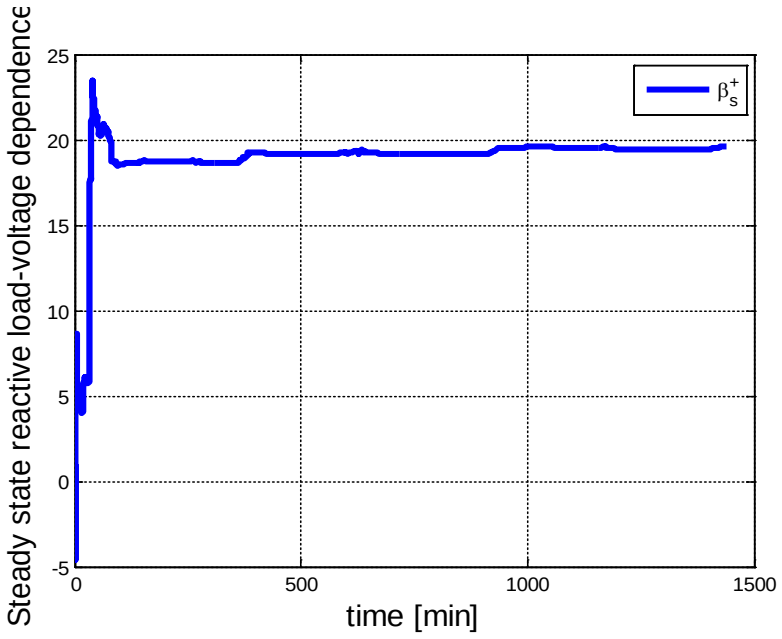

Figure 1.19 Plot of estimated values for β_s.

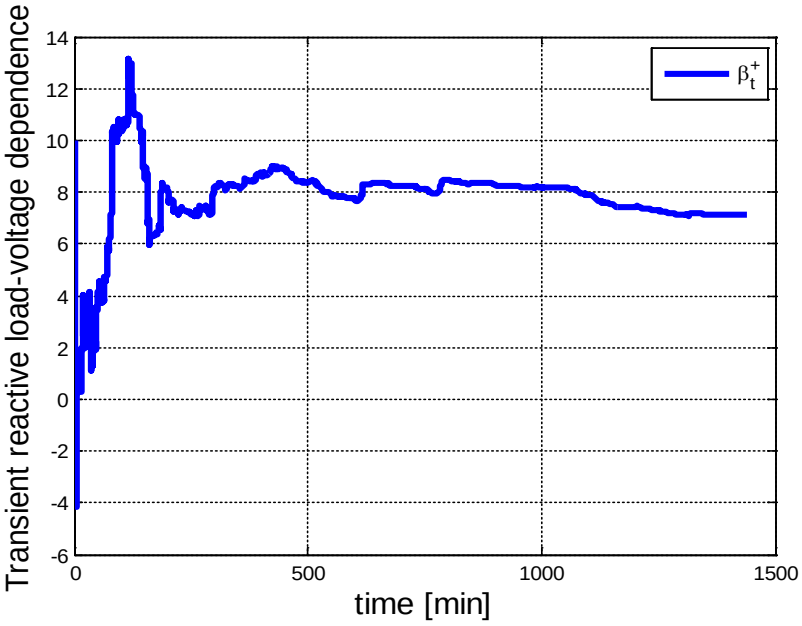

Figure 1.20 Plot of estimated values for β_s.

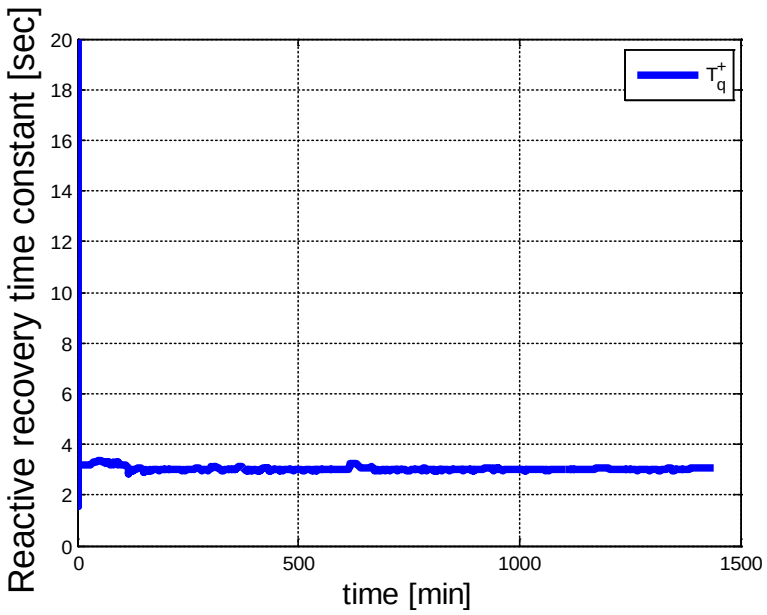

Figure 1.21 Plot of estimated values for T_q.

As in the case of real power model tracking, performance of the UKF is evaluated for tracking the model of the total reactive power. This is accomplished by plotting the total reactive power that is measured and predicted as shown in Figure 1.22.

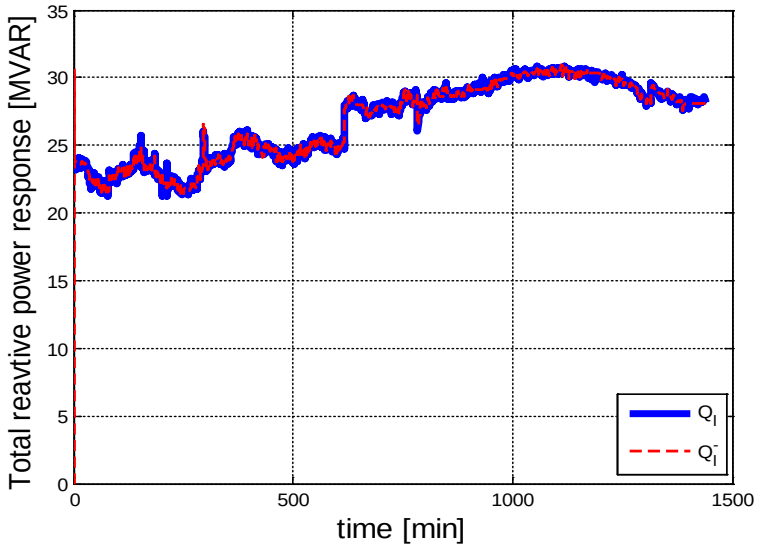

Figure 1.22 Plot of actual and estimated values for Q_l.

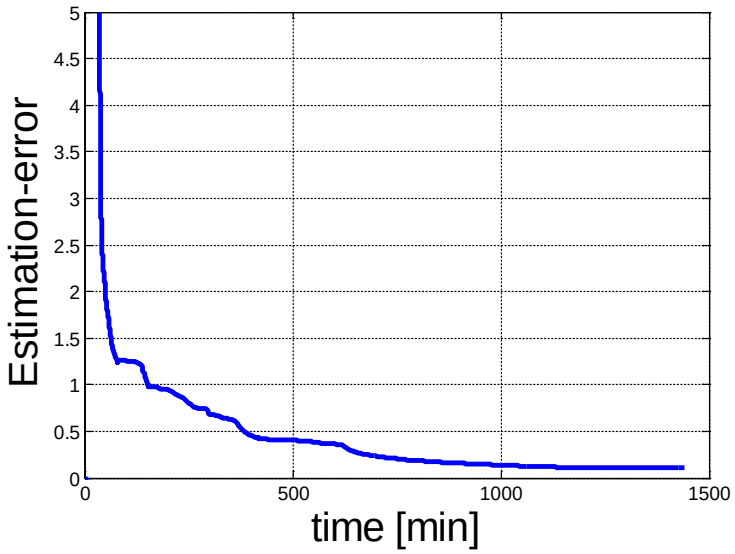

Figure 1.23 Plot of MSE when tracking total reactive power.

Figure 1.23 shows the plot of MSE for the total reactive power and as evident from the figure MSE gradually reduces to an acceptable level.

It should be mentioned that the computational burden per time step for the proposed dynamic estimator is rather modest due to the small number of state and parameters being tracked. So the approach lends itself readily to real-time implementation.

1.4 Conclusions and Future Work

This report is concerned about identification of dynamic load models for power grids. It is recognized that developing and maintaining models that are based on first principles is prohibitively complicated and computationally demanding. Instead, the paper aims to develop a measurement based model whose parameters can be tracked in real-time. A well-studied and accepted exponential dynamic load model is used as a basis in order to develop an UKF to identify and estimate model parameters in real-time. The proposed dynamic estimator is first tested using artificially generated measurement data. Subsequently, the estimator performance is evaluated using actual recorded utility load data captured in 6 second intervals. Both simulation and experimental results indicate that such a dynamic estimator can provide an accurate dynamic load model based on real-time measurements. Given the recent increase in deployment of phasor measurement units in

substations, such measurements are expected to be widely available for implementing the proposed application of this work in the near future.

This study considered development and maintenance of a dynamic load model based on real-time measurements. While the captured model may be considered valid for short term dynamic studies, there is no guarantee that the model will remain valid as the operating conditions change. So, the authors are interested in extending this work to develop dynamic models which will remain valid for extended periods of time.

Part 2. Exploiting Smart Meter Data for Enhanced Load Modeling

2.1 Introduction

2.1.1 Background

As one of the essential elements toward the future smart grid, smart meters have been widely installed in the developed world. It is the first time in history that utilities and system planners gain access to measurements for customers at the building level with great time resolution and at such a large scale. Compared with conventional meters, smart meters generate data that fill the void in the minute-to-hour time scales and couple the spatial and temporal scale.

The massive historical database created by smart meters contains a wealth of information which has not been fully explored or exploited. Currently, most studies on smart meter data are limited to load forecasting [19, 20] and typical load profile (TLP) identifying [21]. However, another critical need for enhanced distribution system operations and planning is a better load model. This project explores a new possibility of building an enhanced load model by implementing data mining techniques on smart meter historical database [22].

2.1.2 Overview of the Problem

From a mathematic point of view, a load model is a formula of the relationship between bus voltage and power (real and reactive) [23]. Compared with the modeling of generators and transmission systems that have been studied in detail, an accurate time-variant load model is difficult to achieve due to electricity load uncertainties and data insufficiency. Traditionally, there are two popular approaches to build a load model: measurement-based approach [24, 25] and component-based approach [26, 27].

The measurement-based approach determines the load model by recording the load responses directly through system voltage stage tests and actual system transients. Although accurate, the measurement-based approach is costly: testers need to perform specific experiments on real system by deliberately changing transformer tap positions,

which may affect energy quality to customers. Moreover, the measurement-based load modeling method cannot capture the time-variant properties of the load. In other words, the load model built through the measurements only reflects the load's property at the time when those measurements are taken. For example, errors are introduced when a daytime load model is used for midnight load analysis.

The component-based approach estimates the system load's P-V and Q-V properties by aggregating typical load components according to certain ratios, which are also the load's ratios of the typical load components in the system. Instead of taking system measurements, this approach builds detailed load model in advance for common load components in the system being studied, such as air conditioners in the residential loads or the electric machines in the industrial loads. Component-based approach hence avoids costly system tests by taking surveys to determine the ratios of typical load components and building load profiles for each load component. However, the accuracy of this approach strongly depends on the accuracy of the load components ratios and the specific models built to represent typical load components. As a result, in most cases, the load model built through component-based approach needs verifications using real system measurements. Table 2.1 lists the strengths and weaknesses of both the measurement-based method and component-based method.

Table 2.1 Comparisons of measurement-based and component-based method

	Measurement-Based Method	Component-Based Method
	Manually change the voltage at the substation and take measurements. Load model is later built through parameter identification.	*Model the real load by aggregating pre-modeled components through some ratios, which are determined through large scale surveys.*
Strengths	*Accurate, no need for verification*	*No costly system tests are needed*
Weaknesses	• *Tests are expensive* • *Tests leads to bad power quality* • *Not available on building level* • *Difficult to get 24 hour model*	• *Need pre-modeled models for different load components* • *Large surveys to set those ratios* • *Survey results ≠ load in reality* • *Model needs further validation* • *Difficult to get 24 hour model* • *Does not account customer behaviors*

The proposed enhanced load model identifies the load's P-V and Q-V properties through data-mining techniques and is time-variant for enhanced accuracy, see Figure 2.1.

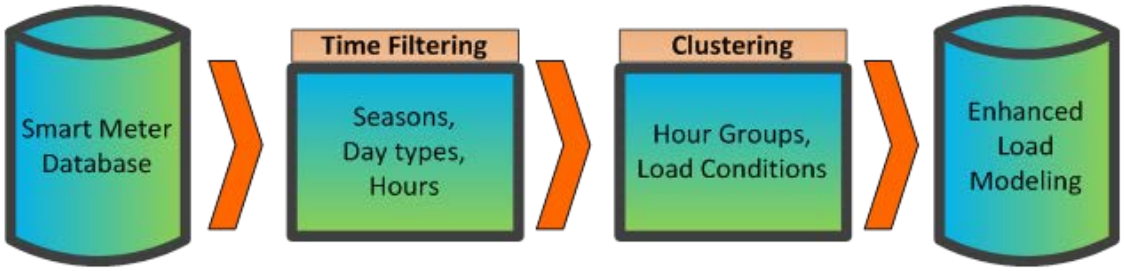

Figure 2.1 Enhanced load modeling method

The new load modeling method has the following merits:

- No expensive system tests or large surveys are needed.
- The load model is available to building level as long as a smart meter is presented.
- It is a time-variant load model taking the customer behaviors into account.
- No further validations are needed since the model comes from real system measurements.

Generally, there are two major barriers for a data-mining-based approach.

First, the load reading resolution for current smart meters ranges from 15 minutes to an hour. Data collected on such a resolution level cannot distinguish the effects from instantaneous load changes and system voltage deviations, both of which are responsible for real and reactive power consumption changes. Based on a previous research by EPRI, this data resolution is not good enough to determine the load composition in detail to allow immediate implementation of the component-based load model [25]. As a result, instead of seeking the decomposition of the load into specific load components, this project introduces the concept of *load condition* and focuses on modeling the P-V and Q-V properties of the load without load disaggregation.

Second, the historical smart meter readings from the massive database need to be clustered to prepare data for meaningful and high-quality load model parameter identification. As a result, multiple data mining techniques such as Davies-Bouldin Index (DBI) and K-subspace method are implemented to facilitate the modeling process.

2.1.3 Report Organization

In the first section, we introduce the smart meter deployment on Georgia Tech main campus as the data testbed for the study. We develop an interactive visualization tool, "Smart Grid Plotter", for easier visualization of the cumulative smart meter database. The visualization tool allows researchers navigate historical data collected by smart meters for all buildings on campus. Users can further configure and save the desired plot through various parameters menus.

In the second section, we propose a novel enhanced load modeling method based on data-mining and machine learning algorithms. The enhanced load model is a time-variant model that allows expressing the load's active and reactive power usage as a function of both time and voltage. The detailed steps for the new load modeling method are further discussed in details through three aspects: data aggregation, hour partition and the load condition assumption.

In the third section, we further explore the smart meter data for both off-line and real-time utility functions. In the report, we show that as a very important information source, smart meter data (both real time data and historical data) can be the core of other 18 potential applications when combined with other data, such as weather data and GIS information. Two exemplary applications, refined power flow analysis and dynamic distribution network reconfiguration, are studied to show how smart meter data and the proposed enhanced load model improve power system analysis results and facilitate advanced energy efficiency operations.

2.2 Data Background and the Visualization Tool

2.2.1 Data Background and Analysis

The tested data used in this project come from the smart meters installed in the distribution system of the Georgia Tech main campus. Nested in the heart of Midtown Atlanta, Georgia Tech is a highly developed and complex university campus with key facilities and services. Georgia Tech owns and maintains its own electrical distribution system in order to:

1) Increase reliability and flexibility, aiming at facilitating maintenance, making the system less susceptible to faults, and reducing faults affecting user numbers.
2) Enjoy benefits such as real time pricing, which are only available to industrial customers.

The Georgia Tech distribution system consists of 15 distribution circuits, all fed from the same substation, which serves more than 200 buildings.

Many of the buildings on campus have extensive instrumentation for control and monitoring of electrical and mechanical signals, including approximately 400 revenue grade smart meters. Every 15 minutes, the measurements reported by the smart meters are recorded and aggregated into a database that allows for comprehensive analyses of events. The general scheme is that a building has a main meter and may include sub-meters, typically for billable tenant load but sometimes applied for specific areas of interest such as a chiller or a PV system, or for measurements and verification (M&V) in a LEED-certified building.

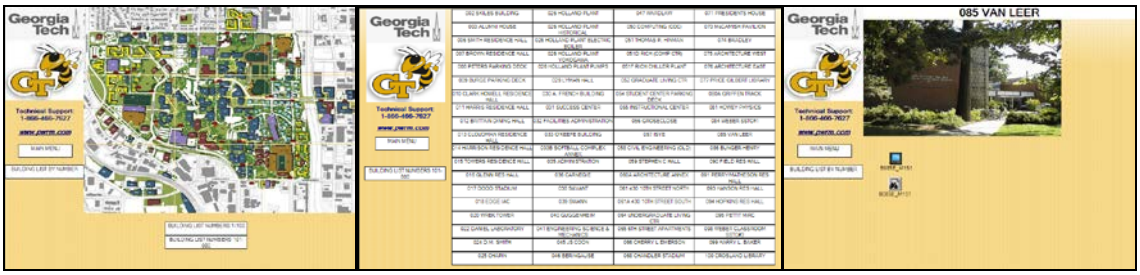

Figure 2.2 ION webreach main menu (left), list of buildings 101-800 (middle), and an example of a building menu

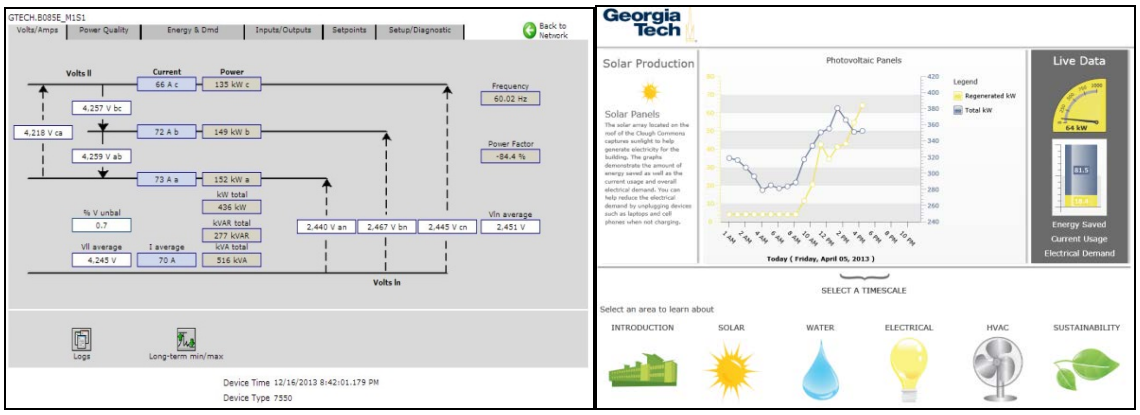

Figure 2.3 ION webreach meter interface

2.2.2 Interactive Visualization Tool

To understant and explore the smart meter data, we developed an interactive visualization tool by the name of "Smart Grid Plotter". To begin with, the smart meter data need to be fetched from the facility's database. To deal with the information more efficiently, all smart meter data have been organized in a SQLite database which is optimized for interactive visualization in Java, as shown in Figure 2.4.

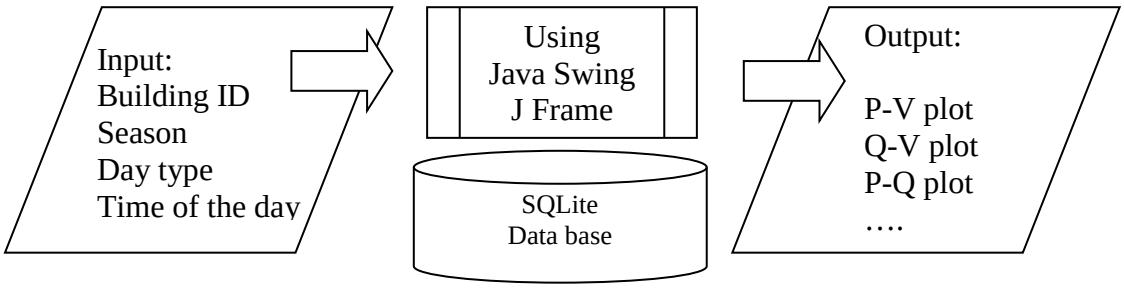

Figure 2.4 Structure of "smart grid potter"

Figure 2.5 shows the interface of the interactive visualization tool. The upper area is the plotting area, where various plots are generated according to the user's specific inputs. The lower area is the parameter input area, where user can insert specific requirements to obtain the desired plot in the plotting area.

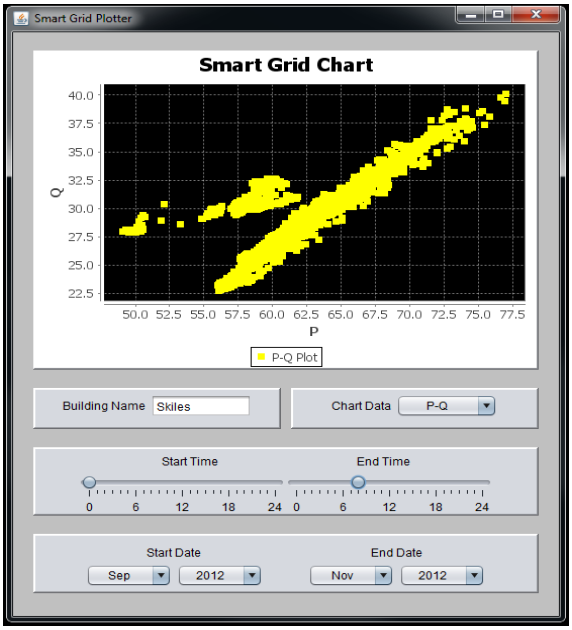

Figure 2.5 Smart grid plotter

As shown in Figure 2.5, we can choose the studied building by inserting building name and pick the plot type by choosing the “Chat Data” manu, which includes P-Q, P-V and Q-V plots. The “Start Time” and “End Time” bar allow users to navigate the data between the desired time span of a day. User can set the studied period by editing the “Start Date” and “End Date” menus at the bottom.

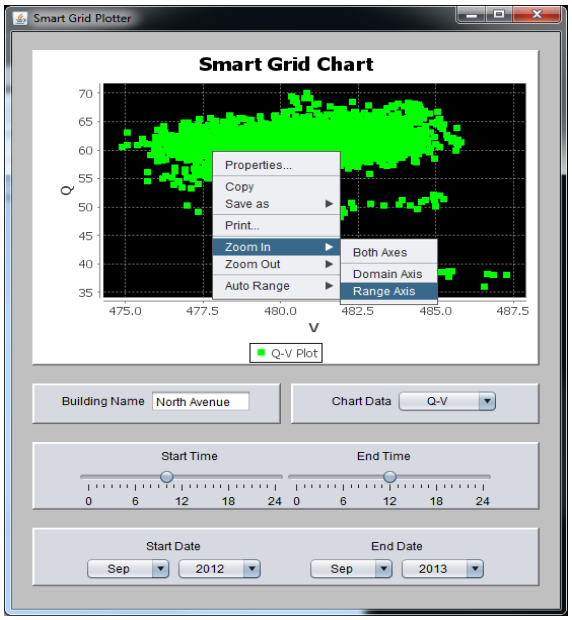

Figure 2.6 Zoom in/out

The interactive visualization tool also allow users to zoom in and out to see the details of the plot by left clicking in the plotting area, as shown in Figure 2.6. Users can also save the plot or print the results as well.

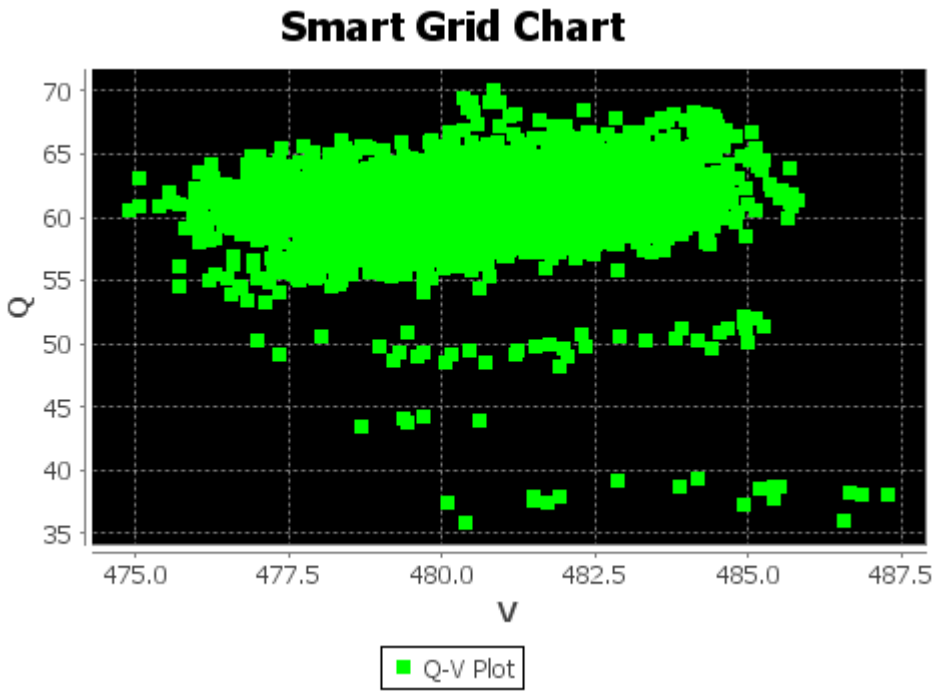

Figure 2.7 Q-V plot of a student residence hall

The visualization tool serves to provide the research with intuitive information on the load behind each smart meter, which is very important in building different load profiles and studying customer behaviors through time. For example, Figure 2.7 shows the annual Q-V plot of a student residence hall from 9am to 1pm (Sep. 2012 to Sep. 2013). From Figure 2.7, we can observe a linear relationship between reactive power usage and voltage. In fact, the reactive power usage is tent to increase as the system voltage increases.

2.3 Enhanced Load Modeling

In this section, we explain the proposed enhanced load modeling method in detail. To counter the barriers mentioned in Section 2.1.2, load condition assumption is introduced. Then various machine learning algorithms will be implemented to filter the original historical smart meter data. Finally, we build the enhanced load model, which is also a

time-variant model through regressions. Figure 2.8 shows the flow chart of the entire load modeling method and its associated algorithms.

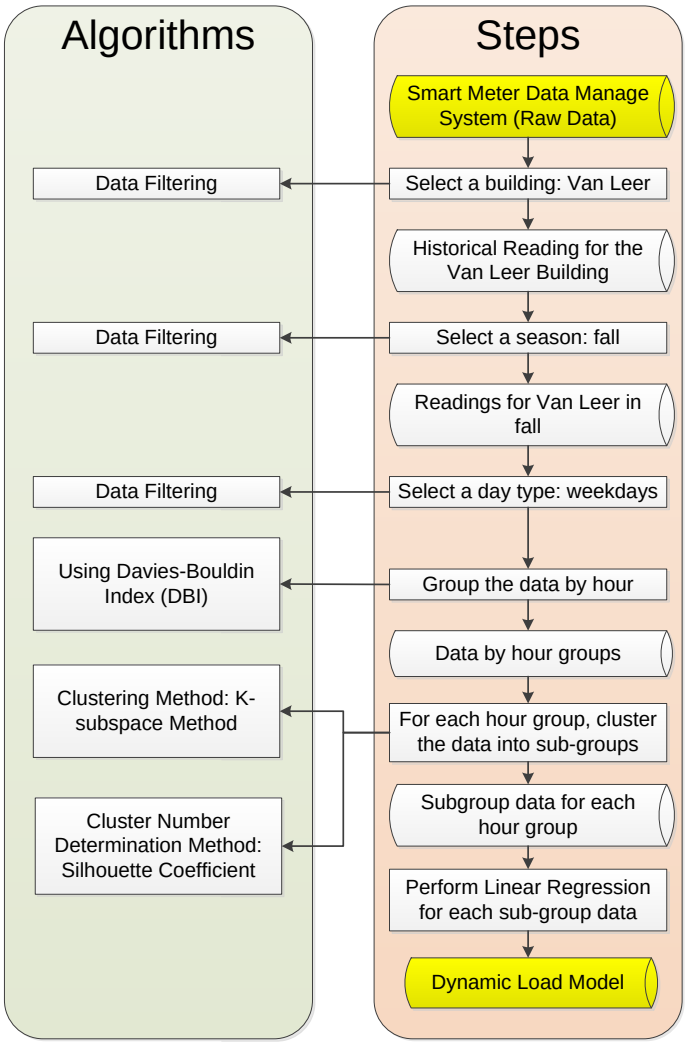

Figure 2.8 The enhanced load modeling method

2.3.1 Data Aggregation

In this section, we aggregate the smart meter data according to the data sources and the time when the readings are collected.

To begin with, the wide spread installation of smart meters allows researchers to gain access to the loading information on every customer and in great detail. Hence, for the first time in history, researchers could look into every load's behaviors based on the

specific building name. In other words, the first step of the data aggregation is to group the data according to the specific source where the data are collected.

Since customer load changes according to time of the day, day of the week and season of the year, it is crucial to establish a time-variant load model to describe the load under various time frames. In this study, time itself is considered as a very important index to aggregate smart meter data. We refer the time-based data aggregation process as time filtering or data filtering. In other words, we prepare the data for time-variant load modeling by grouping the historical smart meter data according their time labels.

This aggregation procedure has multiple benefits:

a) It yields organized data, and consequently models, that are easier to interpret.

b) It increases the volume of data points per static model, allowing for more robust modeling.

Table 2.2 shows how to aggregate smart meter data, and how to label the results of the time filtering process by giving each smart meter reading a data label. The labels in Table 2.2 contain information about the load type, time and load condition. The proposed model has a tree structure that branches through three layers: load type layer, time layer and load condition layer. All the smart meter readings in the database are aggregated and labeled as in Figure 2.9.

Table 2.2 Time-variant load model and data structure

Model Struc.	**First Layer**	**Second Layer**			**Third Layer**
	Load Type	***Season***	***Day***	***Hour***	***Load Cond.***
Data Label	Commercial, Residential, Industrial	Spring, Summer, Fall, Winter	Weekday, Weekend, Holiday	Hr. Group 1, Hr. Group 2, ... Hr. Group K	Condition 1, Condition 2, … Condition K

On the first layer, all loads are classified into commercial, residential and industrial loads. Ideally, a data-mining-based load modeling method does not require users to specify the load type as long as the load is equipped with smart meters. However, marking the data with load types can help us better understand the time-variant properties among different types of load.

On the second layer, for each individual load, all the smart meter readings are marked with time labels. Different time labels are good indicators for customer routine behaviors. For example, season labels can distinguish the cooling and heating loads; day labels can distinguish the loads among different day types, such as weekday and weekends; hour labels can distinguish the working hour load and off-working hour load.

On the third layer, smart meter readings with the same time label will be further clustered and marked with different load conditions. The next section will explain the load condition assumption in detail. On this layer, the model parameters are identified using smart meter data of the same load condition label.

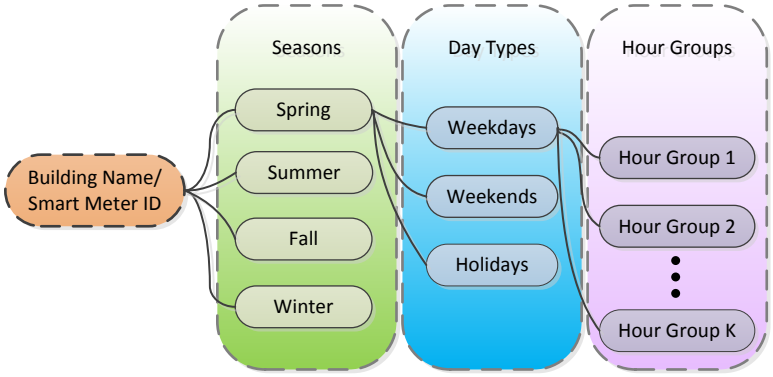

Figure 2.9 Tree-structured data labels

2.3.2 Hour Partition

From the data aggregation process, we know the load properties vary through time. However, it is important to decide to what extent we want to filter the data. In an extreme case, the load model definition could go from 15 minutes per model to 24 hours per model. On one hand, if we try to establish a time-variant model every 15 minutes, we will not have enough data to create a robust model for each model, and there would be 94 models for just one day, which can result in a complicated and inaccurate model. On the other hand, if we try to establish only one load model per 24 hours, we are treating the working hours and off working hours with the same model, which loses the key advantage for a time-variant load model. Hence, we decided to group the 24 hours into a proper number of hour groups according to some machine learning algorithms. The load models are assigned to each of those hour groups. For example, mid night hours from

1am to 6 am might be grouped a single hour group and be represented using one load model, because the activities during the midnight for commercial and residential loads are pretty stable. As is shown in Figure 2.10, the variable of time adds to the complexity of the load modeling algorithm.

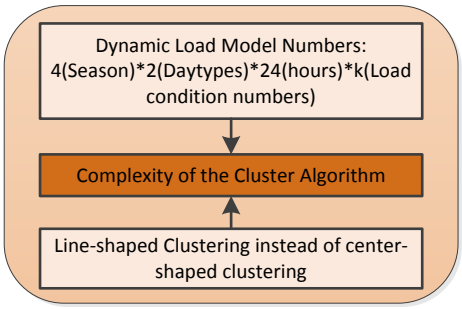

Figure 2.10 Time index adds to the complexity of the model

Since our initial exploratory data analysis (namely through Q-V and P-V plots) revealed that these data points exhibit significantly different patterns across different hours of the day within a given configuration, we would like to build separate models for each group of hours with similar data points. Hence, given 24 sets of P-V (or Q-V) data points, each corresponding to one hour of data for a given configuration, we seek to group different hours exhibiting similar P-V (or Q-V) patterns. The hour partitioning process serves to reduce the complexity of the modeling algorithm, shown in Figure 2.11.

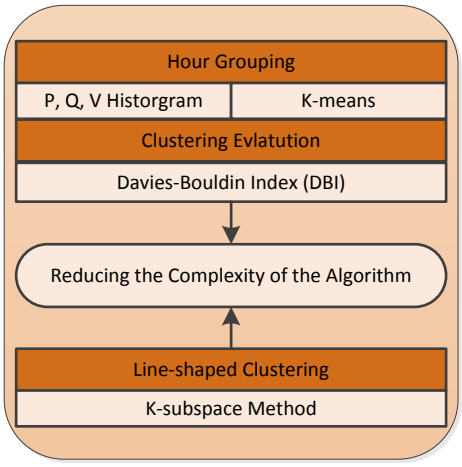

Figure 2.11 Hour partitioning serves to reduce the model complexity

The difficulty here lies in that we are not looking to cluster N data points in some d-dimensional space, as in typical clustering problems; we rather seek to cluster 24 sets of data points into K clusters, which we will refer to as “hour groups” (one cluster could be sleep hours 12am-8am, another could be working hours 9am-5pm, etc.). To accomplish this task, we first bin the P-V data points in B bins along each of the two dimensions (P and V), as is typically done for histograms. As a result of these operations, we obtain 2B frequencies, corresponding to the fraction of data points falling within a certain bin. Applying this procedure to each of the 24 hours, we construct a feature vector of 2B features per hour, where each feature is a bin frequency, see Figure 2.12. Now that we have mapped the hours into this space, we are able to simply apply classical clustering methods, such as k-means, in order to obtain k clusters, i.e. our hour groups.

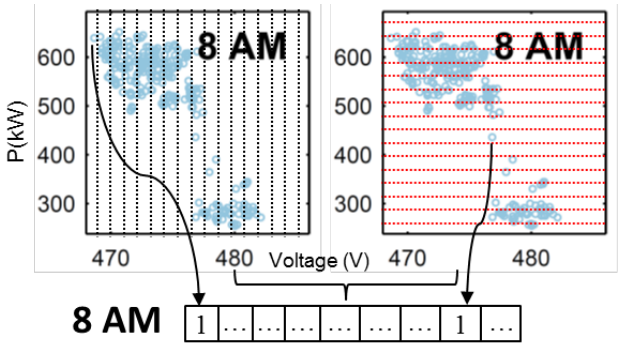

Figure 2.12 Hour grouping method (clustering vector formulation)

The last hurdle lies in choosing the best possible number of clusters k. We adopt the Davies-Bouldin Index (DBI) measure, typically used for cluster quality evaluation in the machine learning literature. In fact, DBI (introduced by David L. Davies and Donald W. Bouldin in 1979) is a metric for evaluating clustering algorithms [28] and is an internal evaluation scheme, where the validation of how well the clustering has been done is made using quantities and features inherent to the dataset.

Assume *k* is the number of the clusters which is to be determined, let *k* ranging from 2 to 6, we run k-means for each of these values, and score the resulting clustering using DBI. Finally, we choose the *k* value corresponding to the smallest DBI score, and save the resulting *k* hour groups. We will aggregate the data points of a given configuration that

belong to the same hour group, and build a model for each such set of data points, see Figure 2.13.

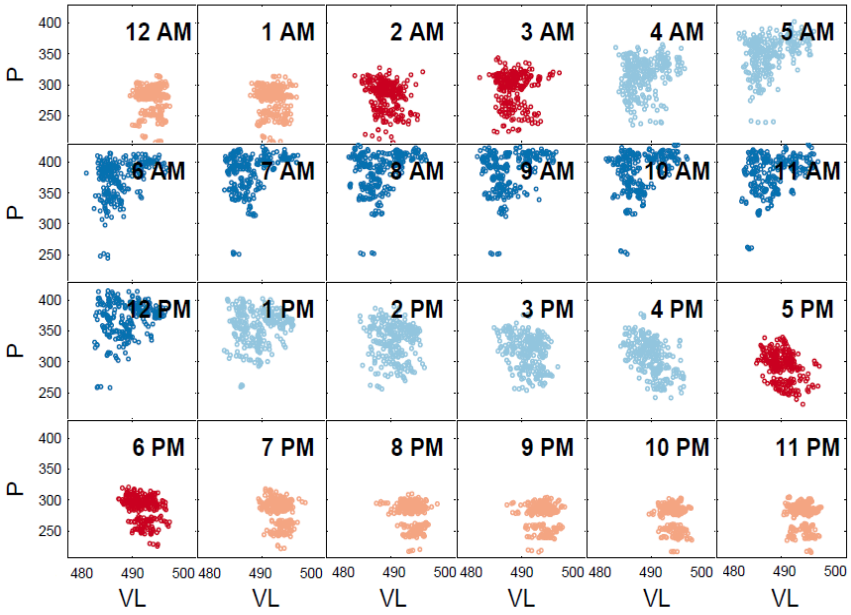

Figure 2.13 Hour grouping results for an apartment building (winter,weekdays)

Various load types are studied to explore their differences in identifying the time label. In the study, a student residential hall and a family apartment are chosen as residential loads; an office building and a student center are chosen as commercial loads; and a chiller plant on campus is chosen as an industrial load. Their time label identification results for weekdays in fall are shown in Table 2.3, where hours with consistent load behaviors are merged. Moreover, results shown in Table 2.3 also indicate that even under the same load type, different customers have their own power consumption patterns, such as the peak hours between the student residential hall and the family apartment. These customized properties can only be captured by performing smart meter data mining.

Table 2.3 Time label identification results (weekdays, fall)

Commercial Loads																								
Office Building	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Student Center	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Residential Loads																								
Residential Hall	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Family Apt.	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Industrial Loads																								
Chiller Plant	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23

Note: ■ stands for working hours (peak hours); ■ for off-working hours (night hours); ■ for daytime hours specifically found in residential loads; ■ for hours in between.

2.3.3 Load Condition Assumption

2.3.3.1 Load Condition Assumption

For the purpose of this project, a load composition is defined as the state of the total aggregated load, including total real/reactive power and the precise connected individual loads that represent this aggregate load value. Technically, each load composition can be modeled by a set of static model parameters. However, due to the number of individual load components in a large building, the number of possible load compositions of which devices are connected is significantly large. In practice, a fixed and rigid static model for a building is not accurate enough to model the dynamically changing nature of the load, because the load composition changes over seasons of the year, days of the week, and hours of the day.

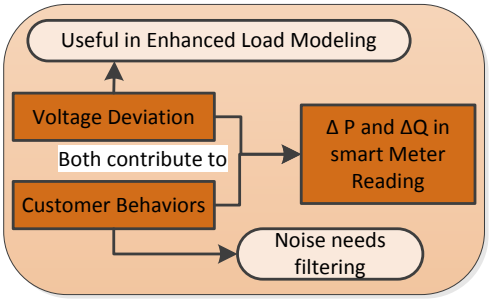

Figure 2.14 Voltage deviations vs. customer behaviors

In traditional measurement-based load modeling, data is collected at a very high frequency (1000Hz) before and after the voltage deviation [31]. As a result, the load composition is assumed to be fixed, and only voltage is responsible for the load's real and reactive power changes. However, for most smart meter databases, the data is logged at a

resolution of several minutes or hourly, and load composition is subject to changes between different readings. In other words, voltage is no longer the only factor that influences the power consumption of the load, as shown in Figure 2.14. Hence, an assumption about the load condition is made to justify that it is possible to filter out the instantaneous load changes, and build the P-V and Q-V model through data mining techniques, shown in Figure 2.15.

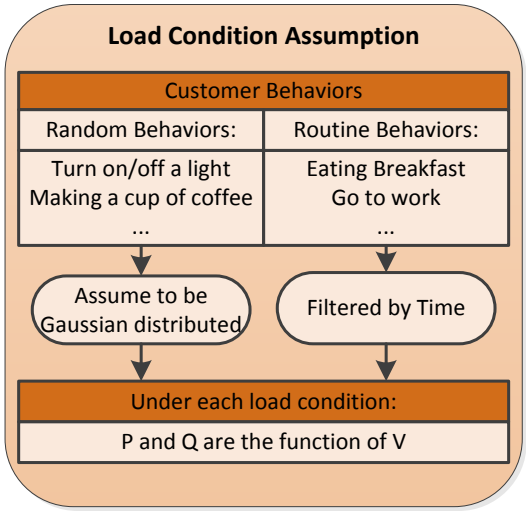

Figure 2.15 Load condition assumption

To begin with, by definition, the number of load compositions is 2^n, where n is the number of appliances. Every smart meter reading for the load is measured under one of those load compositions. In this paper, load condition is defined as a group of load compositions sharing the similar P-V and Q-V properties. As a result, the smart meter readings can be clustered accordingly into several load conditions.

The energy consumption of customers can be separated into random behaviors (such as turning on a light or making a cup of tea) and routine behaviors (such as eating breakfast in the morning or turning on the heater in winter). It is assumed that routine-behavior loads are usually the dominant factor in energy consumption and are strongly correlated to time, such as seasons and working hours. In contrast, random-behavior loads can be interpreted as additional small loads on top of the energy consumption of routine behaviors. Compared with routine behaviors, random behaviors change more frequently and are responsible for the frequent instantaneous load changes. Figure 2.16 shows the

night time Q-V plot of a commercial building during the summer season, from which we can observe two distinct load conditions.

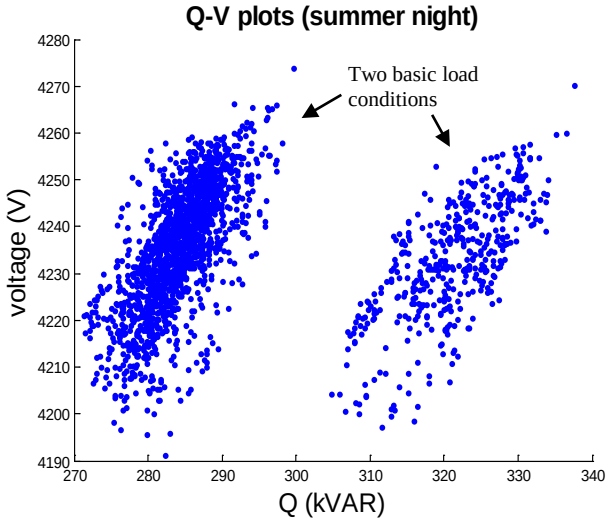

Figure 2.16 P-V and Q-V plot for a commercial building in summer night

Under this assumption, all load compositions within a load condition are considered to be different random behaviors on top of the same routine behavior. As the result, data mining techniques can be implemented to identify different load conditions by clustering all smart meter readings. When all data is clustered, a static model is built for every load condition using.

2.3.3.2 K-Subspace Method

Now, given a set of data points belonging to an hour group of a certain configuration, our goal is to cluster these data points into cohesive groups, and find the parameters (slope and intercept) of the line that best fits those points; these lines will be our Q-V or P-V load models.

Given the assumption that Q and V follow an almost linear relationship, our exploratory data analysis had revealed that the data points tend to be organized in a few separate line-shaped clusters, as shown in Figure 2.16. As a result, traditional clustering algorithms such as k-means fail to capture the correct clusters, shown in Figure 2.17.

To overcome this difficulty, we resort to the K-Subspace clustering algorithm, which is able to discover *k* line-shaped clusters, via minimization of the orthogonal distance from each point to the nearest such line.

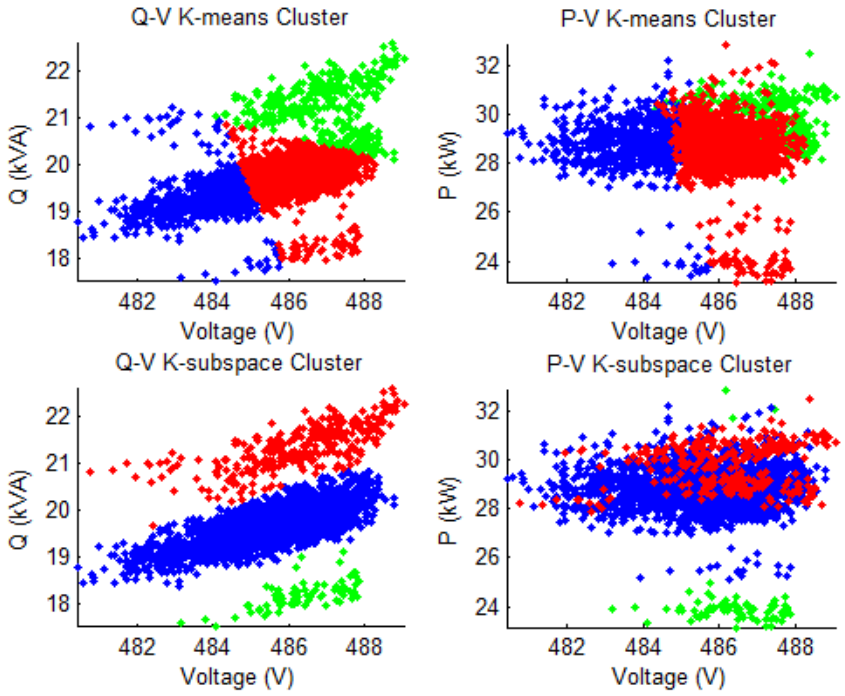

Figure 2.17 Comparisons between K-subspace method and K-means method

In practice, multiple load conditions can exist under the hour group. As a result, on the third layer of the model, smart meter readings are clustered into several load conditions so that each of the load conditions can be modeled by a static model, shown in Figure 2.18.

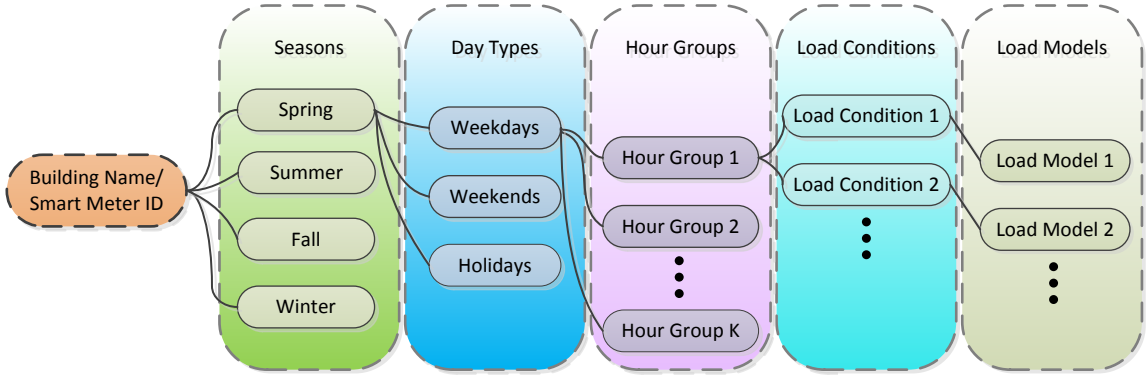

Figure 2.18 The load condition clustering based on hour partition results

Traditional K-means algorithm [29] clusters data based on their relative Euclidean distance to the nearest cluster center with an iterative process to adjust the centroid. The clusters' shapes are determined by the perpendicular lines between centroids. However, the smart meter readings of different load conditions are distributed in a very specific line-shaped pattern close to each other.

K-subspace method [30] allows the detection and clustering of line-shaped data by assigning each cluster C_k with a unit direction vector $\boldsymbol{a}_k$ and a center $\mathbf{c}_k$. The entire algorithm seeks to minimize the perpendicular distance of all the data points $\boldsymbol{x}_{k,i}$ to the line defined by $\boldsymbol{a}_k$ and $\boldsymbol{c}_k$ within each cluster, as shown in Equation 3.1.

$$\min_{\mathbf{c_k},\mathbf{a_k}} \sum_{\mathbf{i}\in\mathbf{C_k}} \mathrm{Dist}(\mathbf{x_i},\mathbf{C_k}) = \sum_{\mathbf{i}\in\mathbf{C_k}} \|\mathbf{x_i} - \mathbf{c_k} - \alpha \mathbf{a_k}\|, \tag{3.1}$$

where $\alpha = (\boldsymbol{x} - \boldsymbol{c}_k)^T \boldsymbol{a}_k$.

Figure 2.17 shows the Q-V and P-V plot of a commercial building during off-working hours on weekdays in the fall 2012. Comparing Q-V plot with P-V plot, it can be seen that reactive power are more sensitive to voltage deviations than active power. As a result, the load conditions are clustered using Q-V plot. In Figure 2.17 the clustering results are marked with different colors, where the cluster number K is set to be 3.

2.3.3.3 Silhouette Coefficient

Similarly to the case of hour partitioning, we propose to automatically detect the number of clusters k. We achieve that goal by adapting the Silhouette Coefficient, another common cluster evaluation method for machine learning, to our line-shaped clusters. Equation 3.2 defines the Silhouette Coefficient.

$$\mathrm{S(i)} = \frac{\mathrm{b(i)} - \mathrm{a(i)}}{\max\{\mathrm{a(i)}, \mathrm{b(i)}\}}, \tag{3.2}$$

where $\mathrm{S(i)}$ stands for the Silhouette coefficient; $a(i)$ and $b(i)$ stand for the dissimilarity of all data within the same cluster a and b.

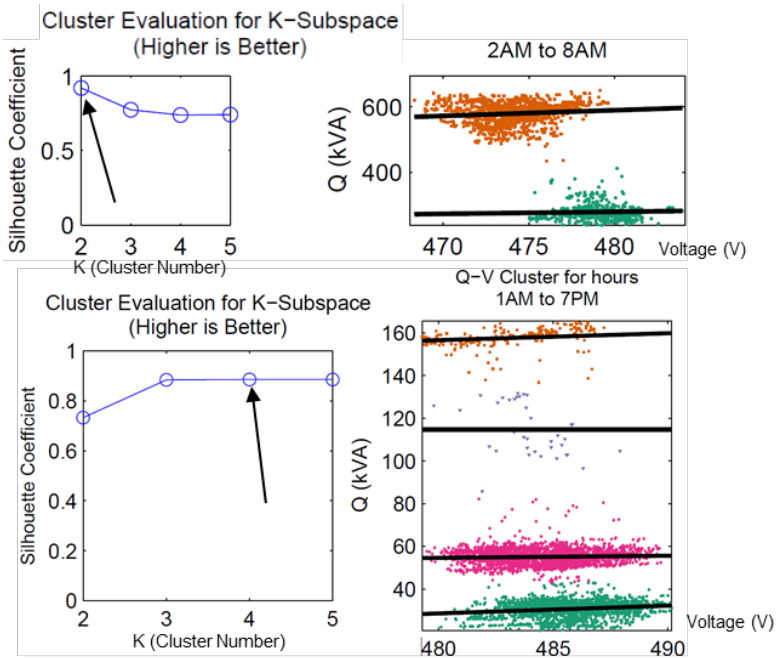

Figure 2.19 Silhouette coefficients for K-subspace method

Figure 2.19 shows how to determine the cluster number by comparing Silhouette coefficient. The upper figure comes from an office building and the lower figure comes from a chiller plant. In both cases, the cluster number with the highest Silhouette coefficient gives the best cluster number estimation. The right side figures are the K-subspace clustering results based on the optimal cluster number determined by Silhouette coefficients.

2.4 Exploiting Smart Meter Data for Utility Functions

An inventory of 18 major potential operational and economic applications of Smart-Meters at the distribution level has been developed. Figure 2.20 shows a brief view of the relationship among those applications.

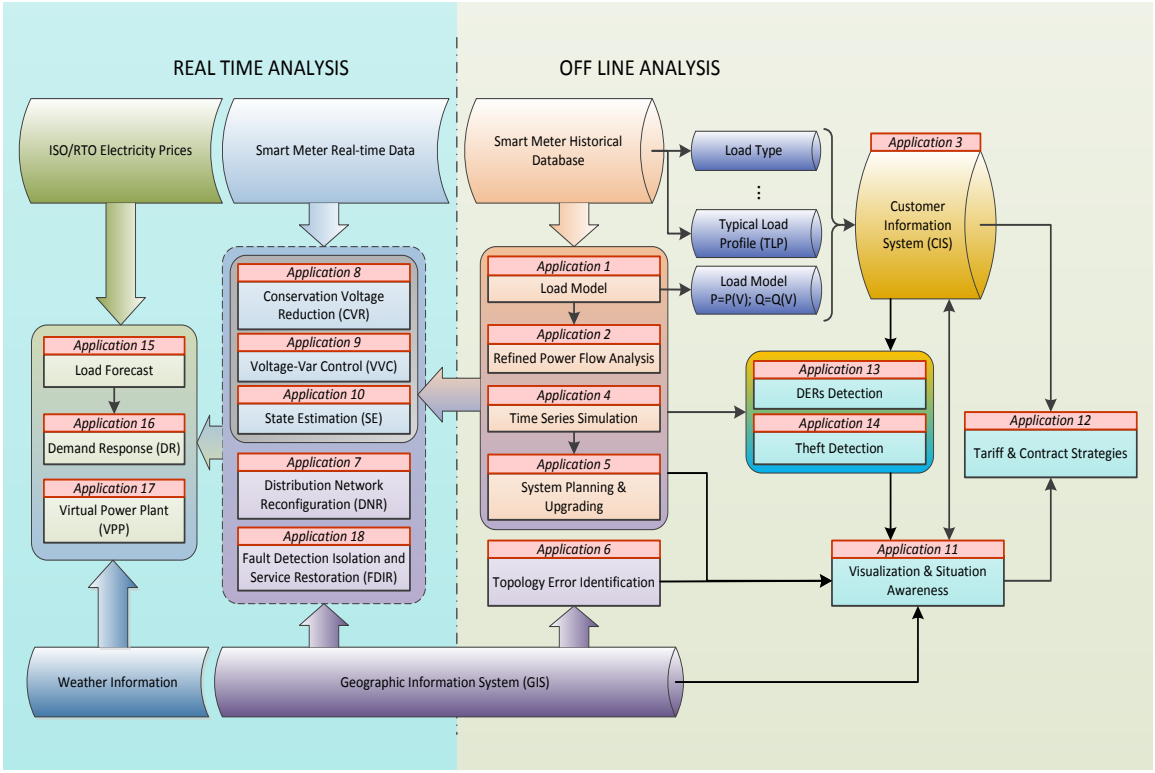

Figure 2.20 Model for advanced smart meter applications

2.4.1 Smart Meter Data Application for Online and Offline Operations

From Figure 2, we can see that the installation of smart meter provides utilities with both real-time data and historical data, both of which are of great importance in both real-time and off-line distribution network analysis.

To begin with, the historical data collected by smart meters is a valuable resource for the enhanced load model. The enhanced load model is superior to the traditional load model in many ways: First, the enhanced load model is built on the building level instead of the substation level, as long as the building has a smart meter installed. Second, the enhanced load model is a dynamic model which varies according to seasons, day types or even hours. Third, the enhanced load model is self-validated because the model is built upon real measurements collected by smart meter, which avoids costly experiments for model verification.

Along with load type data and typical load profile (TLP) information, the enhanced load model can be used to form a new customer information system (CIS), where every customer has its own profile. The new CIS allows utilities to watch out for potential harmful behaviors and to negotiate for better contracts with customers. Moreover, the CIS contains the customer's power energy consumption pattern which facilitates other system functions such as load forecast and demand response.

Since, the enhanced load model writes the real and reactive power usage of a customer as a function of time and system voltage, it enables a refined dynamic power flow analysis. Compared with the traditional fixed-P and fixed-Q model, the enhanced load model provides more accurate power flow results for system operators to understand the system responses to voltage perturbations. These voltage perturbations may be introduced by system faults or some control actions such as LTC taps or shunt capacitors.

The historical smart meter data also make the time-series simulation possible for distribution systems, which is a powerful tool when it comes to power system planning and simulation. Traditionally, there is no way for system operators to perform true time-series simulation since there is no time series data collected by smart meters. Instead of time-series simulation, system planners try to adopt several synthetic scenarios to represent different operation states of the power system. However, these synthetic scenarios will lead to inaccurate conclusions especially in economic related decision makings, which are not based on several synthetic scenarios but true time-series data from the real system. It is known that most of the advanced distribution control functions require various upgrades based on the original distribution system configuration, such as circuit breakers, shunt capacitors, dispatch-able distributed generators and so on. These advanced functions include conservation voltage reduction (VCR), voltage-var control and optimization (VVC&O), and distribution network reconfiguration (DNR). Time-series simulation data provided by smart meter historical database serve well to justify the economic benefits of those system upgrades.

Statistics and data-mining techniques are powerful tools to analyze the smart meter historical data. In many cases, the distribution system is subject to topological errors as

the voltage level gets lower, and many components are packed in a very small area. By studying the time-series voltage correlations, the historical database can provide convincing information to identify system topology errors. Moreover, proper data mining techniques also enable utilities to better manage their customers. Since misbehaviors such as energy theft and unnoticed PV installation result in huge profit losses for utilities throughout the country, it is crucial for utilities to take advantages for the CIS and historical smart meter data to detect those misbehaviors in time and take necessary actions accordingly.

The real-time feedbacks or near real-time data from smart meters can be used as inputs for real time functions such as CVR (Conservative Voltage Reduction), SE (State Estimation), DNR (Distribution Network Reconfiguration) and VVC&O (voltage-var control and optimization). The real-time feedbacks from smart meters are also of great value when it comes to Fault Detection Isolation and Service Restoration (FDIR). This is especially true in the legacy system where system failures can only reported by customer feedbacks and excessive man forces are standby to locate the fault and restore services. The introduction of the smart meter and corresponding advanced metering infrastructure (AMI) will take advantages of various sensors' feedbacks such as the last gasp reading from smart meter and zero voltage distribution to fast identify system failures and accurately locate the fault to save man force.

By analyzing the real time data from smart meter, utilities will be able to play a more active role in the energy market by adopting a highly reliable load forecasting results and managing the controllable loads and distributed energy resources (DERs) through virtual power plant (VPP) technologies.

Finally, equipped with proper visualization tools, smart meter data can benefit both customers and utilities as a win-win business. While the system operators gain an enhanced situational awareness, customers will also be notified of their energy usage patterns and help them saving energy and reshaping their energy consumption patterns.

2.4.2 Application Examples

In this section, two applications are explored to explain the advantages of the enhanced load model in power system control and operations.

2.4.2.1 Refined Power Flow Analysis

Once we established the enhanced load model, we can update the original load model in power flow calculation software to get refined power flow analysis results. In fact, currently, most power flow analysis software use constant P and constant Q model for the system load. The enhanced load model will replace the constant P and constant Q load model with a voltage related PQ model, where P and Q are both functions of voltage. Moreover, the time-variant load model enables time series simulation with load's P and Q as functions of both time and voltage.

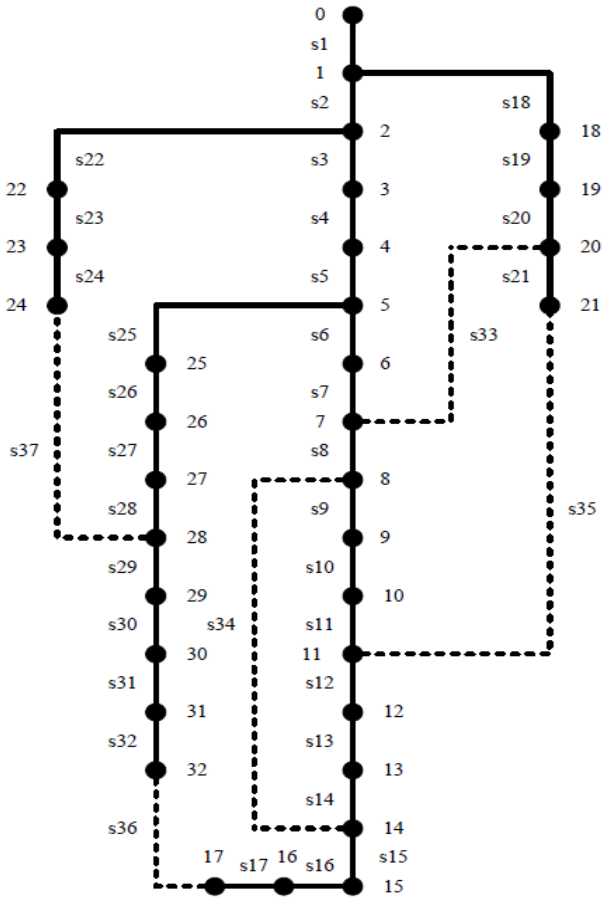

Figure 2.21 Baran & Wu 33-bus test system

To show the differences between the constant P constant Q model and the enhanced load model in power flow analysis, we introduce a 33-bus test system (Baran & Wu test system), which consists of 33 buses and 32 branches, shown in Figure 2.21. The test system information is shown in Table 2.4 and Table 2.5.

Table 2.4 Test system bus data

Bus Data (S_Base = 10MVAR)		
Bus Num.	Real Power (pu)	Reactive Power (pu)
0	0	0
1	0.01	0.006
3	0.009	0.004
4	0.012	0.008
5	0.006	0.003
6	0.006	0.002
7	0.02	0.01
8	0.02	0.01
9	0.006	0.002
10	0.006	0.002
11	0.0045	0.003
12	0.006	0.0035
13	0.006	0.0035
14	0.012	0.008
15	0.006	0.001
16	0.006	0.002
17	0.006	0.002
18	0.009	0.004
19	0.009	0.004
20	0.009	0.004
21	0.009	0.004
22	0.009	0.004
23	0.009	0.005
24	0.042	0.02
25	0.042	0.02
26	0.006	0.0025
27	0.006	0.0025
28	0.006	0.002
29	0.012	0.007
30	0.02	0.06
31	0.015	0.007
32	0.021	0.01
33	0.006	0.004

Table 2.5 Test system line data

Line Data			
In Bus	Out Bus	Resistence(Ohm)	Reactence(Ohm)
0	1	0.0922	0.047
1	2	0.493	0.2511
2	3	0.366	0.1864
3	4	0.3811	0.1941
4	5	0.819	0.707
5	6	0.1872	0.6188
6	7	0.7114	0.2351
7	8	1.03	0.74
8	9	1.044	0.74
9	10	0.1966	0.065
10	11	0.3744	0.1238
11	12	1.468	1.155
12	13	0.5416	0.7129
13	14	0.591	0.526
14	15	0.7463	0.545
15	16	1.289	1.721
16	17	0.732	0.574
1	18	0.164	0.1565
18	19	1.5042	1.3554
19	20	0.4095	0.4784
20	21	0.7089	0.9373
2	22	0.4512	0.3083
22	23	0.898	0.7091
23	24	0.896	0.7011
5	25	0.203	0.1034
25	26	0.2842	0.1447
26	27	1.059	0.9337
27	28	0.8042	0.7006
28	29	0.5075	0.2585
29	30	0.9744	0.963
30	31	0.3105	0.3619
31	32	0.341	0.5302

Two scenarios are created to ensure a fair comparison between the two models, and the simulation results show that the new model is more accurate and shows improvements on the power flow analysis results. It is assumed that in the base case, the voltage level at substation level is 1.05 times of the nominal voltage value. Then a disturbance is introduced by changing the voltage from 1.05 times the nominal voltage to nominal voltage at substation (bus No.0).

In Scenario 1, the real power and reactive power usage of each bus are fixed when the disturbance is introduced. However, in Scenario 2, the reactive power of each bus is a function of bus voltage, and the real power usage is still fixed due to the poor correlation between real power usage and voltage deviation. As a result, to get the updated load after a disturbance, several iterations of power flow analysis is needed. The load is updated every iteration according to the voltage deviation, until the voltage stabilizes, as is shown in Figure 2.22.

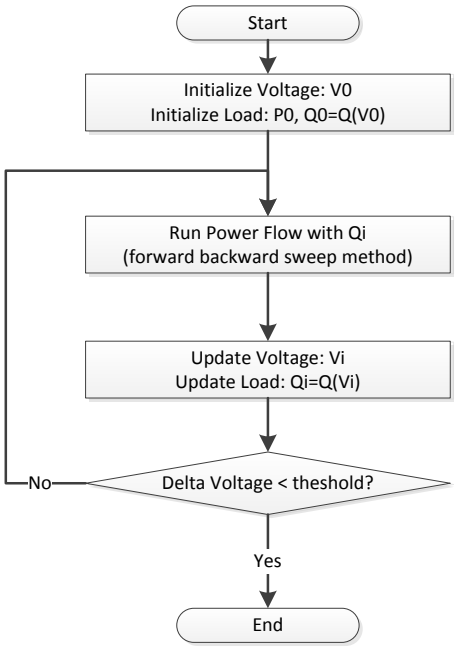

Figure 2.22 Refined power flow analysis flow chart

Figure 2.23 shows the power flow voltage differences between the constant PQ model and the enhanced load model. The results show that the new load model can effectively capture the reactive power consumption changes of the loads at the presence of voltage perturbation. In Scenario 2, the reactive power usage of consumer is a function of system voltage. When the voltage at substation decreases, the voltage at the consumer end will also decrease, which leads to a decreased reactive power consumption. Although the decreased reactive power consumption in return pushes the system voltage up, it cannot completely compensate the voltage drop at the substation, shown in Figure 2.23.

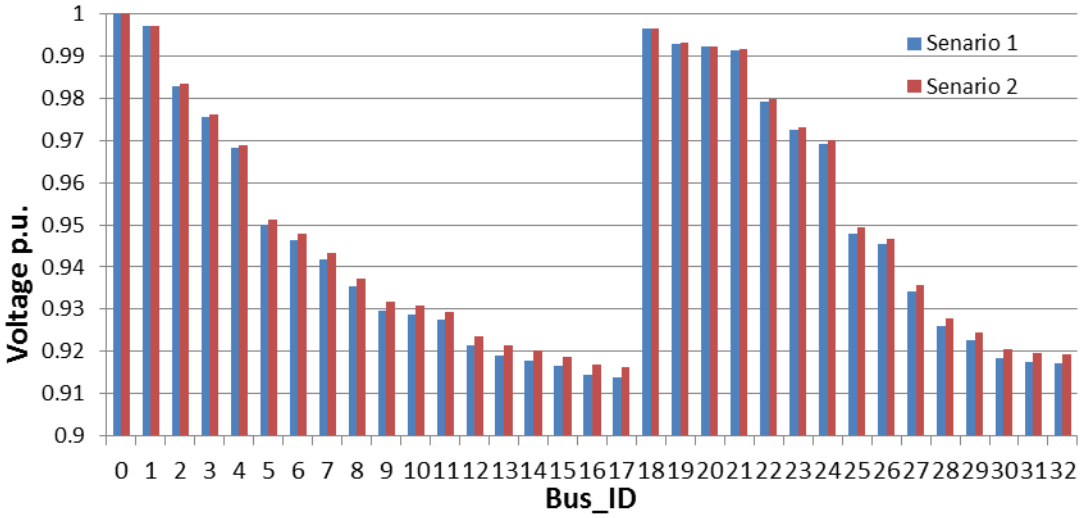

Figure 2.23 Refined power flow results comparison

2.4.2.2 Dynamic Distribution Network Reconfiguration

Distribution network reconfiguration (DNR) is an advanced distribution network application which seeks to minimizing the system operation losses by changing the topology of the tree structured network. Traditionally, the configuration of the distribution network topology is fixed or pre-optimized based on the power consumption condition on each bus. However, it is also known that the load on each bus is not fixed, and varies according to seasons, day types and different hours of the day. As a result, the time-variant load model built through the smart meter database can be used to refine the distribution network configuration, which allows system operators to change the system topology dynamically to reduce system losses over time.

For example, most distribution networks experience load migrations among residential loads and commercial & industrial loads between working hours and off-working hours. Let's assume Figure 2.24 and Figure 2.25 give the load information for both working hours and off-working hours for the abovementioned 33-bus test system.

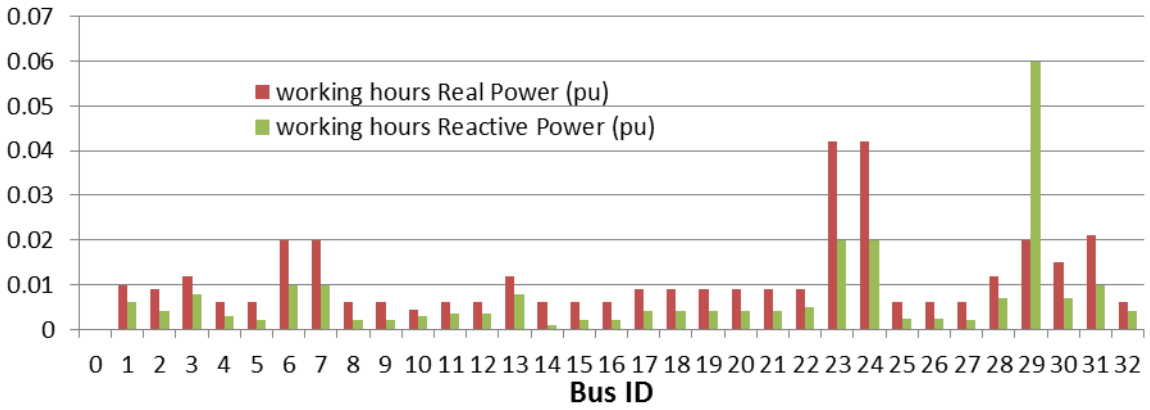

Figure 2.24 System loads during working hours

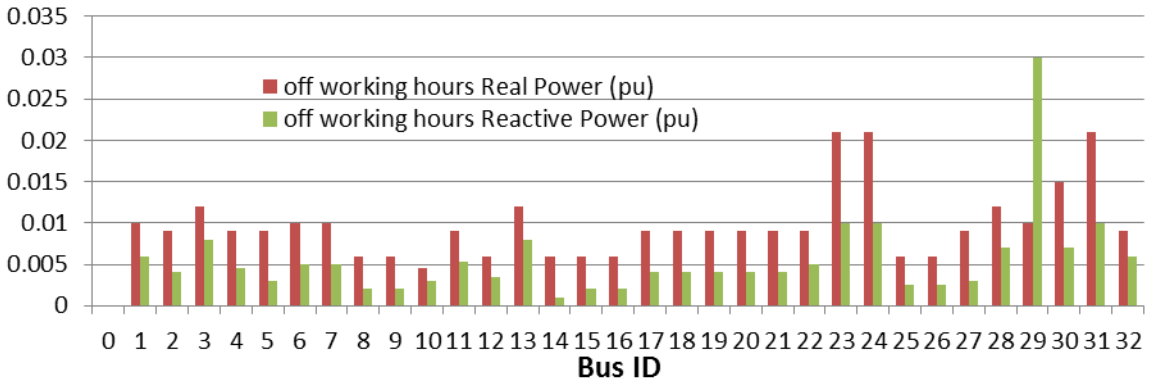

Figure 2.25 System loads during off-working hours

During the working hours, most industrial and commercial loads are at their peak hours which lead to the optimal system configuration, where branch 7, 9, 14, 31 and 37 are opened with the total system losses of 128.72kW. However, during the off-working hours, the load shifts from industrial and commercial load to residential load. This leads to the optimal system configuration, where branch 7, 9, 14, 28 and 31 are opened with the total system losses of 82.45kW. The topologies for the optimal network configurations of the two cases are shown in Figure 2.26 and Figure 2.27.

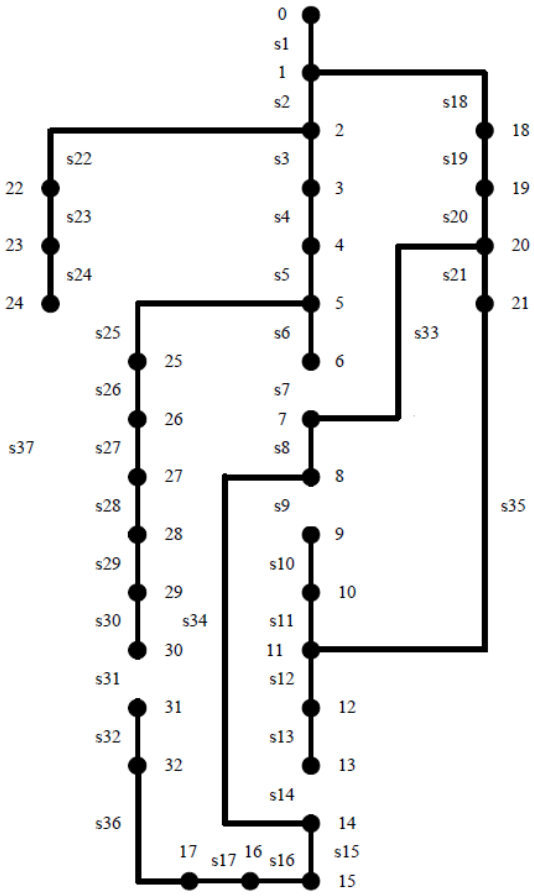

Figure 2.26 Working hours configuration

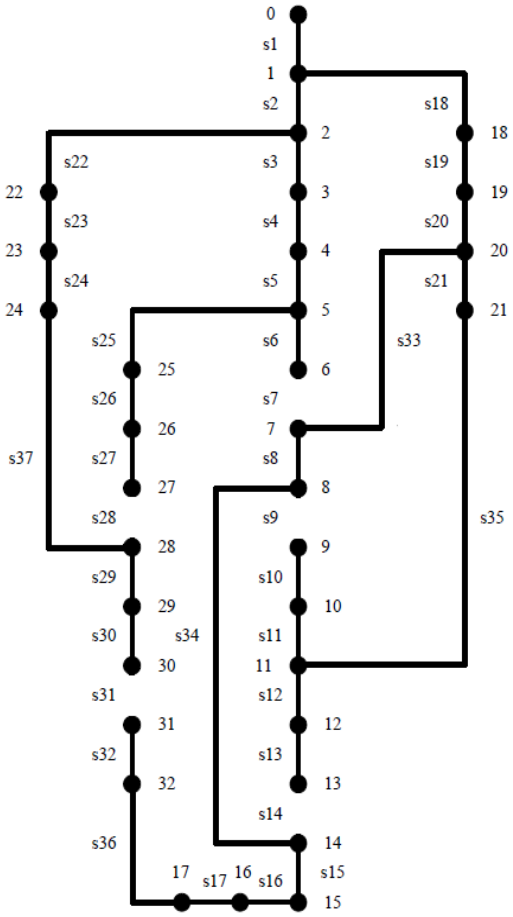

Figure 2.27 Off-working hour configuration

2.5 Conclusions

One of the critical needs for distribution system operations and planning applications is enhanced modeling of the load. In this study, we explore a novel load modeling method for enhanced load models. The new method differentiates itself from traditional measurement-based method and component-based method by taking advantages of the emerging data collected through the widely spread smart meters. Various data-mining algorithms and machine learning techniques are introduced to build the time-variant enhanced load model, which writes the real and reactive power usage of the load as a function of both time and voltage.

In this project, we study the load's time-variant properties by aggregating smart meter historical data through various time labels and hour partitioning processes. We further verify the data-mining-based model by the load condition assumption, which classifies the customer behaviors into routine behaviors and random behaviors. Meanwhile, various data-mining and machine learning algorithms are introduced such as K-subspace method, Davies-Bouldin Index (DBI) and Silhouette Coefficients.

In the first section, we introduce the smart meter deployment on Georgia Tech main campus as the data background of the study. We developed an interactive visualization tool, "Smart Grid Plotter", for easier visualization of the cumulative smart meter database. The visualization tool allows researchers navigate historical data collected by smart meters for all buildings on campus. Users can further configure and save the desired plot through various parameters menus.

In the second section, we propose a novel enhanced load modeling method based on data-mining and machine learning algorithms. The enhanced load model is a time-variant model that writes the load's active and reactive power usage as a function of both time and voltage. The detailed steps for the new load modeling method are further discussed in details through three aspects: data aggregation, hour partition and the load condition assumption.

In the third section, we further explore the smart meter data for both off-line and real-time utility functions. In the report, we show that as a very important information source, smart meter data (both real time data and historical data) can be the core of other 18 potential applications when combined with other data, such as weather data and GIS information. Two exemplary applications, refined power flow analysis and dynamic distribution network reconfiguration, are studied to show how smart meter data and the proposed enhanced load model improve power system analysis results and facilitate advanced energy efficiency operations.

In the future, the smart meter data will be more tightly integrated into vast majority of utility applications for both energy efficiency and reliability improvements.

Part 3. Exploiting Weather and Load Recording Data to Enhance Load Modeling

3.1 Research Objective

The goal of this research is to make a significant contribution toward an accurate dynamic model for the study of fault induced delayed voltage recovery (FIDVR) or other system disturbances of concern. That being the case, the overall goal is to populate the load model shown in Figure 3.1. The load model contains several elements, such as the substation transformer, feeder impedance, electronic load, and motor load. *The particular goal of this research is to accurately specify the air-conditioning motor load.* This will be done by identifying the temperature dependent and temperature independent load fractions. The basic premise is that the temperature dependent load in the summer season is due to air conditioning. The data required for this analysis is weather station data and power recording data of any kind. The weather station data is now freely available from the internet on sites such as www.noaa.gov. The power recording data may be Advanced Metering Infrastructure (AMI) data from a single customer, substation recording data for a feeder or substation, or system load data from an independent system operator such as PJM.

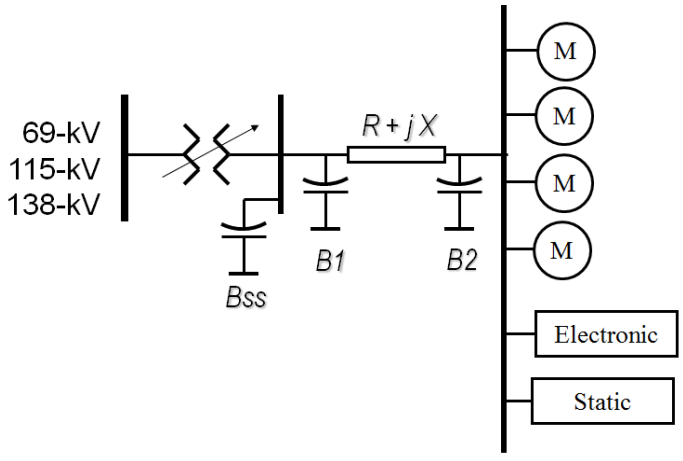

Figure 3.1 Load model for system simulation. All elements need to be specified but this research aims to specify air conditioning motor load.

3.2 Load and Temperature Relationships

Some, but not all, load is temperature dependent. This statement is somewhat intuitive, but we will use an example to illustrate the point. Figure 3.2 shows the normalized load and normalized temperature for a whole year for a power company on the east coast for the U.S.A (Maryland area). The load is highest when the temperature is highest in the summer due to the demand for air conditioning. However, load also increases slightly in the fall and winter when the temperature dips down toward its low extreme and electricity is used for heat.

Another way to view the load and temperature relationship is by plotting the real power load against the outdoor ambient temperature, as illustrated in Figure 3.3. Again, it is obvious that load is highest at the highest temperatures and that load increases as the temperature moves toward its negative extreme. However, Figure 3.3 also illustrates that a portion of the load is independent of temperature. The independent portion can be seen during mild temperatures that typically occur in the spring and fall. At these mild temperatures (around 15°C (59°F) in Figure 3.3), load does not vary with temperature. Examples of temperature-independent load would be lighting, office equipment and manufacturing equipment. Temperature-independent load is always present, whether the temperature is hot, cold, or mild, but it becomes apparent at mild temperatures.

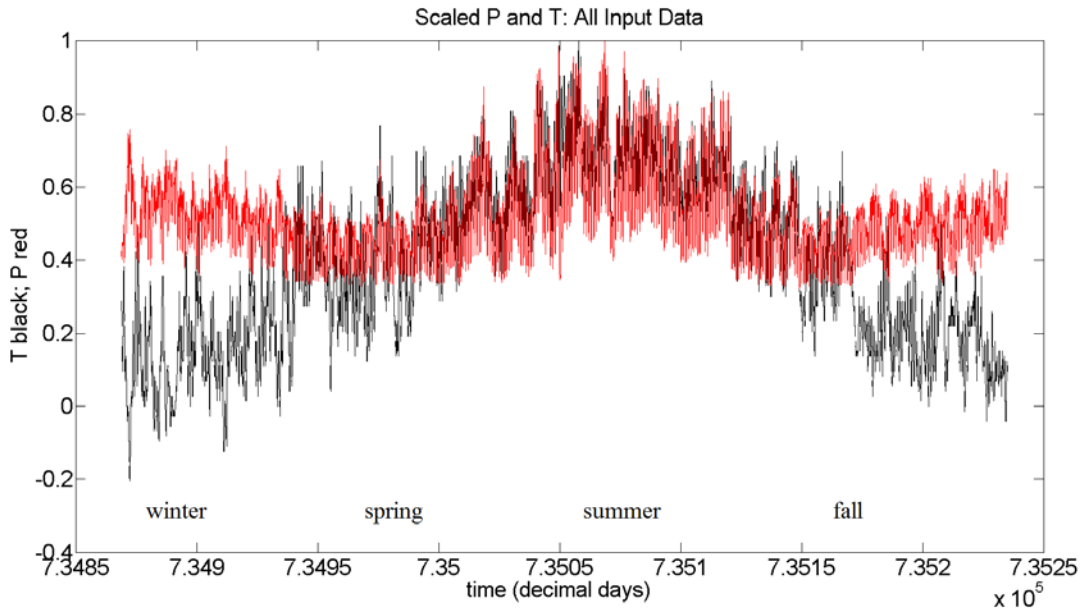

Figure 3.2 Normalized load and temperature for a single company for a full year. Load is highest during the highest temperature but it also rises when the temperature moves to the cold extremes.

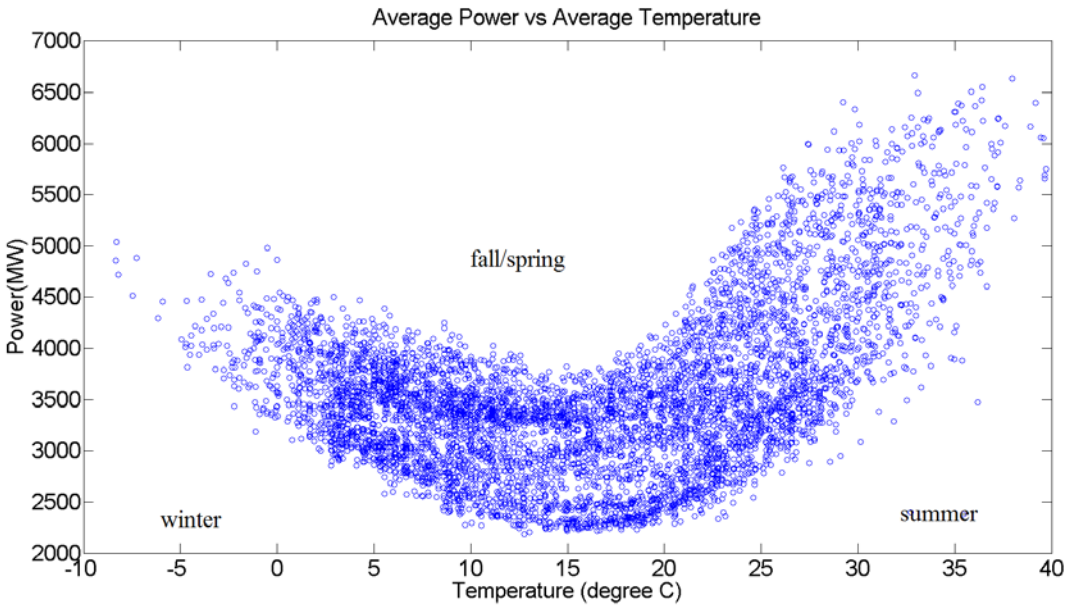

Figure 3.3 Real power versus the outdoor ambient temperature. Temperature-dependent load appears at high or low temperatures and temperature-independent load becomes apparent at mild temperatures (around 15°C (59°F) in this case).

Still further insight about the load-temperature relationship can be gleaned from looking at their variation with time, as shown in Figure 3.4 below. Here we see that load and

temperature both have a daily peak, but the peak in load occurs before the peak in temperature. Thus, the peaks in load and temperature do not occur at the same time. On the other hand, the graph shows that load generally increases when the temperature increases.

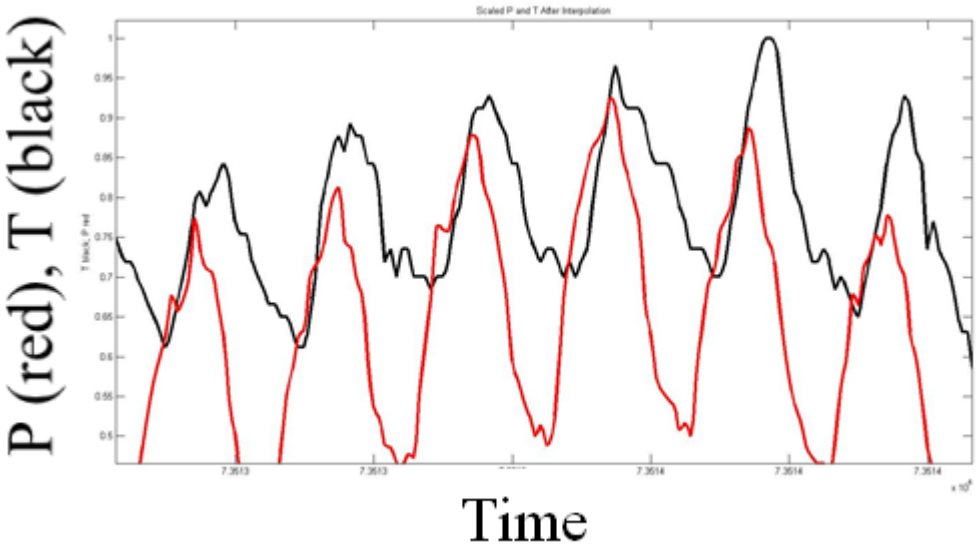

Figure 3.4 Approximately six days of normalized load and normalized temperature plotted against time.

In Figure 3.4, it is very interesting that the peak in temperature occurs *after* the peak in load. We would think that the peak in temperature would come first because high temperature causes high load. However, the outdoor ambient temperature may be acting as a surrogate for the factor that is really causing the rise in load; solar irradiance. We speculate that the sunlight first hits the building and street surfaces and causes them to warm up. This warming increases the demand for air conditioning and also increases the outdoor ambient temperature. Thus, irradiance may be a very useful variable for describing load variations, but solar irradiance data is not widely available like temperature data is.

The magnitude and timing of the peaks in load and temperature are a function of many factors, such as the dynamics of the overall thermal system. As a simple way to compensate for these dynamics, we calculate the cross-correlation between temperature

and power as shown in Figure 3.5. In this case, the highest cross-correlation is at a time lag of approximately 5 hours, indicating that temperature lags behind power by approximately 5 hours. As a simple way to compensate for these dynamics, we can shift the temperature data back in time by 5 hours. After making this shift, it is interesting to re-examine the power-temperature relationship by again plotting load versus temperature. This is shown in Figure 3.6.

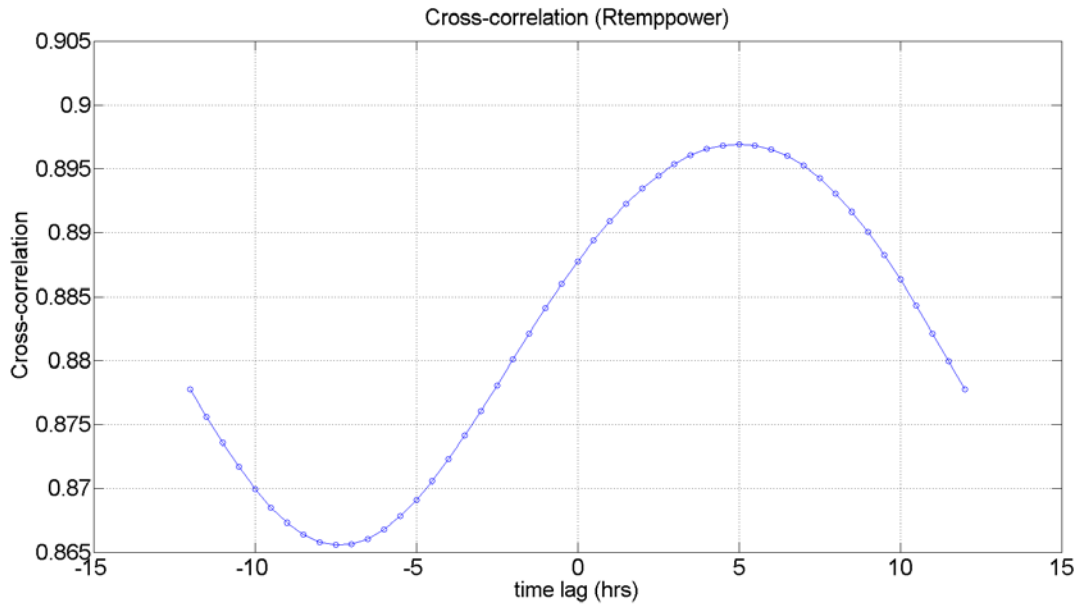

Figure 3.5 Cross-correlation of (real power) load and temperature.

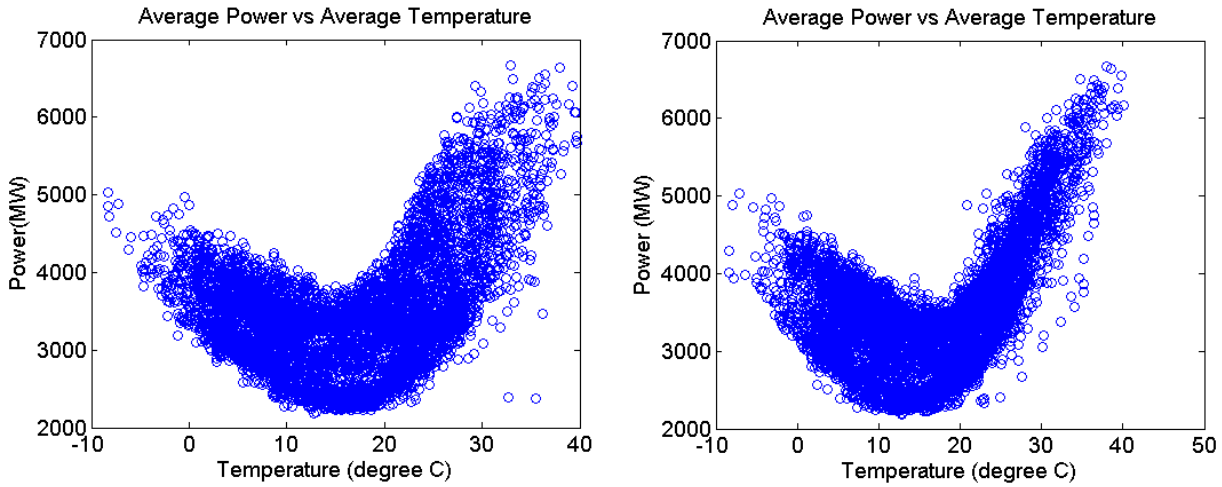

Figure 3.6 Load versus outdoor ambient temperature before (left) and after (right) a shift of the temperature data.

As shown in Figure 3.6, there is a stark change in the shape of the load versus temperature curve after the temperature data is shifted. Namely, the cloud of points becomes more tightly gathered after the time shift.

3.3 Segmented Regression to Create the Change-Point Curve

Ultimately, we intend to use the load versus temperature relationship to estimate the amount of air-conditioning load. In order to do this, we need to fit the data using some form of regression. For this, we chose segmented three line regression because it has physical meaning. This is commonly called a change-point curve in energy efficiency literature [32]. An example of these change-point curves is shown in Figure 3.7. The vertical scale of both graphs is the same so that a visual comparison of the results can be made. It is evident that the fit of the change-point curve is better for the graph on the right, after a shift in temperature has been made. However the improvement can be quantified by calculating the root mean square error (RMSE) of the curve fit. In doing so, we will confine the RMSE calculation to the high temperature portion of the curve because we are most interested in calculating the air-conditioning fraction of the load.

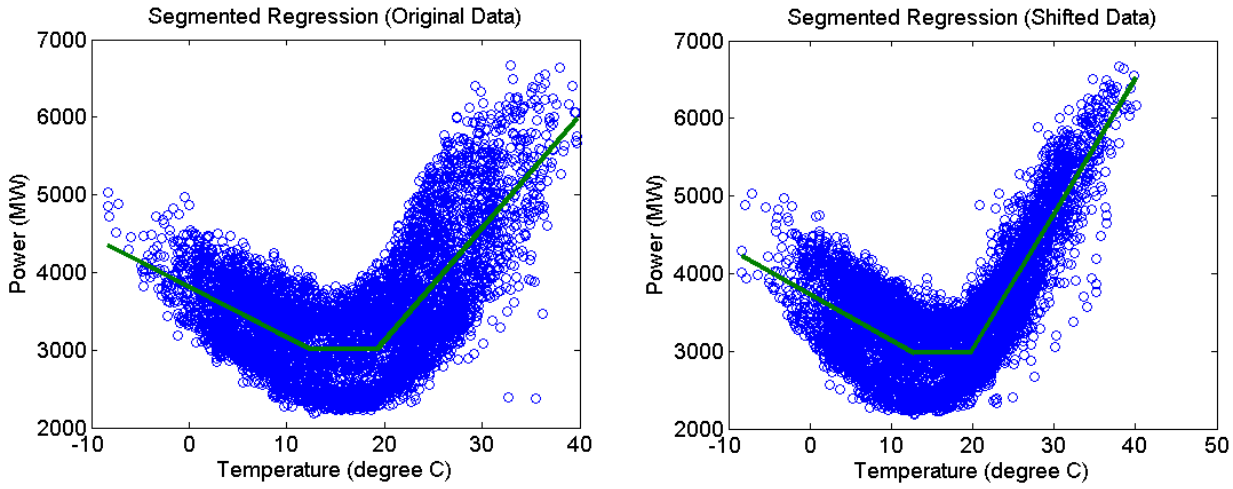

Figure 3.7 Segmented regression with (right) and without (left) shifting the temperature data. The resulting 3 segment line is referred to as a change-point curve.

Figure 3.8 shows the high-temperature portion of the data, which will be used for RMSE calculations. Again, it is apparent by visual inspection that the error is smaller after the temperature has been shifted by the optimal lag distance. However, the RMSE calculation

allows quantification of the difference. Figure 3.9 shows a bar plot comparing the RMSE before and after the temperature shift for 2012 data from five different companies in the PJM interconnection. In all cases, the error is lower after the shift in temperature data. This demonstrates the effectiveness of shifting the temperature data before performing the regression.

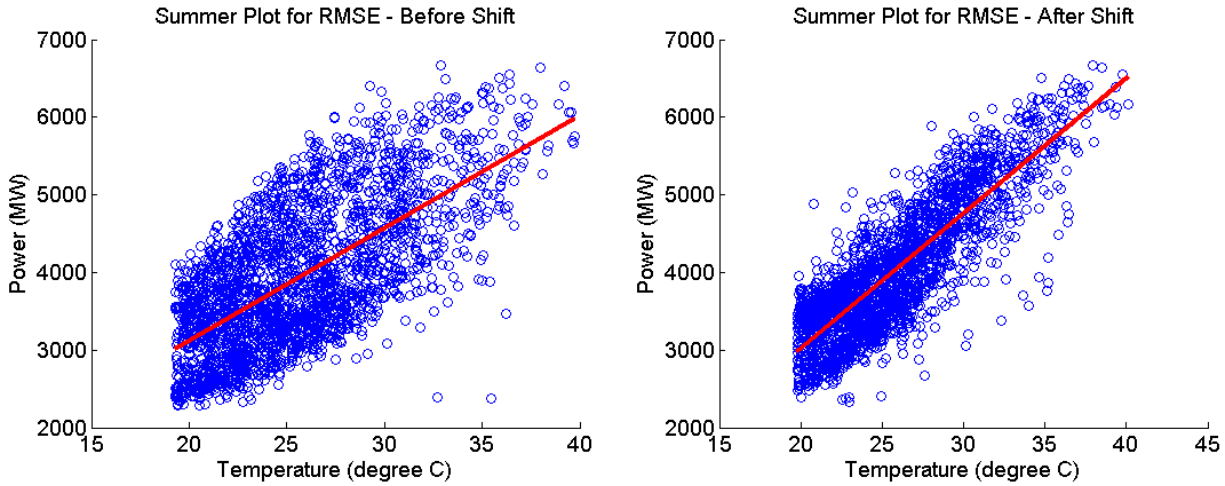

Figure 3.8 The high temperature portion of the data, which is used for RMSE calculations.

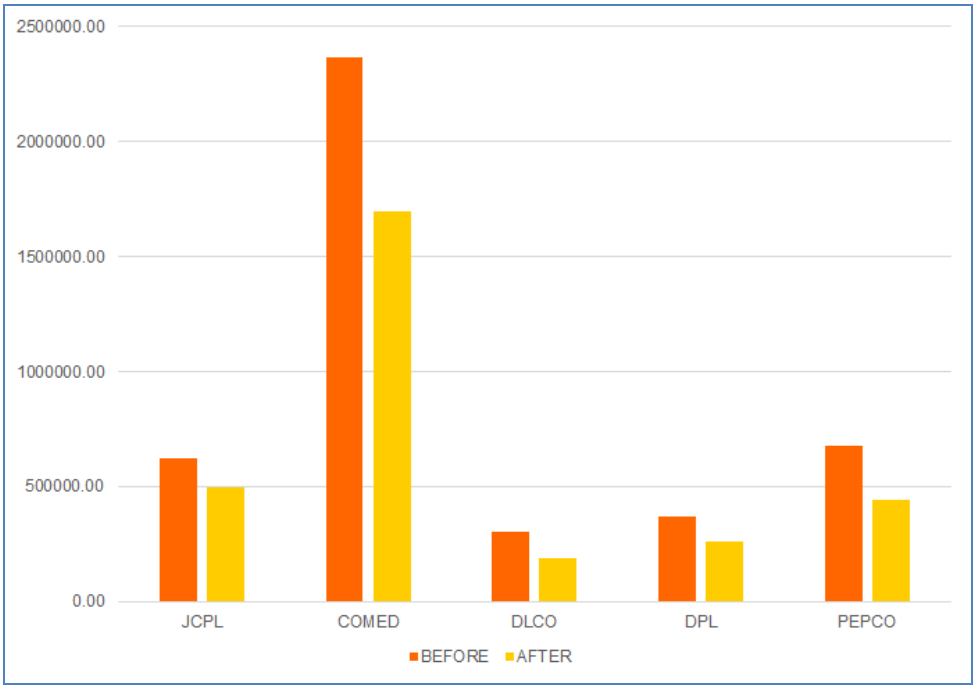

Figure 3.9 Root mean squared error (RMSE) of linear fit before and after shifting temperature data.

3.4 Calculation of Air Conditioning (AC) Load Fraction

The previous sections showed how to create an accurate segmented regression model to describe the relationship between active power load and outdoor ambient temperature. We refer to this model as the change-point curve. The center segment of the change-point curve is (intentionally) flat, indicating the range of temperature where power consumption is independent of temperature. Once the temperature reaches the upper "change-point", power consumption increases linearly with temperature, generally due to cooling demands. At this point, we make the assumption that this increase with temperature is due to cooling and that the majority of cooling load is motor load. As illustrated in Figure 3.10, the change-point curve allows us to divide the load into temperature-independent and temperature-dependent fractions. Since cooling load is of interest to us, the heating load (winter load) is not taken into consideration for this calculation. Once these fractions have been identified, we can calculate the temperature-dependent fraction, which we will refer to as the AC fraction, as follows:

$$AC\ Fraction = \frac{P_{TD}}{P_{TD} + P_{TI}}$$

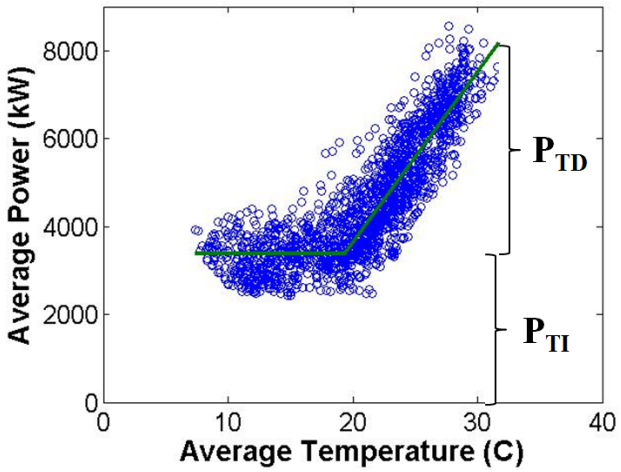

Figure 3.10 Illustration of the temperature-independent and temperature-dependent fractions of load.

Two notes can be made here about refinements that could be made to this calculation. First, air conditioning load generally consists of compressors, fans, and pumps; all devices that are run by motors. Thus, air conditioning load is motor load. Secondly, more

research should be done to make sure that non-motor loads are not contributing to the temperature dependent fraction. For example, we may find that the hotter the weather is, the more people stay indoors and watch television. Thus, it may be discovered that television load is temperature dependent. If that were found, a simple factor can be applied to the calculation to indicate how much of the temperature-dependent load is air-conditioning load. Thus, we introduce a factor γ into the foregoing equation to indicate how much of the temperature-dependent load is air conditioning load:

$$AC\ Fraction = \gamma\left(\frac{P_{TD}}{P_{TD} + P_{TI}}\right)$$

3.5 Temperature-Humidity Index as a Potential Weather Variable

Many of the example data sets used in this research are from PJM because the hourly historical load data is publically available. However, for load forecasting purposes, PJM recommends the usage of a temperature-humidity index (THI) instead of temperature alone [33]. That being the case, the natural question is whether THI is a better weather variable than temperature for calculating the AC fraction of load. In order to answer this question, we repeated our analysis using THI instead of temperature. An example of the results are shown in Figure 3.11 below.

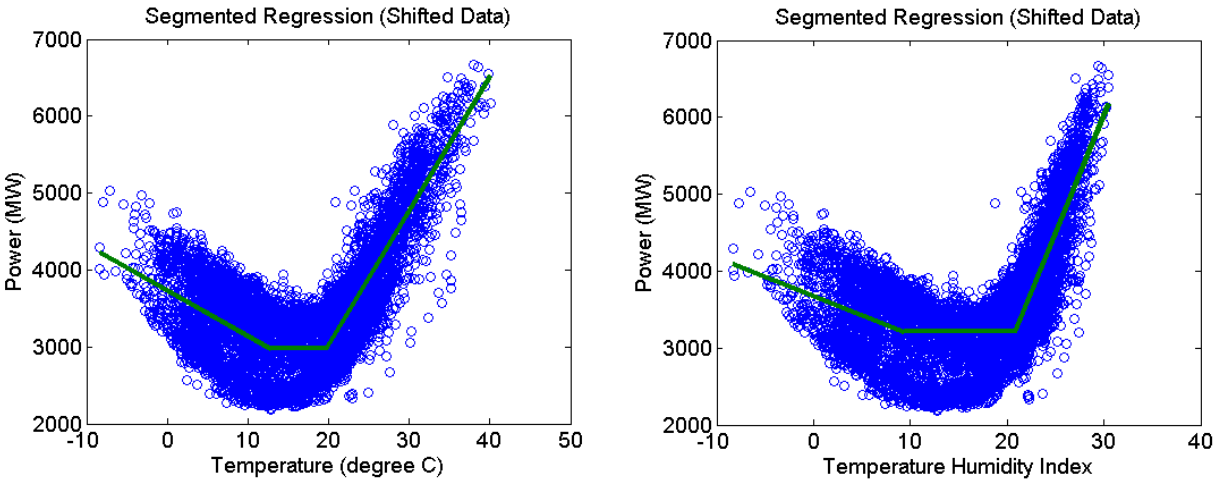

Figure 3.11 Examples of two different weather variables for estimating the AC fraction. Temperature is on the left, and temperature-humidity index (THI) is on the right.

In order to determine which of these weather variables is most descriptive, we repeated the earlier calculations of RMSE for the upper part of the change point curve and compared the error to that we obtained using temperature as the weather variable. The results of this comparison are shown in Figure 3.12. In all cases, temperature actually provides the lowest error in the linear fit and is therefore the more descriptive variable for estimating the AC fraction of the load. Thus, we will continue to use temperature for the independent weather variable in this analysis.

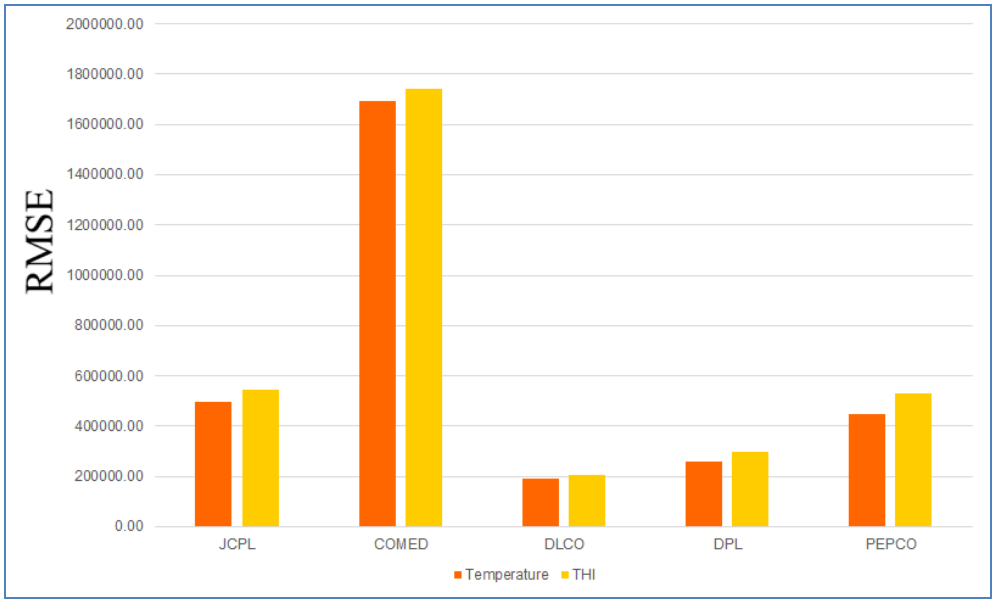

Figure 3.12 Comparison of weather variables by calculation of the RMSE in the cooling portion of the change point curve.

3.6 What Causes Separation in the Data Cloud?

We now turn our attention to refinements in the estimation of the AC fraction of the load. A major refinement came about by investigation of the load versus temperature data cloud. As shown in Figure 3.13, the data points seem to separate into two or more distinct clouds of data at medium to low temperatures. At first, it was not apparent what factor was causing this separation. We tried to divide the data into weekdays, weekends, and holidays, as shown in Figure 3.14. This showed that the weekend power consumption was generally lower, but the difference did not align with the cloud separation. Next, we investigated the difference in load between day and night. Figure 3.15 shows the average daily load profile for the whole year of data. The profile shows that the load is lowest

from 12am - 6am. Thus, we divided the data points in to two sets; the daytime set from 6am-midnight (18 hours), and the nighttime set from midnight to 6am (6 hours). These two sets are graphed in Figure 3.16, where we see that *the nighttime load accounts for the lower cloud of data and the daytime load accounts for the upper cloud of data*.

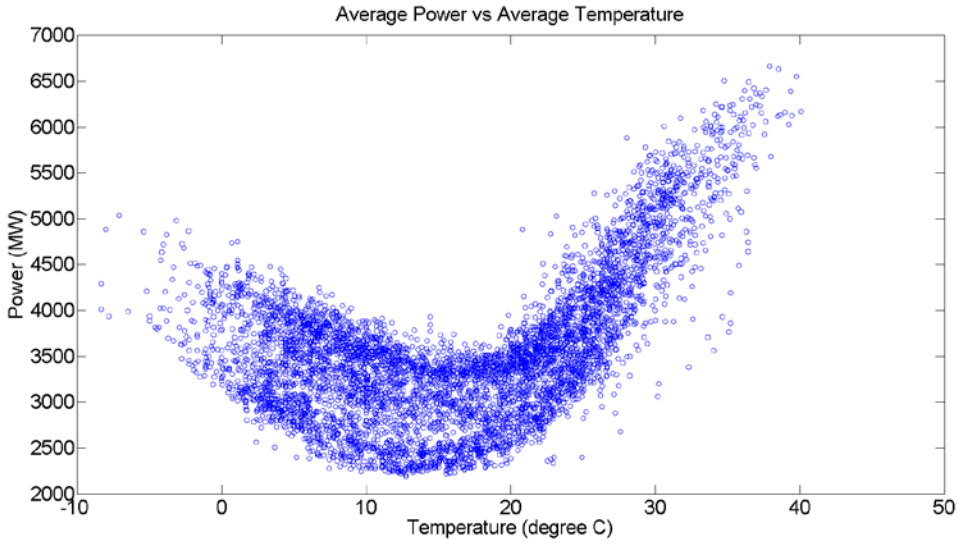

Figure 3.13 An example showing separation of the data cloud at medium to low temperatures.

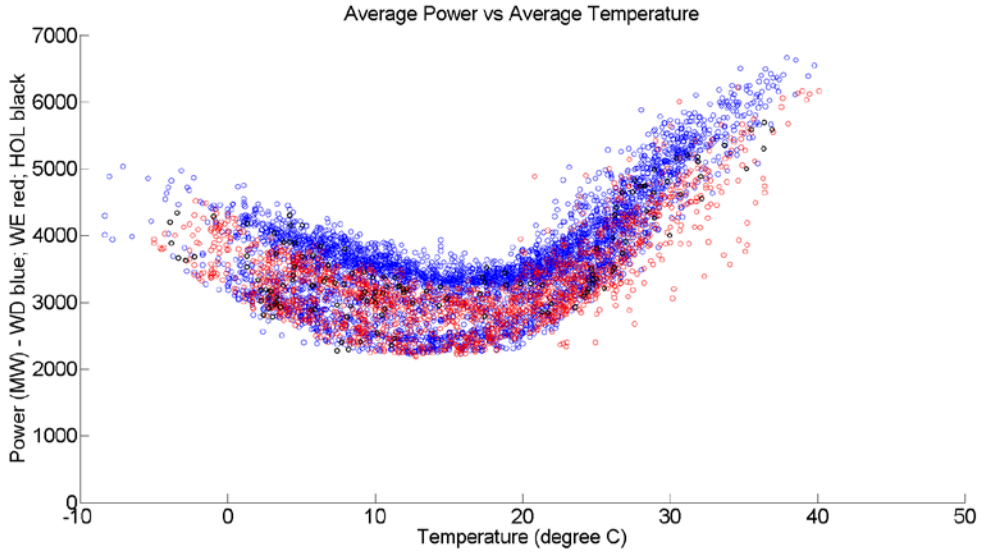

Figure 3.14 Separation of the data points into weekdays (blue), weekends (red), and holidays (black).

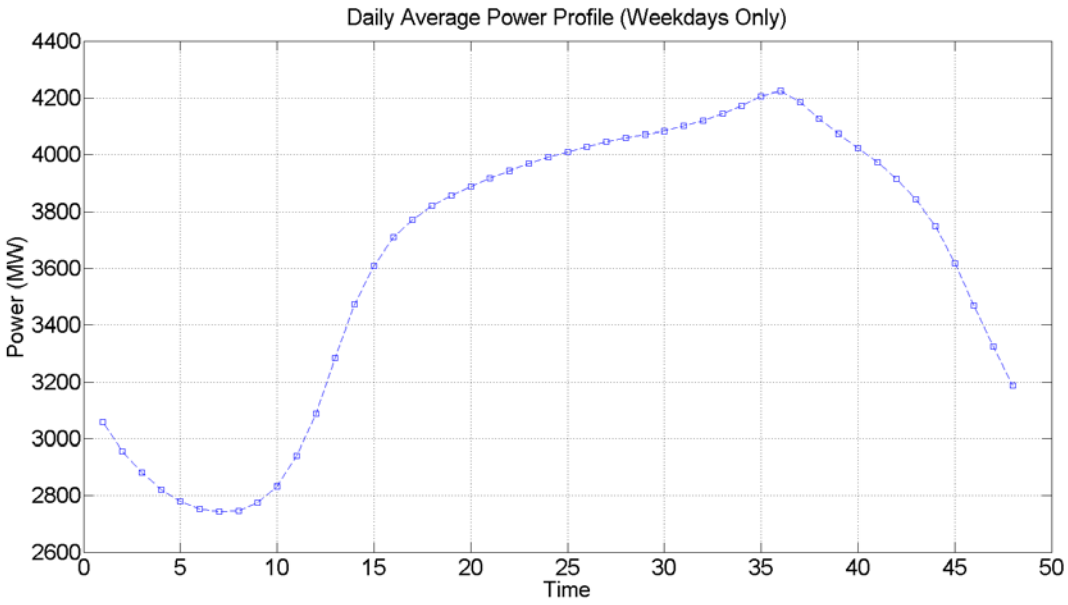

Figure 3.15 Average daily load profile (time in in ½ hour increments).

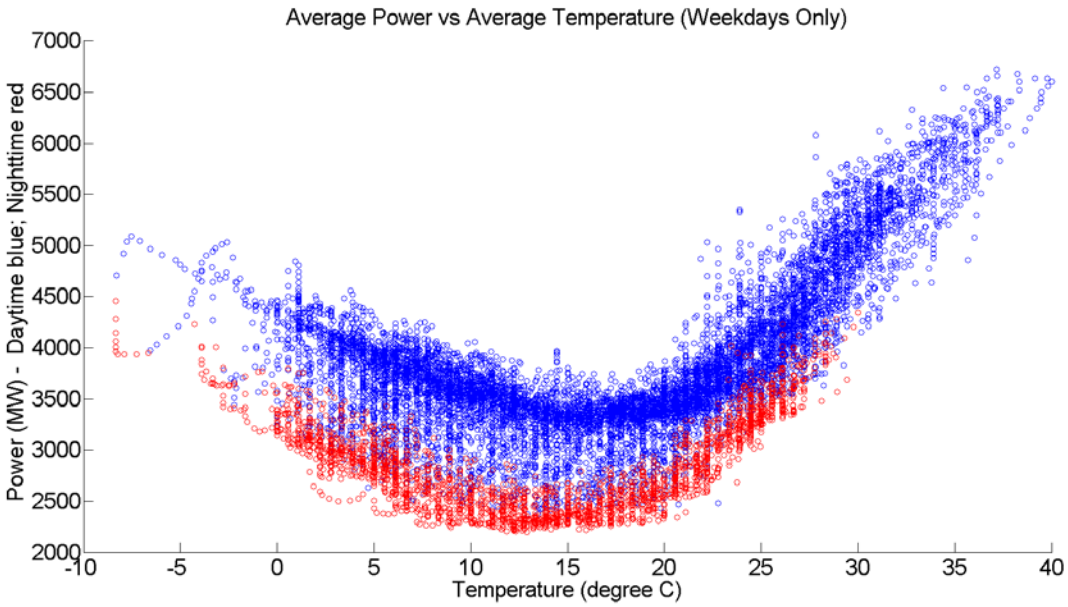

Figure 3.16 Data points separated into daytime load (blue) and nighttime load (red).

3.7 Change-point Curves for Four Different System States

In the previous section we showed that weekend load is lower than weekday load and that nighttime load is lower than daytime load. Next, we combine these variables to define four different system states; weekday daytime, weekday nighttime, weekend daytime,

and weekend nighttime. Data points from each of these system states are shown in Figure 3.17 below. This graph fits with our intuition that load is highest during the weekday daytime and lowest during the weekend nighttime. We also see that weekend daytime is the second highest loading state and weekday nighttime is the third highest (slightly higher than weekend nighttime).

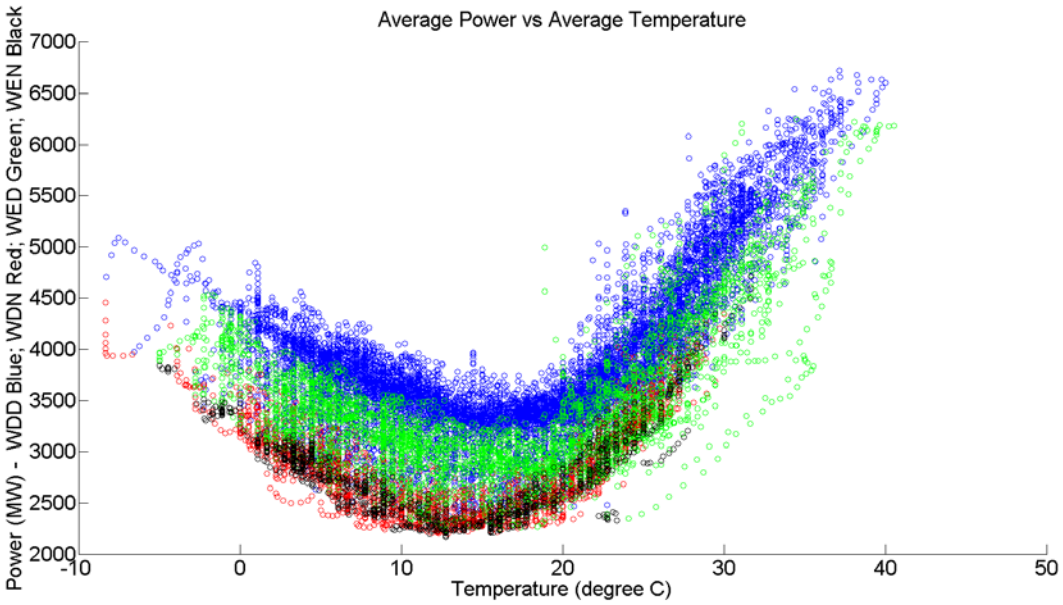

Figure 3.17 Load divided into four different system states; weekday daytime (blue), weekend daytime (green), weekday nighttime (red), and weekend nighttime (black).

Now that these four loading states for the system have been defined, we can go a step further and create a change-point curve for each different system state. This set of curves is shown in Figure 3.18, where we see that the data for each system state is well-behaved. Furthermore, as shown in Figure 3.19, we can overlay these change-point curves on a single graph to compare the behavior of each loading state. Once overlaid, we can clearly see the differences in loading states. We clearly see that the highest loading occurs during the weekday daytime and the lowest load occurs during the weekend nighttime. We also see that each system state has a different temperature independent load, which is the flat portion of the change-point curve. This is important for estimating the AC fraction of load because the temperature-independent load (P_{TI}) is in the denominator of the equation.

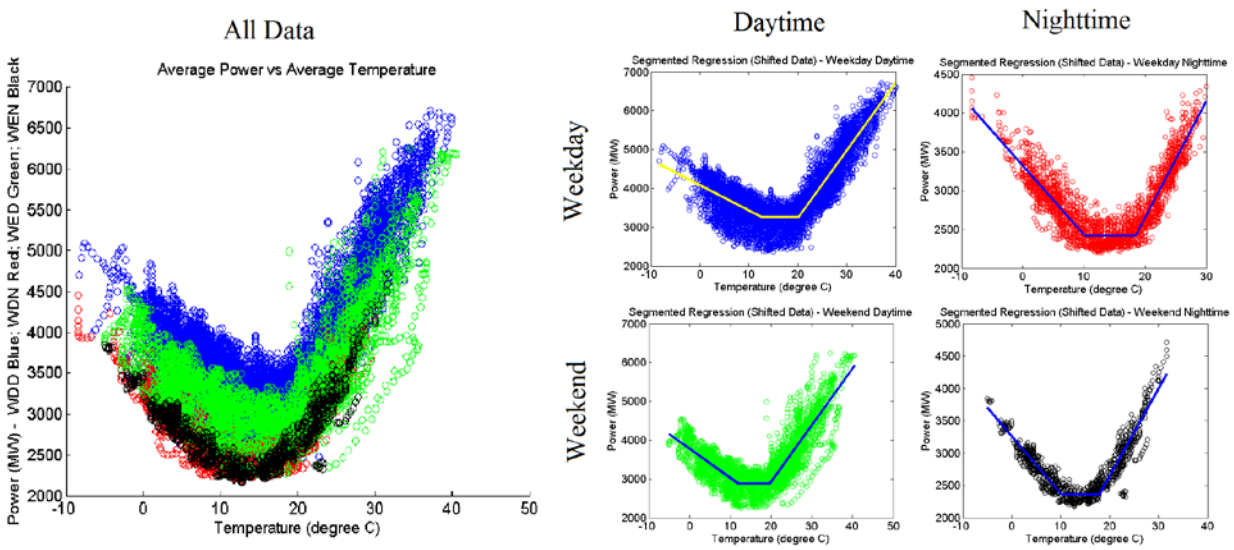

Figure 3.18 A change-point curve created for each different loading state.

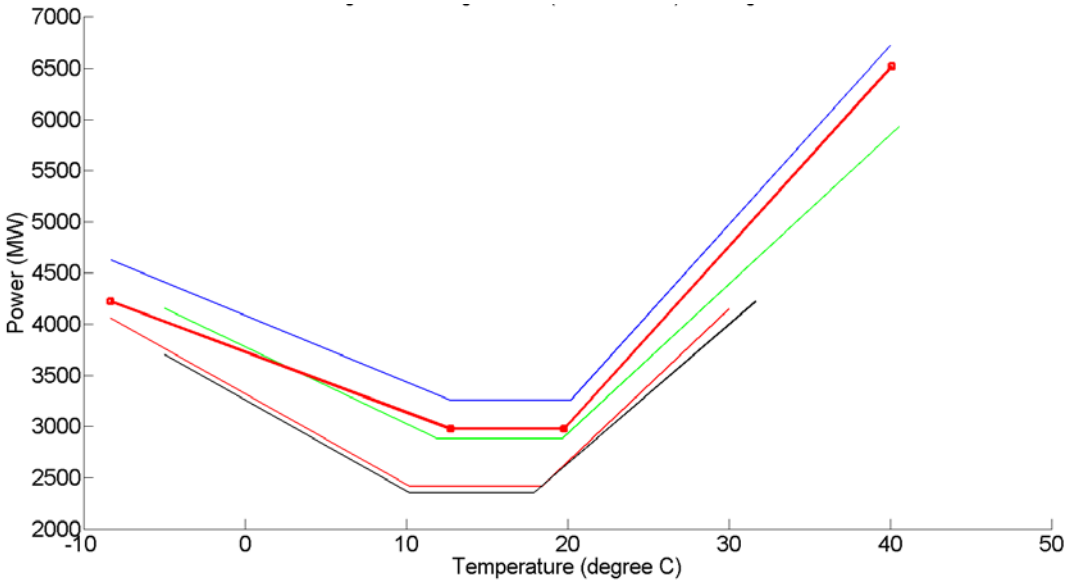

Figure 3.19 Change point curves for all data (red - bolded & w/squares)and the four different states of system loading; weekday daytime (blue), weekend daytime (green), weekday night (red), weekend night (black).

Each different change-point curve shown in Figure 3.19 will provide a different AC fraction of load because the temperature-independent power (P_{TI}) and the high temperature slope are different for each curve. For the sake of comparison, we calculated the AC fraction for the highest temperature encountered in this data set, which is approximately 40°C (104°F). The results are shown in Table 1 below, where we see that the AC fraction is higher for the nighttime loading states than for the daytime loading

states, assuming we calculate all AC fractions at the same temperature. This is because the temperature-independent load is lower for the nighttime states. We present these results only for the sake of comparison, recognizing that the maximum temperature during the night will not be as high as the maximum temperature during the daytime. The key point here is that when we estimate the AC fraction of load, the correct change-point curve should be used. For example, if we want to study an event that occurred in the daytime during the week, we should use the change-point curve corresponding to the weekday daytime as the basis for our calculations.

Table 3.1 AC fraction of load calculated at 40°C (104°F) for each system loading state

Data	**Plot**	AC Fraction at 40°C
All Data	RED (Bolded & W/Squares)	54.36%
Weekday Daytime	BLUE	51.64%
Weekend Daytime	GREEN	50.83%
Weekday Nighttime	RED	57.21%
Weekend Nighttime	BLACK	56.09%

3.8 Application to System Studies

The forgoing discussion has demonstrated how the AC fraction of load can be estimated based on historical load and temperature data. Now we turn our attention to how these concepts can be applied for the purposes of a system study. We assume that we wish to simulate a past event, so we will need to determine the AC fraction of load at the time of the event. The steps that must be taken to estimate the AC fraction are:

1) Specify the time of the event, t_E.
2) Calculate the offset time for the corresponding temperature: $t_E' = t_E + \Delta t$, where Δt is the lag time for optimal cross-correlation, as illustrated in Figure 3.5.
3) Obtain the temperature at the offset time calculated in step 2 above: $T(t_E')$.
4) Calculate the temperature dependent fraction of the power using the upper portion of the change-point curve (see Figure 3.20): $P_{TD} = mT(t_E') + b$.
5) Calculate the AC fraction:

$$AC\ Fraction = \gamma\left(\frac{P_{TD}}{P_{TD} + P_{TI}}\right)$$

These steps can easily be programmed and that is what we would suggest in order to make the concepts presented here practical and useful.

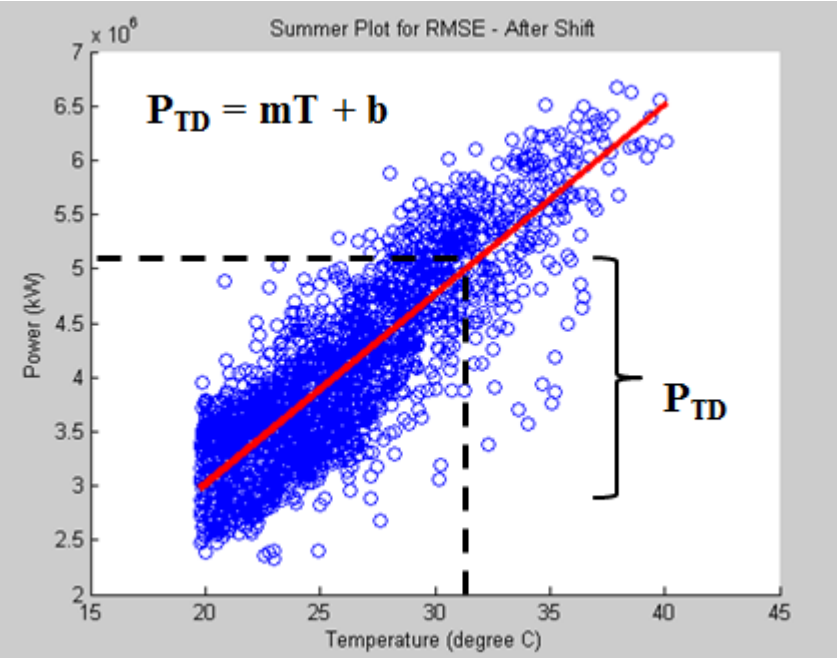

Figure 3.20 Application for system studies. The slope intercept equation for the high temperature portion of the change-point curve is used.

3.9 Conclusions

This work has shown a method for calculating the AC motor load using historical load and temperature data. Historical temperature data is now readily available from internet sources, and historical load data is more and more available through various recording means. The concepts have been illustrated using the load for various companies in the PJM interconnection, but the same concepts can be applied at the feeder level or even at the individual customer level.

Estimation of the AC fraction of load was performed using change-point curves and it was shown that separate change-point curves should be constructed for each different system loading state: weekday daytime, weekday nighttime, weekend daytime, and weekend nighttime. Of course, construction of these curves has been programmed so that the process is automated. Furthermore, a straightforward program can be written to

estimate the AC fraction of load at any particular time of interest. The steps for this application were shown.

Naturally, the amount of motor load will depend on the outdoor ambient temperature at the time of the event. The higher the temperature, the higher the AC motor load will be. These results use historical load and temperature data to take the guess work out of estimating the AC motor fraction of load.

3.10 Future Work

The foregoing results are meant to inform the load modeling process but they do not comprise a complete load modeling system. Some additional steps that will promote a complete load modeling process are:

1) Apply the foregoing analysis to some particular events and compare the derived motor load with that required to re-create the event.
2) Break the air-conditioning motor load into two basic fractions; a compressor fraction and a pump/fan fraction. This is important because compressors tend to have a linear torque-speed curve but fans and centrifugal pumps have a torque that varies with the square of the speed. Consequently, when a fan or centrifugal pump slows down, a great deal of load relief is obtained. On the other hand, when a compressor slows down, much less reduction is load torque is seen. Thus, these two loads should be modeled as separate motor loads.
3) Apply this analysis to a single customer where the size of the air-conditioning load is known (or at least the upper bound is known). This would allow quantification of how much of the temperature-dependent load is actually air-conditioning load.

References

[1] A. Rouhani and A. Abur, “Improving performance of dynamic state estimators under unknown load changes,” in Proc. IEEE Power & Energy Society General Meeting, July 21-25, 2013, Vancouver, CA.

[2] A. Rouhani and A. Abur, “Distributed implementation of an augmented state dynamic estimator,” in Proc. IEEE North American Power Symposium Sep. 22-24, 2013, KS, USA.

[3] W. Kao, C. Lin, C. Huang, Y. Chen, and C. Chiou, “Comparison of simulated power system dynamics applying various load models with actual recorded data,” IEEE Trans. Power Syst., vol. 9, no. 1, pp.248-254, Feb. 1994.

[4] J. V. Milanovic and I. A. Hiskens, “Effects of load dynamics on power system damping,” IEEE Trans. Power Syst., vol. 10, no. 2, pp.1022-1028, May. 1995.

[5] P. Kundur, Power System Stability and Control. New York: McGraw Hill, 1993.

[6] S. Liu, “Dynamic-data driven real-time identification for electric power systems,” Ph.D. dissertation, Dept. Elect. Eng., University of Illinois, Urbana-Champaign, 2009.

[7] Load Modeling for Power Flow and Transient Stability Computer Studies, EPRI, Rep., Jan. 1987, vol. 1-4, General Electric Company.

[8] D. Karlsson and D. Hill, “Modeling and identification of nonlinear dynamic loads in power systems,” IEEE Trans. Power Syst., vol. 9, no. 1, Feb. 1994.

[9] J. Wang, H. Jiang, C. Chang, and A. Liu, “Development of a frequency-dependent composite load model using the measurement approach,” IEEE Trans. Power Syst., vol. 9, no. 3, pp.1546-1556, Aug. 1994.

[10] D. SAGI, “Physically based electric load estimation using fuzzy-neuro systems,” Ph.D. dissertation, Dept. Elect. Eng., New Mexico State Univ., Las Cruces, New Mexico, May 2009.

[11] C. Lin, A. Chen, C. Chiou, C. Huang, H. Chiang, J. Wang, and L. Fekih-Ahmed, “Dynamic load models in power system using the measurement approach,” IEEE Trans. Power Syst., vol. 8, no. 1, pp.309-315, Feb. 1993.

[12] I. R. Navarro, “Dynamic load models for power systems,” Ph.D. dissertation, Dept. of Ind. Elect. Eng. and Auto., Lund University, Lund, 2002.

[13] H. Renmu, M. Jin, and D. Hill, “Composite load modeling via measurement approach,” IEEE Trans. Power Syst., vol. 21, no. 2, pp.663-672, May 2006.

[14] D. Han, J. Ma, R. He, and Z. Dong, “A real application of measurement-based load modeling in large-scale power grids and its validation,” IEEE Trans. Power Syst., vol. 24, no. 2, pp.1756-1764, Nov. 2009.

[15] A. Najafabadi and A. Alouani, “Real time estimation of sensitive parameters of composite power system load model,” in Proc. IEEE T&D Conf. May. 07-10, 2012, FL, USA.

[16] H. Bai, P. Zhang, and V. Ajjarapu, “A novel parameter identification approach via hybrid learning for aggregate load modeling,” IEEE Trans. Power Syst., vol. 24, no. 3, pp.1145-1154, Aug. 2009.

[17] S. Guo and T. Overbye, “Parameter estimation of a complex load model using phasor measurements,” in Proc. IEEE PECI Conf. Feb. 24-25, 2012, IL, USA.

[18] D. Simon, Optimal State Estimation: Kalman, H Infinity and Nonlinear Approaches. New Jersey: John Wiley & Sons, 2006.

[19] Asare-Bediako, B., et al. (2013). Day-ahead residential load forecasting with artificial neural networks using smart meter data. PowerTech (POWERTECH), 2013 IEEE Grenoble.

[20] Koponen, P. (2012). Short-term load forecasting model based on smart metering data: Daily energy prediction using physically based component model structure. Smart Grid Technology, Economics and Policies (SG-TEP), 2012 International Conference on.

[21] Stephen, B.; Mutanen, A.J.; Galloway, S.; Burt, G.; Jarventausta, P., "Enhanced Load Profiling for Residential Network Customers," Power Delivery, IEEE Transactions on , vol.29, no.1, pp.88,96, Feb. 2014

[22] Xiaochen Zhang, Grijalva, S., Mathew Reno, “A Time-Variant Load Model Based on Smart Meter Data Mining”, accepted by IEEE PES 2014 General Meeting.

[23] IEEE Committee, Load representation for dynamic performance analysis. IEEE Transections on Power Systems, 1993, (2),472-482.

[24] T. Frantz, T. Gentile, S. Ihara, N. Simons, M. Waldron, “Load Behaviour Observed in LILCO and RG&E Systems”, IEEE Trans., Vol. PAS-103, No. 4, April 1984.

[25] S.A.Y. Sabir, D.C Lee, “Dynamic Load Models Derived from data Acquired During System Transients,” IEEE Trans., Vol. PAS-101, September 1982, pp 3365 to 3372.

[26] Vaahedi, E., et al. (1987). "Load Models for Large-Scale Stability Studies from End-User Consumption." Power Systems, IEEE Transactions on 2(4): 864-870.

[27] Price, W. W., et al. (1988). "Load modeling for power flow and transient stability computer studies." Power Systems, IEEE Transactions on 3(1): 180-187.

[28] Davies, David L.; Bouldin, Donald W. (1979). "A Cluster Separation Measure". IEEE Transactions on Pattern Analysis and Machine Intelligence. PAMI-1 (2): 224–227.

[29] J. B. MacQueen (1967): "Some Methods for classification and Analysis of Multivariate Observations, Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability", Berkeley, University of California Press, 1:281-297

[30] Dingding Wang (Sch. of Comput. Sci., Florida Int. Univ., Miami, FL, USA); Ding, C.; Tao Li Source: Machine Learning and Knowledge Discovery in Databases. Proceedings European Conference, ECML PKDD 2009, p 506-21, 2009

[31] Chiang, H.-D., et al. (1997). "Development of a dynamic ZIP-motor load model from on-line field measurements." International Journal of Electrical Power & Energy Systems 19(7): 459-468.

[32] S. Doty and W.C. Turner, Energy Management Handbook, 7th ed., GA: The Fairmont Press, 2009.

[33] http://www.pjm.com/~/media/documents/manuals/M19.ashx

[34] EPRI, Load Modeling for Power Flow and Transient Stability Computer Studies, Volume 1: Summary Report, 1987.

[35] A. Meklin, J. Undrill, B. Lesieutre, W. Price, D. Chassin, R. Bravo, S. Yang, “Load Modeling in Power System Studies: WECC Progress Update,” IEEE Power and Energy Society General Meeting, Pittsburgh PA, 2008.