

Electricity Load Modeling in Frequency Domain

By

Shiyin Zhong

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Electrical Engineering

Robert P. Broadwater

Kwa Sur Tam

Richard A. Ashley

Virgilio A. Centeno

Saifur Rahman

November 28th, 2016

Blacksburg, VA

Keywords: Load Modeling, Load Profile, AMI, Spectral Analysis, Load Forecast,

Classification

Copyright © 2016, Shiyin Zhong

Electricity Load Modeling in Frequency Domain

Shiyin Zhong

Abstract

In today's highly competitive and deregulated electricity market, companies in the generation, transmission and distribution sectors can all benefit from collecting, analyzing and deep-understanding their customers' load profiles. This strategic information is vital in load forecasting, demand-side management planning and long-term resource and capital planning.

With the proliferation of Advanced Metering Infrastructure (AMI) in recent years, the amount of load profile data collected by utilities has grown exponentially. Such high-resolution datasets are difficult to model and analyze due to the large size, diverse usage patterns, and the embedded noisy or erroneous data points. In order to overcome these challenges and to make the load data useful in system analysis, this dissertation introduces a frequency domain load profile modeling framework. This framework can be used a complementary technology alongside of the conventional time domain load profile modeling techniques.

There are three main components in this framework: I) the frequency domain load profile descriptor, which is a compact, modular and extendable representation of the original load profile. A methodology was introduced to demonstrate the construction of the frequency domain load profile descriptor. II) The load profile Characteristic

Attributes in the Frequency Domain (CAFID). Which is developed for load profile characterization and classification. III) The frequency domain load profile statistics and forecasting models. Two different models were introduced in this dissertation: the first one is the wavelet load forecast model and the other one is a stochastic model that incorporates local weather condition and frequency domain load profile statistics to perform medium term load profile forecast.

7 different utilities load profile data were used in this research to demonstrate the viability of modeling load in the frequency domain. The data comes from various customer classes and geographical regions. The results have shown that the proposed framework is capable to model the load efficiently and accurately.

Electricity Load Modeling in Frequency Domain

Shiyin Zhong

General Audience Abstract

In today's highly competitive and deregulated electricity market, companies in the electricity power generation, transmission and distribution sectors can all benefit from collecting, analyzing and deep-understanding their customers' electricity consumption behavior. This strategic information is vital in forecasting and managing the future electricity demand. This information is also very important in utility company's long-term resource and capital planning.

With the proliferation of Advanced Metering Infrastructure (AMI) in recent years, the amount of electric load profile data collected by utilities has grown exponentially. Such high-resolution datasets are difficult to model and analyze due to the large size, diverse usage patterns, and the embedded noisy or erroneous data points. In order to overcome these challenges and to make the load data useful in system analysis, this dissertation introduces a frequency domain load profile modeling framework. This framework can be used a complementary technology alongside of the conventional time domain load profile modeling techniques.

There are three main components in this framework: I) the frequency domain load profile descriptor, which is a compact, modular and extendable representation of the original load profile. A methodology was introduced to demonstrate the construction of

the frequency domain load profile descriptor. II) The load profile Characteristic Attributes in the Frequency Domain (CAFID). Which is developed for categorizing the load profile data. III) The frequency domain load profile statistics and forecasting models..

7 different utilities load profile data were used in this research to demonstrate the viability of modeling load in the frequency domain. The data comes from various customer classes and geographical regions. The results have shown that the proposed framework is capable to model the load efficiently and accurately.

Dedication

This work is dedicated to

my dear family,

my loving parents,

and my teachers.

Acknowledgments

First, I would like to express my sincerely gratitude to my advisor and mentor, Dr. Robert Broadwater. His warm, upbeat personality, can-do attitude and constant encouragement have made huge impact in my life, both professionally and academically. Without his generous support and guidance, it will be impossible for me to finish this long journey.

I also want to thank my co-advisor, Dr. Kwa Sur Tam. He has guided me through this long and difficult research path with earnest and constructive opinions. His emphasis on academic rigor and relentless pursue of original ideas have shown me what research is all about. I am grateful to have him as a mentor.

Thanks also to Dr. Richard Ashley, Dr. Virgilio Centeno and Dr. Saifur Rahman for serving on my committee.

Last but not least, I would like to dedicate this thesis to my wife, who has been my biggest supporter over this long endeavor. Her devotion to our family has provided the perfect environment for me to work on my research. Her patient, wisdom and encouragement make this unusually long journey that much more enjoyable.

CONTENTS

Abstract	ii
General Audience Abstract	iv
Dedication	vi
Acknowledgments.....	vii
List of Figures	xi
List of Tables	xiv
Chapter 1: Introduction.....	- 1 -
1.1 Introduction to Load Profile Modeling in Frequency Domain	- 1 -
1.2 Challenges in Load Profile Modeling in the Frequency Domain.....	- 4 -
1.3 Research Objectives	- 6 -
1.4 Literature Review	- 6 -
1.4.1 Electricity Load Pattern Characterization	- 7 -
1.4.2 Electric Load Pattern classification.....	- 8 -
1.4.3 Electric Load Estimation.....	- 9 -
1.5 Dissertation Outline.....	- 11 -
Chapter 2: Load Profile in Frequency Domain.....	- 13 -
2.1 Methodology for Modeling Load Profiles in the Frequency Domain.-	13 -
2.2 Prerequisites for Load Profile Data in the Frequency Domain	- 17 -
2.2.1 Data Adequacy	- 17 -

2.2.2	Band-limitedness.....	- 18 -
2.2.3	Benefits of Representing Load with Frequency Components.....	- 35 -
2.3	Load Profile Statistics in Frequency Domain	- 37 -
2.3.1	Consistent Harmonic Patterns	- 40 -
2.3.2	High Degree of Certainty	- 41 -
2.3.3	Similar Patterns for Historical Data at Different Utilities.....	- 43 -
2.4	Summary	- 45 -
Chapter 3:	Hierarchical Classification of Load Profiles Based on Their Characteristic Attributes in Frequency Domain.....	- 47 -
3.1	Hierarchical Classification Tree.....	- 50 -
3.2	Load Profile Characteristic Attributes in the Frequency Domain.....	- 51 -
3.3	Classification Procedure.....	- 57 -
3.4	Hierarchical Frequency Domain Load Profile Classification	- 59 -
3.5	Comparison between Frequency Domain Hierarchical Load Profile Classification and Utilities' Current Practices	- 64 -
3.6	Load Profile Classification.....	- 69 -
3.7	Summary	- 70 -
Chapter 4:	Wavelet based Load Model for AMI Data	- 72 -
4.1	Discrete Wavelet Transformation	- 76 -
4.1.1	Multi-Resolution DWT	- 77 -

4.1.2	2D Multi-Resolution Wavelet Transformation	- 91 -
4.2	AMI Load Data Modeling using Wavelet.....	- 92 -
4.2.1	Multi-Resolution Wavelet Load Model	- 93 -
4.2.2	Classified Wavelet Load Model.....	- 97 -
4.3	Load Estimation with Wavelet Based Load Model	- 103 -
4.4	Summary	- 106 -
Chapter 5: Stochastic Load Model in Frequency Domain.....		- 108 -
5.1	Frequency Domain Stochastic Load model	- 110 -
5.2	Load Estimation Experiment.....	- 119 -
5.3	Load Estimation Envelopes.....	- 122 -
5.4	Load Model Performance Evaluation	- 123 -
5.4.1	Model Performance Metrics.....	- 123 -
5.4.2	Load Model Performance.....	- 125 -
5.4.3	Comparison with Other Models	- 127 -
5.5	Summary	- 130 -
Chapter 6: Contributions and Further Research		- 132 -
6.1	Contributions.....	- 132 -
6.2	Future Research.....	- 135 -
Reference		- 136 -

List of Figures

Figure 1-1: Average daily load profiles for commercial customers	- 1 -
Figure 1-2: Average daily load profiles for residential customer: non-winter months	
.....	- 1 -
Figure 1-3: Average daily load profiles for commercial customers	- 2 -
Figure 1-4: Average daily load profiles for residential customer: non-winter months	
.....	- 2 -
Figure 1-5: TLP for a typical office building.....	- 4 -
Figure 1-6: Time Domain Load Profile Description using 5 Parameters	- 5 -
Figure 2-1: Average daily load profiles for utility A commercial customers...-	13 -
Figure 2-2: Original and reconstructed (by using 7 harmonics) average daily load profiles for the Commercial customers in Utility A for March (PMEI = -0.27%, PTE = 0 hour, MME = 3.73% and MAPE=1.02%)	- 32 -
Figure 2-3: Original and reconstructed (by using 3 harmonics) average daily load profiles for the residential customers in Utility A in January (PMEI = 1.70%, PTE = 0 hour, MME = 4.35% and MAPE = 1.42%)......	- 33 -
Figure 2-4: Original and reconstructed (by using 6 harmonics) average daily load profiles for the residential customers in Utility A in July (PMEI = -0.51%, PTE = 0 hour, MME = 4.65% and MAPE = 1.65%)......	- 33 -
Figure 2-5: Plots of coefficient of variation for the average daily load profiles for two classes of customers of Utility B	- 42 -

Figure 2-6: Utility A, B residential & large commercial customer class multi-year monthly harmonics comparison.....	- 44 -
Figure 3-1: CAFD-Based Classification Method	- 48 -
Figure 3-2: a_{714H}/a_{51H} Box Plot	- 54 -
Figure 3-3: a_{357H}/a_{51H} Box Plot	- 55 -
Figure 3-4: Hierarchical Classification Tree.....	- 58 -
Figure 3-5: Partition of the feasible space created by θ_{102} and a_{714H}/a_{51H}	- 60 -
Figure 3-6: CAFD-Based Signatures	- 62 -
Figure 3-7: XML document for CAFD-based Signature.....	- 63 -
Figure 4-1: Sample AMI Load Profile Data	- 74 -
Figure 4-2: Multi-resolution DWT Frequency Band.....	- 78 -
Figure 4-3: Multi-Resolution IDWT	- 79 -
Figure 4-4: Sample 24-hour load profile DTW	- 81 -
Figure 4-5: Synthesized 24hour load profile at different resolution.....	- 86 -
Figure 4-6: 3 level synthesized load profile.....	- 87 -
Figure 4-7: Synthesized load profile	- 88 -
Figure 4-8: Synthesized load profile	- 88 -
Figure 4-9: Load Profile Detail Coefficient from 2-Level DWT	- 89 -
Figure 4-10: 2D Multi-Resolution DWT	- 91 -
Figure 4-11: Error of Synthesized Load Profile	- 95 -
Figure 4-12: Synthesized Load Profile Percentage Error	- 96 -
Figure 4-13: 5 AMI load profiles' pattern visualization	- 98 -
Figure 4-14: K-Mean Clustering Algorithm	- 99 -

Figure 4-15: Hourly Load Estimation % Error Histogram	- 104 -
Figure 5-1: Sample Commercial and Residential Transformer Normalized Daily Load Profiles under different WI.....	- 112 -
Figure 5-2: FDWSSM Flow Chart.....	- 120 -
Figure 5-3: Model Performance Metrics Comparison	- 125 -
Figure 5-4: FDWSSM vs. M1-M5 MAPE, CV Comparison	- 128 -

List of Tables

Table 2-1: Harmonics for the Average Daily Load Profile for the Commercial Customers in Utility A	- 29 -
Table 2-2: Number of Harmonics Needed to Achieve Different Levels of Performance Indices.....	- 32 -
Table 2-3: Number of Harmonics Needed to Reconstruct the Load Profile for Different Types of Customers in Utility B	- 34 -
Table 2-4: Statistics for the Magnitudes and Phase Angles of the First Three Frequency Components for the Residential Customers of Utility B and Utility C	- 38 -
Table 2-5: Statistics for the Magnitudes and Phase of the First Three Frequency Components for the Large Commercial Customers of Utility B and Utility C	- 39 -
Table 2-6: Harmonic Pattern for Different Customer Classes.....	- 40 -
Table 3-1: Common Load profile patterns.....	- 52 -
Table 3-2: Comparison of CAFD-based classes and Utility Commercial & Residential Load Classes	- 65 -
Table 3-3: Comparison of CAFD-based Classes and Utility Commercial Load Classes.....	- 67 -
Table 3-4: Load Profile Classification Test.....	- 69 -
Table 4-1: DB1 QMF Coefficients	- 78 -
Table 4-2: Sample 24hour load profile DTW	- 82 -
Table 4-3: $s_{A_3}(n)$ vs. Original Load Profile Energy.....	- 94 -
Table 4-4: Load profile classification	- 101 -

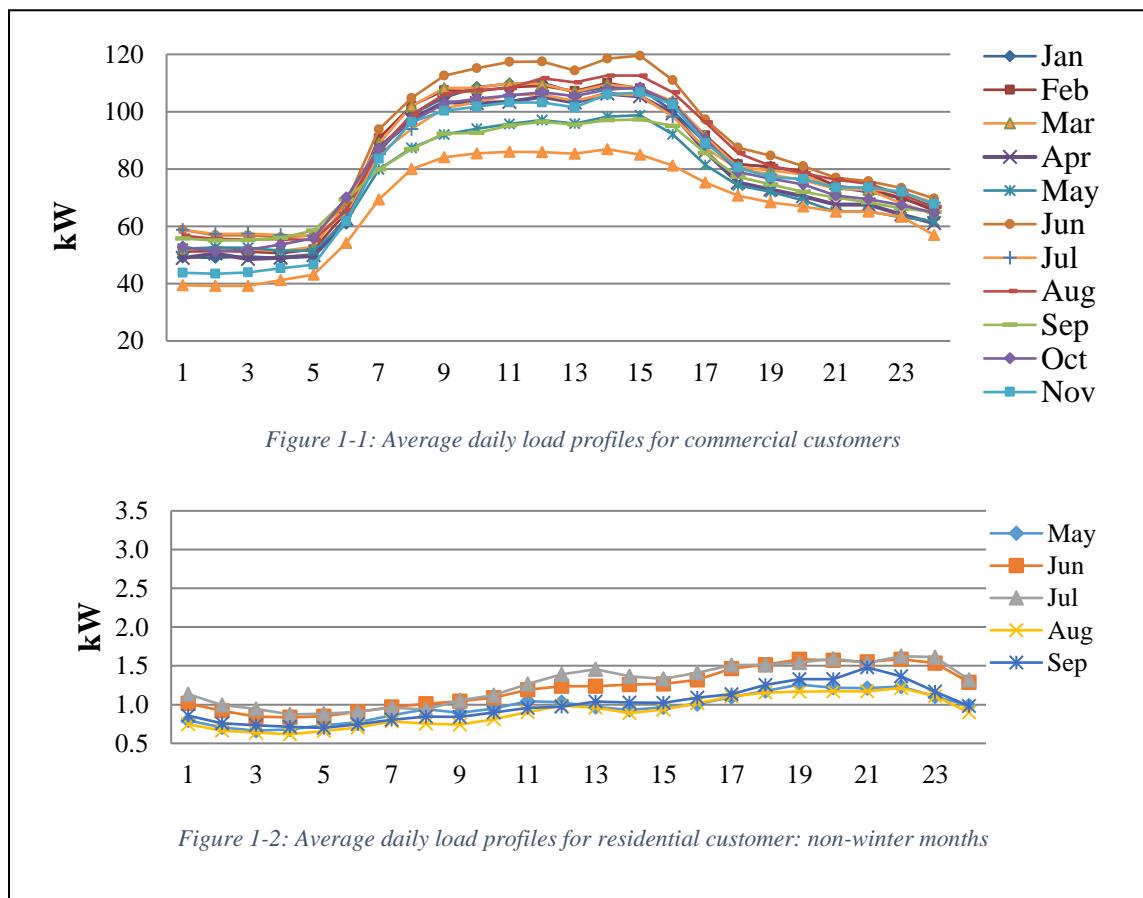
Table 4-5: Synthesized Load Profile Evaluations with PE(n)	- 102 -
Table 4-6: Data Reduction	- 103 -
Table 4-7: System Peak Load Estimation.....	- 104 -
Table 4-8: System Total Energy Estimation.....	- 105 -
Table 5-1: Daily Average Temperature \overline{Temp}_D Range for WI	- 111 -
Table 5-2: Normal Distribution Tests for Daily Peak, Load Shape Descriptor Magnitude and Phase	- 116 -
Table 5-3: Residential Transformer Daily Peak Partitioned by WI.....	- 116 -
Table 5-4: Residential Transformer Frequency Domain Daily Normalized Load Profile Coefficients Partitioned by WI	- 118 -
Table 5-5: The Number of Feeders/Transformers within Different Average Hourly Load Ranges.....	- 121 -
Table 5-6: Estimation Envelope vs. Original AMI Data	- 123 -
Table 5-7: Error Range Distributions for MAPE and Daily Peak MAPE	- 126 -
Table 5-8: Error Distributions for MAPTE	- 127 -
Table 5-9: Execution Time for FDWSSM Model Construction & Load Estimation	- 129 -

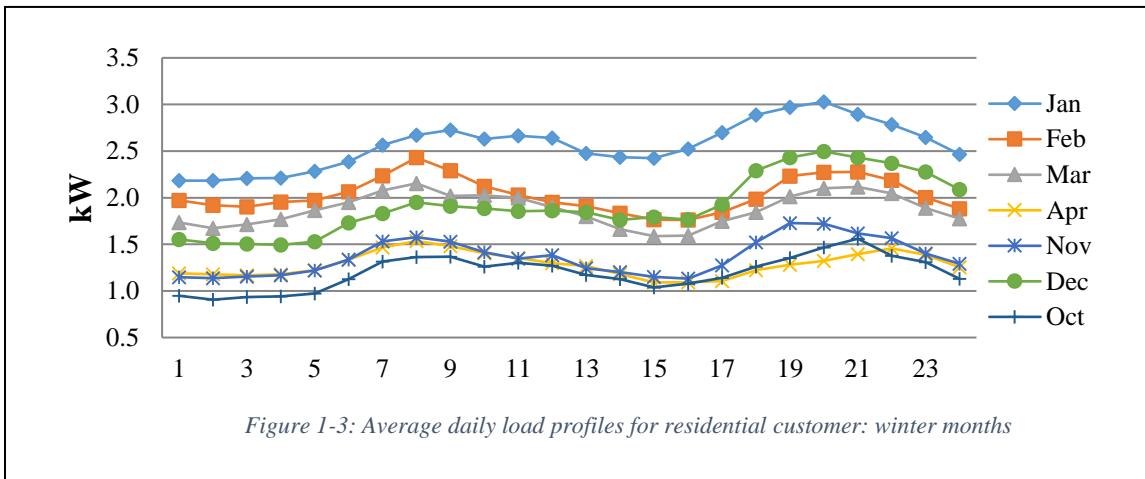
Chapter 1: Introduction

1.1 Introduction to Load Profile Modeling in Frequency Domain

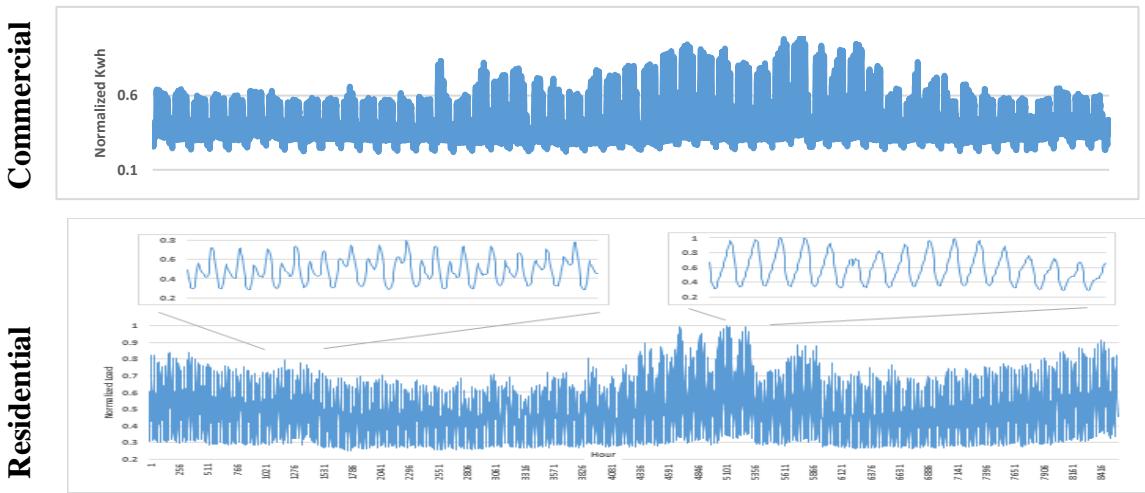
With the proliferation of AMI in recent years, utilities have accumulated a large amount of high-resolution electricity consumption data. Deep-mining this data will help utilities to gain strategic advantages in areas such as load forecasting, demand-side management, and resource planning.

Customer electricity usage patterns are very diverse and can be effected by many factors including geographical, demographical, and meteorological factors. For example, Figures 1-1 to 1-3 present daily load profiles for some sample commercial and residential customers from a large utility in the Midwest region of the United States.





The commercial customer has a set of very consistent load patterns throughout the year. The load pattern exhibits a long and flat high-load period during typical business hours (9am to 5pm). On the other hand, the residential customer has a more diverse set of load patterns. During the winter time there are two relative short and distinctive high-load periods (morning and evening). During non-winter months the high-load morning period shifts to a later time of the day. In addition to studying the daily load profiles, utilities often use the annual load profiles in load research and modeling. In Figure 1-4, there are two very different normalized annual load profile plots: the one on the top is from a commercial customer, the bottom one is from a residential customer. The plots exhibit very different



load patterns characteristics. For example, the residential customer load pattern has a strong correlation with both the hot and the cold weather days while the commercial customer load pattern is mainly affected by the hot weather. The residential customer has a double-peak daily load pattern in January and single-peak daily load pattern in July. The commercial customer has a consistent single peak, load daily pattern throughout the year, where the total daily energy consumption varies with the season.

The annual load profile of residential or commercial customers can be viewed as being composed of many different periodical components, such as the seasonal component with higher electricity usage in the summer and winter, weekly component with different weekend and weekday electricity usage, and daily component with a double peak pattern for residential customers and a single peak pattern for commercial customers. Because of these periodicities in electric load profiles, the frequency domain, or spectral analysis, is suitable for analyzing and modeling electric load data. Spectral analysis techniques have been applied in many different fields, such as understanding periodic patterns in economics, meteorology, geology, and astronomy. Spectral analysis is also applied in visual and audio signal data processing.

Using frequency domain characterization techniques, complex time-domain load profile data can be decomposed into a series of frequency domain components. These components can be used as a set of building blocks to model distinctive load pattern characteristics. These building blocks have the potential to serve as a load profile signature for each customer. Because of the modular and orthogonal nature of these components, statistics can be calculated for each component. The frequency domain statistics then can be used to construct electric load forecasting models.

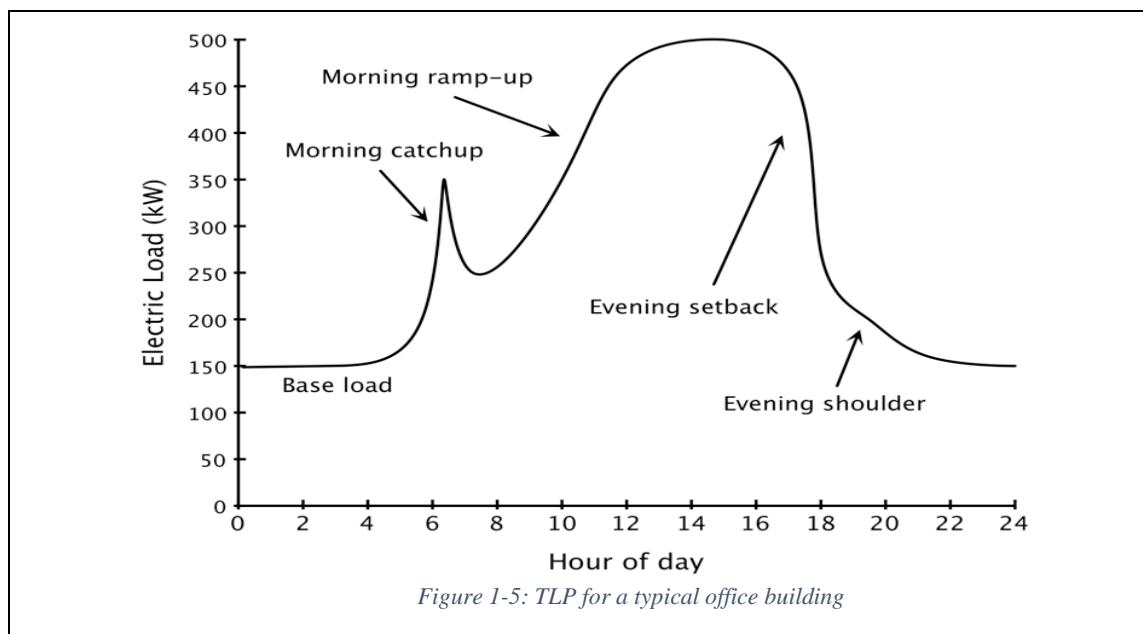
1.2 Challenges in Load Profile Modeling in the Frequency Domain

Load profile data is typically stored as time-series data that is indexed by the collection timestamp. The properties of this data set include:

- High data dimension.
- “Noisy” because of meter reading errors and missing measurements
- It contains various periodical data patterns
- It is correlated with geographical, demographical, climatological and sociological data.

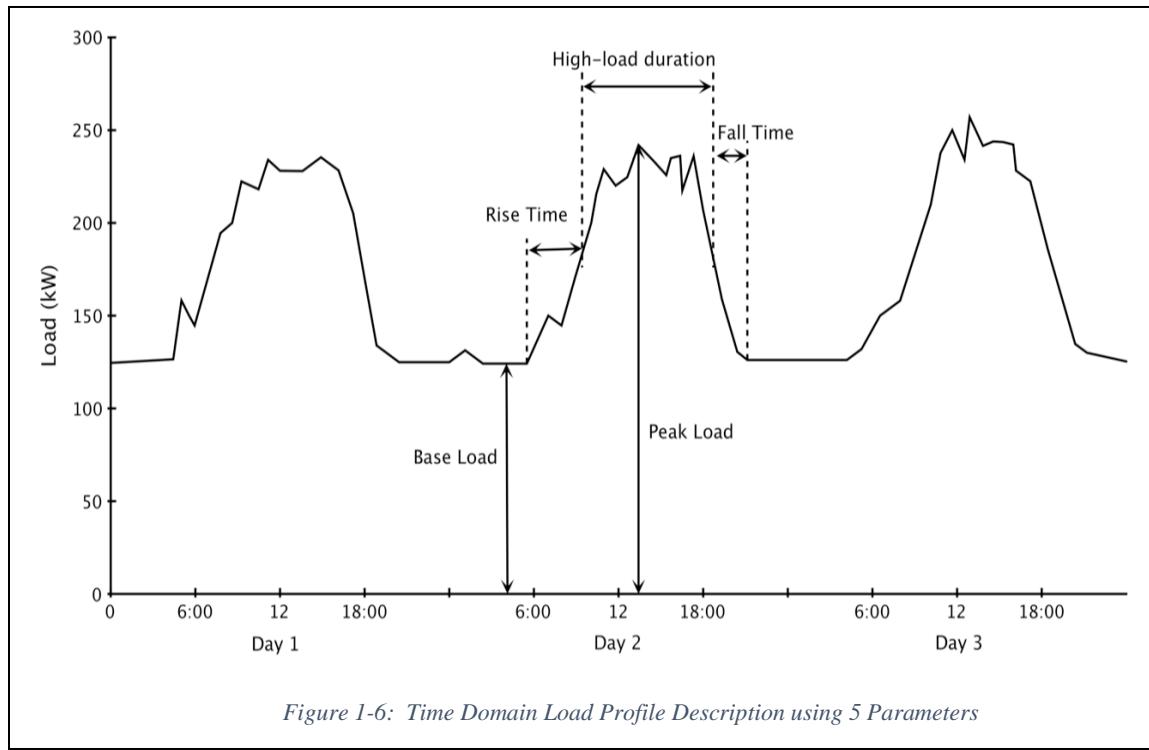
The above properties pose some challenges for modeling load patterns in the time-domain.

In the time domain it is difficult to systematically identify or quantify load pattern characteristics. In the time domain load profiles are typically modeled using a Typical Load Profile (TLP) within a time window (daily, monthly or annually), as shown in Figure 1-5 [1]. Recent research [1] has proposed to describe time domain load profile characteristics with parameters such as



- Base Load
- Peak Load
- Rise Time
- Fall Time
- High Load Duration.

The above parameters are illustrated in Figure 1-6 [2]. These parameters define the temporal relationship between the load profile magnitude values and the time indices. For a high-dimensional dataset, such as an annual load profile consisting of 8760 load measurements, it is very complicated to use such an approach to modeling the load.



Load estimation and load forecast models have been built for various types of customers [2-5] by extracting statistical properties from historical data. Some research refines such models by introducing other factors, such as weather conditions [6-7]. Various load models (regression [8], fuzzy approaches [9-11], and neural networks [12-13]) have

been used for load forecasting and estimation.

The performance of load models is dependent on a load profile classification process. This classification process should not be based on customer electric rate charges. Although customers are generally grouped by utilities into residential, commercial and industrial classes, and respective subclasses, there is no systematic framework that can be used to characterize usage patterns of different classes and subclasses with signatures that are both human-readable and machine-readable.

1.3 Research Objectives

This dissertation seeks to investigate and contribute to three major aspects of load research.

First, a theoretical framework for load profile representation in the frequency domain is sought. The predominant load profile representation today is in the time domain.

Second, a way to hierarchically classify loads based on significant usage patterns is sought. The predominant way that loads are classified today is based upon customer electric rate classifications.

Third, an efficient, scalable and parallel-processing approach to constructing frequency domain load models is sought.

1.4 Literature Review

A literature review of load research has been performed primarily in the following three areas

- Electricity load pattern characterization
- Electricity load pattern classification

- Electricity load estimation

Whereas this dissertation is dependent on the aforementioned fields, the research seeks to find a new framework to systematically model, classify and estimate electric load profiles. The following subsections highlight differences in approaches between the existing literature and the approach taken by this research.

1.4.1 Electricity Load Pattern Characterization

With the advent of AMI systems, high-resolution electric usage data is now being collected at individual customer and/or equipment levels. In order to successfully utilize this high resolution load data [1], the utility must have a systematic and efficient way to describe the unique and significant patterns and/or properties contained in the load data.

Conventional methods for analyzing load profile shape are detailed in reference [1] by Price. This reference discusses load profile characterization methods using various time-domain parameters (illustrated in Figure 1-5 and 1-6). Espinoza [2] introduced a unified framework to model the stationary properties of daily customer profiles. Currently utilities do not have a standard method or terminology to characterize and model customer load profiles. This is because the time domain load profile characterization methods are verbose and imprecise. They are not suitable for modeling the significant patterns contained in the load profiles.

Because consistent and predictable schedules exist in the majority of residential, commercial, and industrial customers, there are many “*hidden periodical patterns*” in load profiles. Preceval [14] and Stoica [15] both listed a set of spectral analysis applications which show that frequency domain analysis methods are good complementary to the traditional time domain methods. Carpneto [16] and Verdu [17] utilize the frequency

domain load pattern data to classify load profiles based on the values of the load profiles' frequency domain components.

This dissertation seeks to formalize the methodology for characterizing and analyzing load profiles in the frequency domain using the Discrete Fourier Transform. The dissertation addresses two issues, data sampling rate and signal bandwidth, which are not adequately addressed in the literature [16-17], but are important for the validity of frequency domain load profile characterization.

1.4.2 Electric Load Pattern classification

In order to effectively manage and plan for the electric power system load demand, it is important for a utility to be able to correctly and efficiently classify and categorize customers based on their electric usage patterns.

Conventionally load profile classification is performed in the time domain using various data mining techniques. Espinoza [2] introduces a time-domain, uniform framework for customer profile classification using the stationarity properties of the time-series Periodic Auto-regression (PAR) model to identify typical daily customer profiles. The basic PAR model template consists of 1176 parameters. This template is further extended to include exogenous variables to account for temperature effects, as well as monthly and weekly seasonal variations. Using the PAR model, the Typical Daily Profile (TDP) model can be computed. The load profile classification is accomplished by applying the unsupervised K-mean clustering algorithm on the TDP model.

In order to discover the unique customer load profile groups with significant load patterns, Ramos [18] applies four different data mining techniques - k-means, normalized N-Cut, Pairwise Constrained (PC k-means) and Metric Pairwise Constrained (MPC k-

means) - to a set of normalized average TDP models (which represent a set of annual load profiles). Once all possible typical load profiles (TDP) are discovered by the unsupervised clustering algorithm, then a supervised approach, supported by cluster validity indices as well as expert opinions, consolidate the large number of TDP into a set of more meaningful groups.

Kim [19] summarizes recent advances in Typical Load Profile generation for a customer class, which included the Self-Organizing Map (SOM), K-mean, Fuzzing C-means, hierarchical, Follow the Leader, and Fuzzy relations. The goal of these classification methods is to produce suitable TLP.

Additional literature [2, 4, 24-25] utilizes several variations of k-means, fuzzy-statistic, neural network and support vector machines [1-2,18-23] to classify load profiles using customer load profiles within different time intervals, such as annual load profiles or daily load profiles.

This dissertation seeks a hierarchical classification method based on the frequency domain characteristics of a load profile. The major difference between the frequency domain approach investigated here and the traditional time-domain classification methods is that the approach here depends on the new frequency domain load pattern characteristic parameters.

1.4.3 Electric Load Estimation

Load pattern characterization and classification focuses on how load profiles behave based on prior load profile data. These impact load estimation which in turn impacts utility financial, generation, transmission, distribution and integrated resource planning.

Load forecasting is important to cost-of-service allocation, rate design, demand response, and energy efficiency [1].

Load forecast models can generally be divided into two groups. The first group is the traditional time series and statistical methods. The major methods include [6-13]

- Time series - ARMA, Box-Jenkins ARIMA, regression, and transfer function (dynamic regression), expert systems; neural networks
- Fuzzy logic
- Support vector machine.

Almeshiae [24] describes a time-series method that uses load pattern decomposition to model the load profiles as two main components: the noise component and the smooth moving, average component. It utilizes the Self-Organizing Map (SOM) to identify segments in the load profiles that have similar behavior. The method combines the region similarity, contour and proposed related points to forecast the load profile.

Torkzadeh [25] proposes a method which combines the Principle Component Analysis (PCA) and Multi-Linear Regression (MLR) to forecast medium term load profiles. Espinoza [2] introduces Periodic Autoregressive (PAR) models to forecast the substation aggregated load profile. Chang [5] presents an updated version of the fuzzy load model to forecast transformer load profiles.

The second grouping for load forecasting are the machine learning methods. For example, Sevlian [26] describes three different load forecasting models: Seasonal ARMA models (SARMA), a Support Vector Regression (SVR) model, and a Feed Forward Neural Network (FFNN) model. The author indicates the performance indices of these models improve as the size of the load increases.

The majority of load modeling literature focuses on short term, from minutes to several days, and long term, from a year to a decade, load estimation at the system or substation level, where the load can range from several MW to several GW [24-26]. These load models are generally constructed using time-domain load data. However, in the time-domain it is difficult to describe and quantify the diverse and complicated load pattern characteristics that exist in AMI load data [27-28].

The aforementioned load forecast research is conducted in the time-domain. Yao [29] proposes a method that utilizes wavelet transformation to decompose the load profile and use neural networks to forecast each component's future value. The estimated wavelet component values are synthesized back to time domain load profiles.

This dissertation proposes to use frequency domain transformation techniques to decompose the load profile into frequency domain components. Frequency domain statistics can be computed for each significant frequency domain component. Conditional statistics will also be computed based on weather conditions. These frequency domain statistics can then be utilized in load estimation.

1.5 Dissertation Outline

Chapter II presents the frequency domain load profile characterization approach and the prerequisite conditions for its application. Analysis of daily load profiles using their frequency components is discussed.

Chapter III introduces the concept of frequency domain characteristic attributes. A frequency domain load profile classification method is presented. New frequency domain attributes are discussed and a general load profile classification procedure is presented.

Chapter 3 presents a load profile classification test that evaluates the efficacy of the frequency domain classification procedure.

Chapter IV introduces another frequency domain load profile modeling approach - the wavelet based AMI load profile model. Two different wavelet based load profile models are introduced. System load analysis results based on the wavelet models are presented.

Chapter V introduces a medium-term weather-dependent stochastic load model using the frequency domain statistics calculated from the AMI data set. The model performance metrics comparison with other models have also been presented.

Chapter VI summarizes the findings, contributions and future research topics.

Chapter 2: Load Profile in Frequency Domain

2.1 Methodology for Modeling Load Profiles in the Frequency Domain

Each daily load profile in Figure 2-1 consists of 24 hourly load data points $x_m(t)$, each of which is the monthly average for a group of commercial customers' load measurements recorded at that hour.

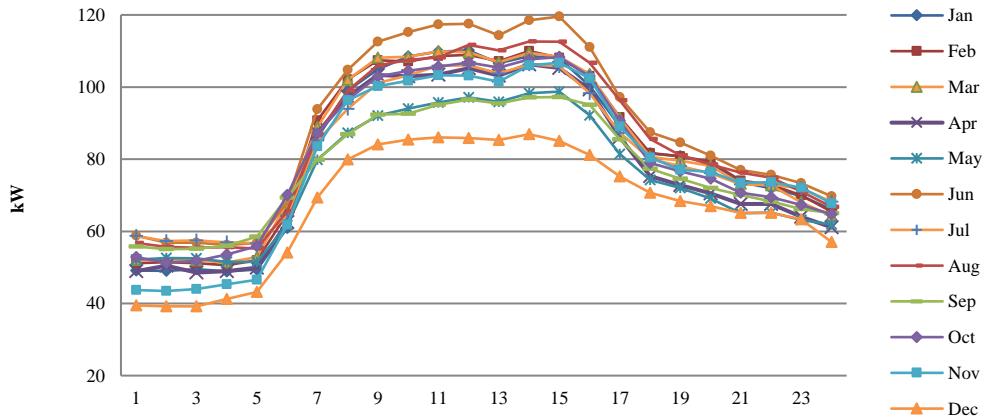


Figure 2-1: Average daily load profiles for utility A commercial customers

In order to focus on the shape of the load profile, the hourly data is normalized by the daily peak x_{peak} as in equation (2.1).

$$x_n(t) = x_m(t)/x_{peak} \quad (2.1)$$

The normalized daily load profile can be expressed by equation (2.2).

$$x(t) = \sum_{n=0}^{23} x_n \delta(t - n\Delta t) \quad (2.2)$$

Where x_n is the peak normalized hourly data indexed by n

Δt (1 hour) is the sampling time interval

The time-domain load profile can be transformed into the frequency domain by using the Discrete Fourier Transform (DFT). The frequency domain representation is shown by equation (2.3).

$$X(f) = \sum_{k=0}^{23} X_k \delta(f - k\Delta f) \quad (2.3)$$

Where X_k is a member of the frequency spectrum indexed by k

Δf is the frequency resolution.

Each member of the frequency spectrum, X_k , can be expressed by equation (2.4).

$$X_k = \sum_{n=0}^{23} x_n e^{-\frac{j2\pi kn}{24}}, \quad k = 0, 1, 2, \dots, 23 \quad (2.4)$$

Where X_k is a complex number that has magnitude a_k and phase angle θ_k .

Using the Inverse Discrete Fourier Transform (IDFT) shown by equation (2.5) and equation (2.6) the time-domain profile, $x^r(t)$, can be reconstructed by using some or all of the frequency components .

$$x^r(t) = \sum_{n=0}^{23} x_n^r \delta(t - n\Delta t) \quad (2.5)$$

$$x_n^r(t) = \frac{1}{24} \sum_{k=0}^{23} X_k e^{\frac{j2\pi kn}{24}}, \quad n = 0, 1, 2, \dots, 23 \quad (2.6)$$

Using Euler's identity to combine exponentials into cosine functions, $x^r(t)$ can be expressed as a sum of frequency components as shown in equation (2.7), which represents

the normalized load profile's shape information in term of magnitudes and phase angles of the harmonics.

$$x^r(t) = x_{DC}(t) + x_{1H}(t) + \cdots + x_{pH}(t) + \cdots + x_{11H}(t) \quad (2.7)$$

Where $x_{DC}(t) = a_{DC} = \frac{a_0+a_{12}}{24}$ is the DC component;

$x_{1H}(t) = \frac{a_1}{12} \cos\left(\frac{\pi n \Delta t}{12} + \theta_1\right)$ is the 1st harmonic with amplitude $a_{1H} = \frac{a_1}{12}$ and phase angle θ_1 ; $x_{pH}(t) = \frac{a_p}{12} \cos\left(\frac{p\pi n \Delta t}{12} + \theta_p\right)$ is the pth harmonic with amplitude $a_{pH} = \frac{a_p}{12}$ and phase angle θ_p .

Each harmonic is associated with an amount of energy consumed by activities performed at the corresponding harmonic frequency. For different customer classes, each harmonic can be interpreted with different physical meanings. For example, the commercial classes' 1st harmonic can be interpreted as representing their single-peak usage pattern from 8am-5pm (normal business operation hours). For residential classes, the 2nd harmonic can be interpreted as representing a typical family's daily household activity pattern, where the higher household activities in the morning and evening leads to higher electricity usage during those periods. The higher frequency harmonics may be correlated with other periodical factors that will be considered in what follows.

The original load profile can be fully reconstructed by using the product of the reconstructed normalized load profile and the peak of the profile as shown in equation (2.8).

$$x_m^r(t) = x_{peak} \times x^r(t) \quad (2.8)$$

Once the original load profile is transformed into the frequency domain, the profile can be described by the following description system:

$$S = k_{\alpha_{DC}} (\alpha_{1H}, \theta_{1H})_-(\alpha_{2H}, \theta_{2H})_- \dots_-(\alpha_{kH}, \theta_{kH}) \quad (2.9)$$

Where S is the shape descriptor

k is the number of frequency components used in the shape descriptor

α_{DC} represents the magnitude of the DC component

α_{kH} represents the magnitude of the k th harmonic normalized by x_{Peak}

θ_{kH} is the phase angle of the k th harmonic.

2.2 Prerequisites for Load Profile Data in the Frequency Domain

Before applying DFT to the time-domain hourly load data, two issues need to be considered. The first issue is whether the hourly data is adequate for representing the load profile. The second issue is whether the information represented in the hourly data is band limited, which can lead to aliasing problems. These issues will now be considered.

2.2.1 Data Adequacy

For individual residential or small commercial customers, the majority of load is consumed by various appliances and electronic devices that operate in duty cycles. The sampling rate commonly used by utilities (from 1 to 4 samples per hour) is too low to capture the true essence of each individual customer’s “true load curve”, which typically can be characterized as “needle peaks” due to the instantaneousness of the turn-on and turn-off processes. The hourly data (even 15-minute data) is not frequent enough to represent the load profile of individual residential or commercial customers [30, 31].

The diverse and non-coincident usage of electricity by a large group of customers (even with the same classification) has a smoothing effect on the volatility of individual load behavior. Consequently, hourly data can be frequent enough to represent the overall load profile for a group of 100 or more individual customers [31]. The hourly data shown in Figure 2-1 are obtained by taking the average of the hourly data from 119 commercial customers. Before the DFT is applied to time-series data, it is important to recognize this relationship between group size and sampling rate. If the DFT is to be applied to the load profile of an individual customer, measurements should be taken every minute [31]. In this paper, the load profile frequency domain characterization and analysis are performed only

on the aggregated customer class load profiles from three utilities, which will be referred to as utilities A, B & C.

2.2.2 Band-limitedness

Using load data from three different utilities, it will be demonstrated for the load data samples that the aggregated customer class load profiles studied in this paper is practically band-limited by showing that 1) the Nyquist frequency component (a_{12}) is negligibly small with 0 degree phase angle. This means that the signal is practically band-limited even though theoretically no signal can be both frequency and time limited (the daily load profile is time limited to 24 hourly measurements) [32], and 2) The time-domain profile can be reasonably well reconstructed by using only the frequency components below the Nyquist frequency.

Table 2-1 (page 32) presents the magnitudes and the phase angles for utility A commercial customer's monthly load profiles (normalized by its peak load value). It also presents the peak load, the load factor, and the load profile's shape descriptor string. Using the normalized magnitudes, the load factor is related to a_0 by equation (2.10).

$$\text{Load Factor} = \frac{\text{Average Load}}{\text{Peak Load}} = \frac{a_0}{24} \quad (2.10)$$

As shown in Table 2-1, the load factor is numerically very close to the DC component's magnitude value, a_{DC} , because the value of a_{12} is negligibly small. The DC component is expressed in equation 2.7 as the sum of a_0 and a_{12} normalized by the daily sample size, 24. a_0 is the average value of the normalized load profile. The shape descriptor string in table I follows the format introduced in equation 2.9 and contains DC and the

following 4 harmonic components' values. The descriptor string can be expanded to include more harmonics as needed.

Table 2-1: Harmonics for the Average Daily Load Profile for the Commercial Customers in Utility A

Month	Peak (kW)	Load Factor	Normalized Harmonic Magnitude a_{pH}											
			a_{DC}	a_{1H}	a_{2H}	a_{3H}	a_{4H}	a_{5H}	a_{6H}	a_{7H}	a_{8H}	a_{9H}	a_{10H}	a_{11H}
1	110.138	0.742	0.744	0.266	0.084	0.042	0.024	0.008	0.030	0.006	0.002	0.004	0.010	0.006
2	109.982	0.754	0.756	0.262	0.076	0.050	0.026	0.012	0.026	0.006	0.002	0.000	0.010	0.006
3	109.923	0.755	0.758	0.264	0.072	0.048	0.026	0.008	0.022	0.006	0.000	0.004	0.006	0.006
4	106.289	0.740	0.743	0.268	0.066	0.056	0.028	0.010	0.022	0.004	0.002	0.002	0.012	0.006
5	98.743	0.762	0.764	0.232	0.052	0.042	0.022	0.006	0.024	0.004	0.004	0.002	0.008	0.006
6	119.548	0.740	0.740	0.258	0.064	0.046	0.024	0.004	0.024	0.004	0.002	0.004	0.008	0.004
7	106.291	0.772	0.772	0.226	0.066	0.036	0.022	0.006	0.022	0.004	0.002	0.004	0.008	0.006
8	112.619	0.753	0.755	0.252	0.058	0.044	0.024	0.012	0.016	0.004	0.002	0.002	0.008	0.002
9	97.226	0.796	0.798	0.212	0.038	0.038	0.010	0.004	0.014	0.006	0.006	0.004	0.006	0.004
10	108.316	0.753	0.755	0.254	0.054	0.050	0.018	0.002	0.020	0.008	0.004	0.002	0.006	0.004
11	106.726	0.742	0.749	0.268	0.084	0.066	0.020	0.010	0.032	0.010	0.006	0.006	0.010	0.008
12	86.911	0.775	0.780	0.248	0.092	0.048	0.014	0.018	0.020	0.008	0.006	0.004	0.008	0.006
Month	Shape Descriptor (S)			Harmonic Phase Angle θ_{pH} (Radian)										
				θ_{1H}	θ_{2H}	θ_{3H}	θ_{4H}	θ_{5H}	θ_{6H}	θ_{7H}	θ_{8H}	θ_{9H}	θ_{10H}	θ_{11H}
1	4_0.74_(0.27, 3.1)_(0.08, 1.5)_(0.04, 1.0)_(0.02, 4.5)			3.070	1.480	0.960	4.480	3.030	2.510	2.020	2.690	3.670	2.800	2.550
2	4_0.76_(0.26, 3.1)_(0.08, 1.5)_(0.05, 1.0)_(0.03, 4.6)			3.100	1.570	1.020	4.610	3.580	2.450	2.360	4.750	4.270	3.090	2.260
3	4_0.76_(0.26, 3.1)_(0.07, 1.5)_(0.05, 1.0)_(0.03, 4.5)			3.130	1.540	0.990	4.460	4.050	2.470	2.840	2.600	0.920	3.050	2.660
4	4_0.74_(0.27, 3.2)_(0.07, 1.4)_(0.06, 1.0)_(0.03, 4.5)			3.150	1.400	1.040	4.530	3.840	2.480	2.390	0.240	4.630	3.140	2.390
5	4_0.76_(0.23, 3.1)_(0.05, 1.2)_(0.04, 1.0)_(0.02, 4.6)			3.120	1.210	0.990	4.640	3.660	2.650	2.150	1.180	4.460	2.880	2.490
6	4_0.74_(0.26, 3.1)_(0.06, 1.3)_(0.05, 0.9)_(0.02, 4.4)			3.100	1.290	0.910	4.430	3.710	2.570	1.230	0.790	4.900	3.020	1.970
7	4_0.77_(0.23, 3.1)_(0.07, 1.3)_(0.04, 1.0)_(0.02, 4.4)			3.100	1.200	0.970	4.370	3.110	2.570	1.400	0.370	4.910	3.120	1.940
8	4_0.76_(0.25, 3.1)_(0.06, 1.3)_(0.04, 1.0)_(0.02, 4.3)			3.050	1.290	0.960	4.310	3.620	2.280	1.810	1.150	5.100	2.940	2.570
9	4_0.80_(0.21, 3.1)_(0.04, 1.4)_(0.04, 1.2)_(0.01, 4.2)			3.090	1.410	1.150	4.160	4.890	2.300	3.040	1.930	1.440	3.550	2.860
10	4_0.76_(0.25, 3.1)_(0.05, 1.4)_(0.05, 1.2)_(0.02, 4.3)			3.140	1.430	1.150	4.270	4.860	2.390	2.690	1.420	0.310	3.130	3.050
11	4_0.75_(0.27, 3.1)_(0.08, 1.6)_(0.07, 1.2)_(0.02, 4.0)			3.000	1.590	1.210	3.990	3.050	2.490	2.310	2.600	2.780	3.210	2.770
12	4_0.78_(0.25, 3.0)_(0.09, 1.7)_(0.05, 1.3)_(0.01, 3.9)			2.970	1.700	1.280	3.860	2.890	2.630	2.390	3.080	3.010	3.260	3.090

With regard to the first point that if the load profile is band-limited, Table 2-1 has shown that as the frequency increases, the magnitude of higher frequency component decreases rapidly. The magnitude values decrease to a negligibly low value (<0.01) beyond the 6th harmonic, which clearly shows the characteristic of a band-limited dataset.

To validate the second point of the band-limited issue, this thesis will introduce the following four indices to evaluate how well the reconstructed load profiles represents the original time domain profile.

1) Peak Magnitude Error Index (PMEI) (in %)

This index is used to evaluate the difference between the value of the peak of the original profile and the reconstructed profile as a percentage of the former

$$PMEI = \frac{x_{o\max} - x_{R\max}}{x_{o\max}} \times 100 \% \quad (2.11)$$

Where $x_{o\max}$ is the peak of the original load profile $x_m(t)$

$x_{R\max}$ is the peak of the reconstructed load profile $x_m^r(t)$.

2) Maximum Magnitude Error (MME) (in %)

This index is used to evaluate the biggest percentage difference between the original and the reconstructed load profile.

$$MME = \text{maximum of } \left(\left| \frac{x_m(t) - x_m^r(t)}{x_m(t)} \right| \times 100 \% \right) \quad (2.12)$$

3) Mean Absolute Percentage Error (MAPE) (in %)

This index is used to evaluate the average difference between the original and the reconstructed load profile.

$$MAPE = \text{mean of} \left(\left| \frac{x_m(t) - x_m^r(t)}{x_m(t)} \right| \times 100 \% \right) \quad (2.13)$$

4) Peak Time Error (PTE) (in Hours)

This index is used to evaluate the time difference between the peaks of the original and the reconstructed load profiles.

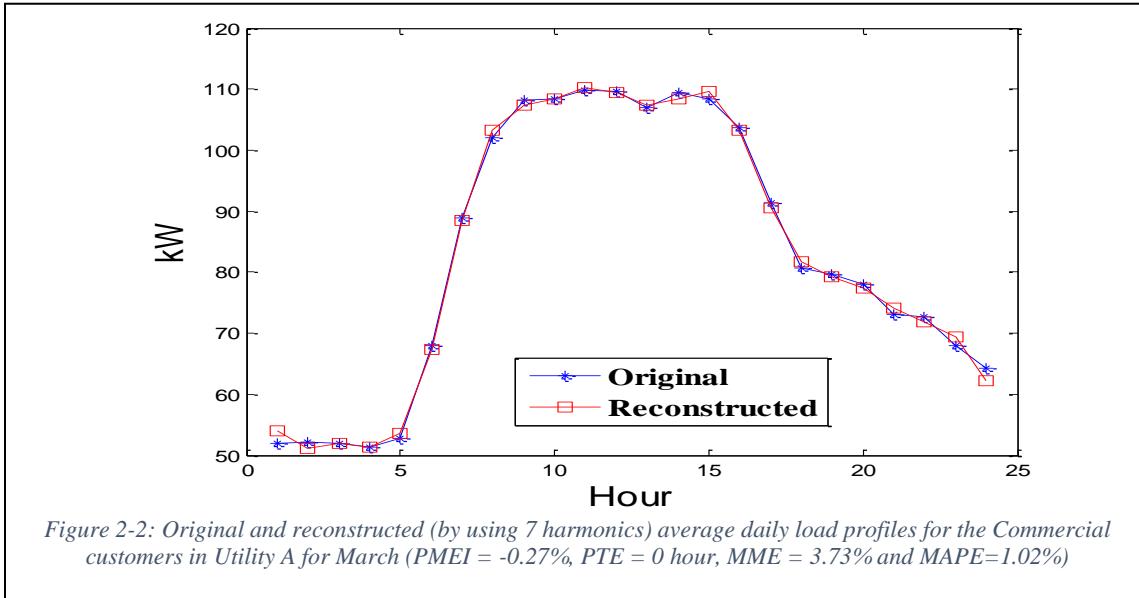
$$PTE = t_{o\max} - t_{R\max} \quad (2.14)$$

Where $t_{o\max}$ is the time at which $x_{o\max}$ occurs

$t_{R\max}$ is the time at which $x_{R\max}$ occurs.

The following procedure is used to handle load profiles with more than one peak. $T_{o\max}$ and $T_{R\max}$ are two sets of hours at which the loads are within $\pm 5\%$ of the absolute peaks of the original and the reconstructed load profiles, respectively. The intersection of these two sets is removed from each set. PTE is the element-by-element sum of differences of the two sets minus their intersection (the length of this new set is n). For example, for Figure 2-2, $T_{o\max} = \{11, 12, 15, 16\}$. For an intermediate reconstructed load profile, $T_{R\max} = \{10, 11, 16, 17\}$. The intersection of two sets are $\{11, 16\}$, which are omitted from the PTE calculation. $PTE = |12-10| + |15-17| = 4$. More harmonics are needed to reduce the PTE below 4.

Here a reconstructed load profile is considered satisfactory if the PMEI, MME, and MAPE are all less than 5% and the PTE is 2 hours or less.

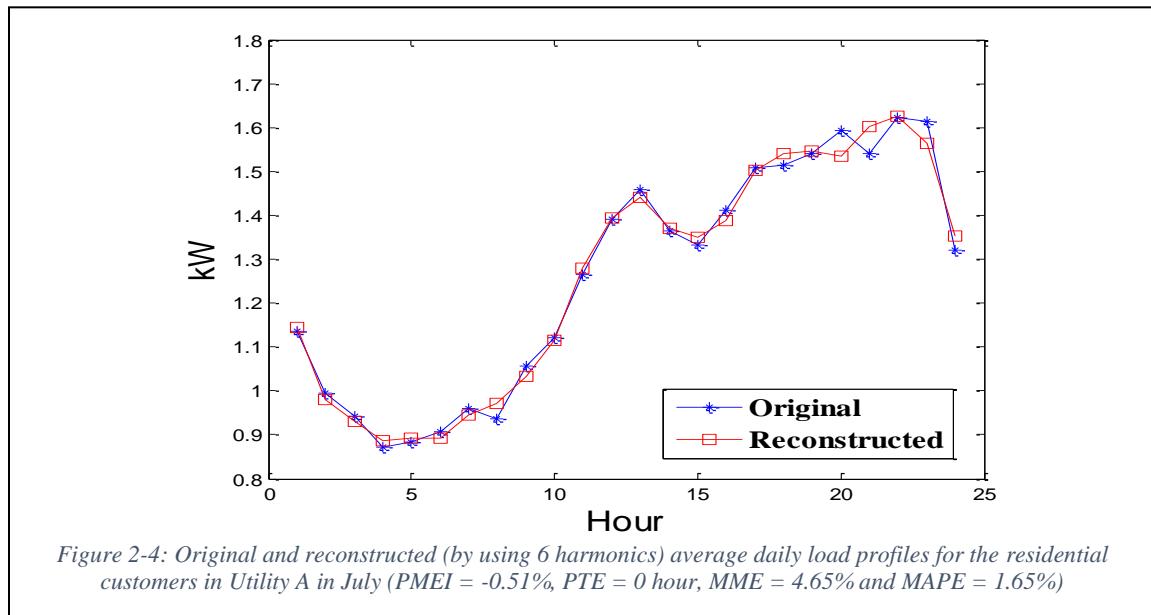
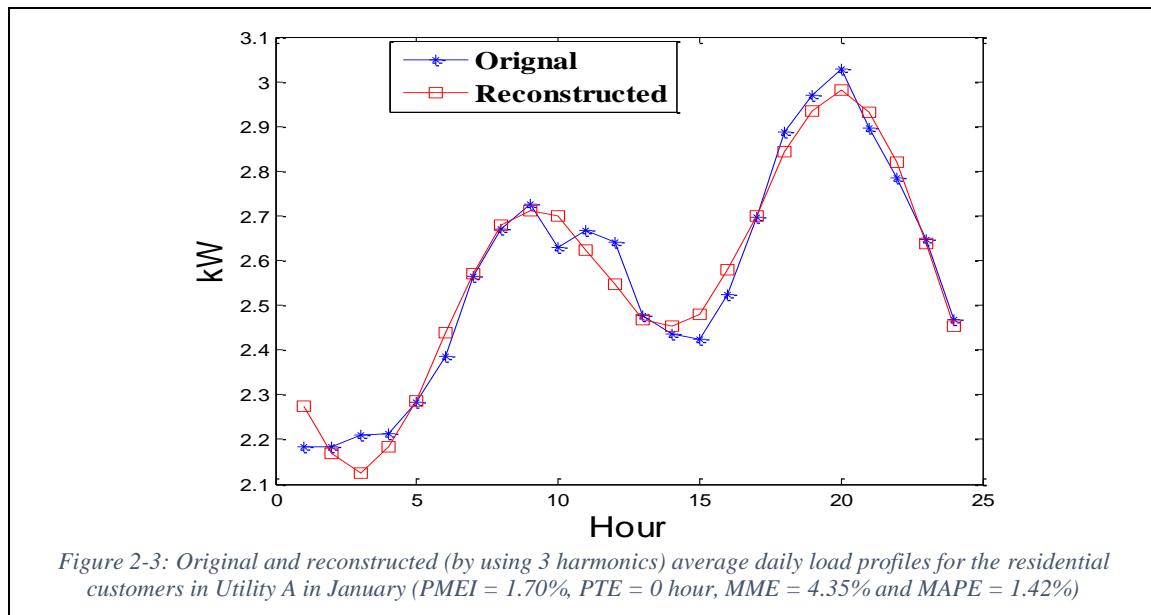


The January residential class load profile from utility A is used to demonstrate the IDFT reconstruction process. Table 2-2 presents the number of harmonics that are needed to achieve different levels of performance indices. For Table 2-1 load profile reconstruction, 3 harmonics will be sufficient to satisfy the aforementioned performance requirements.

Table 2-2: Number of Harmonics Needed to Achieve Different Levels of Performance Indices

Number of Harmonics Used	PMEI (%)	PTE (Hour)	MME (%)	MAPE (%)
1	15.27	4.00	17.56	8.08
2	8.43	2.00	13.14	6.84
3	1.70	0.00	4.35	1.42
4	0.60	0.00	3.83	1.11
5	0.65	0.00	4.28	1.09
6	1.55	0.00	2.94	0.90
7	0.94	0.00	2.09	0.72
8	0.78	0.00	1.80	0.70
9	0.45	0.00	0.88	0.50
10	0.34	0.00	0.90	0.49
11	0.65	0.00	0.67	0.38
12	0.15	0.00	0.20	0.17

Figures 2-3 to 2-4 are three original-vs.-reconstructed load profile overlay plots for commercial and residential customer classes in utility A for different months. The profiles reconstructed from the frequency domain representations are compact (ranging from 25% to 50% reduction in terms of harmonics used compared with the original forms). These plots illustrate that such compact representations are capable of representing the original profile without sacrificing major pattern characteristics.



Beside the examples from utility A, Table 2-3 shows the number of harmonics needed to reconstruct a load profile that satisfies the performance indices requirements for utility B. From this table, Utility B's customer class load profile frequency domain representations at most need 9 harmonics. These profile reconstruction examples, combined with the data presented in Tables 2-1 to 2-3, show that the commercial and residential customer class group load profiles in utilities A and B are band-limited, and therefore the DFT can be applied to these datasets.

Table 2-3: Number of Harmonics Needed to Reconstruct the Load Profile for Different Types of Customers in Utility B

Number of Harmonics Used	Percentages of Reconstructed Load Profiles with PMEI, MME, MAPE all < 5% and PTE <= 2 hours						
	Commercial		Multiple Dwelling	Public Building	Religious Entity	Residential	Mass Transit
	Large	Small					
1	4.2%	0.0%	20.8%	29.2%	0.0%	0.0%	0.0%
2	45.8%	12.5%	100.0%	95.8%	12.5%	8.3%	0.0%
3	95.8%	29.2%	100.0%	100.0%	29.2%	16.7%	0.0%
4	100%	50.0%	100.0%	100.0%	50.0%	33.3%	29.2%
5	100%	70.8%	100.0%	100.0%	70.8%	75.0%	37.5%
6	100%	79.2%	100.0%	100.0%	79.2%	87.5%	45.8%
7	100%	91.7%	100.0%	100.0%	91.7%	91.7%	91.8%
8	100%	91.7%	100.0%	100.0%	91.7%	100.0%	95.0%
9 or more	100%	100%	100.0%	100.0%	100.0%	100.0%	100.0%
Number of Customers	3,871	10,267	64	286	448	14,8107	8
Potential Storage Saving	67.0%	25.0%	92.0%	75.0%	33.0%	33.0%	25.0%

2.2.3 Benefits of Representing Load with Frequency Components

Figures 2-3 to 2-4 show that using a subset of frequency components to reconstruct the load profiles does not result in significant loss of accuracy. Table 2-2 and 2-3 demonstrate that applying similar analyses to the real-world data from utility A and B also shows that the higher frequency harmonics act as noise and have negligible effect on the load profile. It's possible to represent the load profile with a more compact frequency domain representation using the shape descriptor defined in equation 2.9, where this approach is illustrated in table 1 by the shape descriptor column. The frequency domain load profile representation can be more robust and more resistant to error/outliers than the original time-domain data.

The shape descriptor has the flexibility not only to represent the daily load profile, but also the yearly load profile of a customer class. It just needs to incorporate the necessary frequency components from the yearly load profile's frequency domain representation using the procedure introduced earlier in this chapter. This approach has the advantage of reducing the load profile data dimensions by only incorporating the major frequency components with significant magnitude values, thereby realizing a substantial savings in the size of the load data model (Table 2-3). It also has the flexibility to satisfy the need for additional accuracy by including more frequency components.

Each shape descriptor describes the signature of a customer class's load profile and is suited for automated machine processing. For example, the hamming-distance between a load profiles' descriptor strings can be used to quantify the level of differences between different load profiles, and naturally can be adopted in load profile classification. Each component in the descriptor can be independently analyzed, which can be implemented as

sets of parallel processes for more efficient analysis routines.

“A prerequisite to developing an accurate load-forecasting model is an in-depth understanding of the characteristics of the load to be modeled” [33]. The proposed method intends to provide an alternative view of the characteristics of load in terms of frequency components. One possible use of this approach is to use the frequency components from the past to forecast frequency components in the future, and then recombine them using IDFT. The robustness of significant harmonics and the orthogonal relationship between is beneficial to constructing effective load models.

2.3 Load Profile Statistics in Frequency Domain

The dataset used in this section comes from 3 utilities. It ranges from year 2002 to year 2007, and contains almost 1.5 million daily records (24 hourly measurements). The work here does not attempt to re-cluster the sample customers. That is, the customer class grouping used have been defined by their respective utilities' load research / customer information departments.

Because of their significance in IDFT reconstruction process, as shown in previous chapter, the analysis in this section focuses on the first 3 frequency components of the normalized daily load profiles.

Tables 2-4 and 2-5 present the DC, 1st and 2nd harmonic statistics, which are used to evaluate the certainty and consistency of each component's magnitude and phase. The frequency components' phase angle statistics presented in Table 2-4 to 2-6 and Figure 2-6 are circular statistics [34]. Normal statistics are not suitable for analyzing phase angle data sets. For example, the average of 5° and 355° is not 180° since both angles point approximately in the same direction, and the designation of the high and low values in phase angle is arbitrary.

Table 2-4: Statistics for the Magnitudes and Phase Angles of the First Three Frequency Components for the Residential Customers of Utility B and Utility C

Utility	Month	Mean						Coefficient of Variation (%)				Standard Deviation	
		Peak (kW)	a_{DC}	a_{IH}	a_{2H}	θ_{IH}	θ_{2H}	Peak	a_{DC}	a_{IH}	a_{2H}	$\sigma_{\theta IH}$	$\sigma_{\theta 2H}$
B	Jan	1.49	0.64	0.18	0.16	1.7	2.29	15%	7%	19%	17%	0.21	0.19
	Feb	1.36	0.64	0.16	0.17	1.63	2.24	16%	6%	15%	13%	0.19	0.2
	Mar	1.37	0.62	0.17	0.17	1.6	2.23	16%	7%	16%	11%	0.19	0.18
	Apr	1.28	0.64	0.16	0.16	1.61	1.99	19%	7%	24%	18%	0.22	0.25
	May	1.2	0.64	0.17	0.15	1.71	2	18%	9%	22%	15%	0.27	0.29
	Jun	1.61	0.67	0.21	0.11	1.75	1.93	32%	9%	27%	35%	0.22	0.39
	Jul	1.99	0.7	0.21	0.07	1.67	1.88	26%	7%	29%	46%	0.31	0.63
	Aug	2.04	0.71	0.2	0.07	1.7	1.78	28%	8%	31%	44%	0.28	0.62
	Sep	1.56	0.67	0.19	0.12	1.71	2.03	21%	9%	25%	30%	0.25	0.33
	Oct	1.6	0.61	0.18	0.16	1.77	2.2	31%	13%	20%	21%	0.23	0.26
	Nov	1.58	0.62	0.22	0.16	1.79	2.23	16%	9%	15%	19%	0.23	0.26
	Dec	1.78	0.61	0.21	0.17	1.77	2.24	18%	8%	14%	20%	0.21	0.24
C	Jan	4.1	0.77	0.07	0.14	2.33	2.66	9%	4%	58%	19%	0.61	0.18
	Feb	3.62	0.77	0.07	0.15	2.19	2.55	11%	3%	47%	12%	0.49	0.18
	Mar	3.23	0.74	0.06	0.16	2.79	2.35	13%	5%	60%	13%	1.03	0.16
	Apr	2.64	0.71	0.11	0.15	2.11	2.25	13%	6%	62%	14%	0.74	0.15
	May	2.55	0.68	0.15	0.15	1.91	2.07	15%	6%	36%	12%	0.19	0.12
	Jun	3.04	0.72	0.18	0.14	2.06	1.93	16%	7%	24%	15%	0.2	0.19
	Jul	2.02	0.76	0.21	0.06	2.14	1.52	9%	5%	15%	31%	0.14	0.4
	Aug	3.53	0.74	0.22	0.1	2.1	1.9	15%	4%	24%	16%	0.12	0.17
	Sep	2.65	0.7	0.16	0.14	2.04	2.08	10%	4%	18%	10%	0.1	0.08
	Oct	2.29	0.71	0.12	0.14	1.91	2.13	5%	4%	24%	8%	0.06	0.1
	Nov	2.5	0.71	0.13	0.16	1.72	2.51	7%	4%	28%	13%	0.14	0.13
	Dec	3.26	0.76	0.11	0.14	1.65	2.48	12%	6%	35%	18%	0.3	0.16

Table 2-5: Statistics for the Magnitudes and Phase of the First Three Frequency Components for the Large Commercial Customers of Utility B and Utility C

Utility	Month	Mean						Coefficient of Variation				Standard Deviation	
		Peak (kW)	a_{DC}	a_{IH}	a_{2H}	θ_{IH}	θ_{2H}	Peak	a_{DC}	a_{IH}	a_{2H}	$\sigma_{\theta IH}$	$\sigma_{\theta 2H}$
B	Jan	1482.5	0.84	0.16	0.03	2.84	1.16	13%	6%	32%	41%	0.14	0.65
	Feb	1452.17	0.84	0.16	0.03	2.87	1.08	10%	5%	29%	34%	0.14	0.54
	Mar	1450.2	0.84	0.17	0.03	2.86	1.02	9%	5%	24%	40%	0.09	0.51
	Apr	1488.27	0.83	0.17	0.02	2.81	0.9	13%	4%	21%	41%	0.08	0.52
	May	1579.53	0.82	0.19	0.02	2.79	0.79	12%	5%	23%	41%	0.11	0.5
	Jun	1797.48	0.81	0.2	0.02	2.82	1.03	10%	4%	16%	39%	0.1	0.61
	Jul	1768.9	0.83	0.18	0.02	2.83	1.32	8%	2%	12%	36%	0.12	0.59
	Aug	1876.85	0.82	0.19	0.02	2.84	1.27	9%	3%	16%	71%	0.25	0.55
	Sep	1712.99	0.81	0.2	0.02	2.8	1.09	11%	4%	19%	36%	0.11	0.55
	Oct	1507.95	0.83	0.18	0.02	2.83	1.03	10%	4%	19%	41%	0.1	0.49
	Nov	1402.67	0.85	0.15	0.02	2.85	1.31	12%	6%	33%	44%	0.31	0.75
	Dec	1399.73	0.86	0.14	0.03	2.84	1.41	10%	5%	30%	44%	0.25	0.54
C	Jan	568.01	0.84	0.18	0.05	2.73	1.73	9%	4%	7%	13%	0.05	0.08
	Feb	501.73	0.83	0.18	0.05	2.8	1.61	11%	3%	6%	16%	0.04	0.13
	Mar	547.7	0.82	0.2	0.05	2.74	1.62	14%	5%	5%	14%	0.04	0.11
	Apr	522.53	0.8	0.22	0.05	2.67	1.65	13%	6%	8%	22%	0.05	0.09
	May	533.1	0.79	0.23	0.04	2.65	1.56	15%	6%	7%	17%	0.08	0.39
	Jun	591.8	0.79	0.23	0.04	2.68	1.67	19%	7%	6%	17%	0.07	0.19
	Jul	837.25	0.8	0.23	0.04	2.62	1.76	10%	5%	7%	20%	0.09	0.26
	Aug	699.58	0.79	0.23	0.04	2.7	1.74	16%	4%	7%	17%	0.07	0.22
	Sep	630.49	0.78	0.24	0.04	2.68	1.56	10%	4%	7%	14%	0.07	0.14
	Oct	548.12	0.8	0.22	0.05	2.71	1.61	5%	4%	6%	15%	0.05	0.08
	Nov	540.34	0.81	0.21	0.05	2.73	1.71	8%	4%	9%	15%	0.06	0.12
	Dec	533.66	0.84	0.19	0.05	2.74	1.67	12%	6%	11%	29%	0.11	0.28

The followings are the major findings through the statistical analysis of the first 3 frequency components

2.3.1 Consistent Harmonic Patterns

The residential and large commercial classes' frequency domain statistics in Table 2-4 and 2-5 show that the relative contribution from the 3 frequency components to the overall load profile displays a consistent pattern for the entire year for both utility B and C. For example, the DC magnitude for both utility A and B residential customers is around 0.7; the 1st and 2nd harmonic magnitudes are around 0.2; and the 1st and 2nd harmonic phase angles are around 2 radians.

Such consistency is summarized in Table 2-6 using the mean and the range of the frequency components' value. For example, Utility B residential customer's a_{DC} ranges from 0.61 to 0.71 with mean equal to 0.66. The narrow ranges for the frequency component values in table V indicates a high consistency.

Using the aforementioned descriptor system in equation 2.9 ($S = k_{\alpha} \alpha_{DC} - (\alpha_{1H}, \theta_{1H}) - (\alpha_{2H}, \theta_{2H}) - \dots - (\alpha_{kH}, \theta_{kH})$). Utility B residential classes' first 3 frequency components typical pattern can be represented using the mean of the magnitude and the phase angles listed in table V as: $S = 2_0.7_0.2, 2_0.1, 2$, while its commercial

Table 2-6: Harmonic Pattern for Different Customer Classes

	Residential		Commercial	
	Utility B	Utility C	Utility B	Utility C
a_{DC}	0.66±0.05	0.73±0.05	0.84±0.03	0.81±0.03
a_{1H}	0.19±0.03	0.14±0.08	0.17±0.03	0.21±0.03
θ_{1H}	1.70±0.10	2.22±0.57	2.83±0.04	2.71±0.09
a_{2H}	0.12±0.05	0.11±0.05	0.03±0.01	0.05±0.01
θ_{2H}	2.04±0.26	2.09±0.57	1.10±0.31	1.66±0.10

class can be represented as: $S = 2_0.8_(0.2, 3)_(0.03, 1)$.

Commercial customers' frequency component values generally have a narrower range than their residential customer counterparts, which makes their load profile patterns more consistent than residential customer profiles.

2.3.2 *High Degree of Certainty*

The coefficient of variation COV, is used to evaluate how well the frequency domain approach can extract useful information out of data polluted by noise. COV can be expressed as equation (2.15):

$$COV = \text{standard deviation}/\text{mean} \times 100\% \quad (2.15)$$

Customer load profile data collected by meters will be unavoidably tainted by various noise/errors. In the time domain, it is difficult to separate such noise/error influences on the COV, which is high across all time points for two different Utility B customer class daily load profiles shown in Figure 2-5.

In the frequency domain, it has been shown in Table 2-4 and 2-5 that the majority of the DC component COVs are less than 10% for both types of customer in two utilities. COVs for the 1st and 2nd harmonics are larger, but since the magnitudes of these components are smaller than DC, the impact of these increased levels of uncertainty is correspondingly smaller. The COVs for the first 3 frequency component magnitudes are smaller than that of the time-domain data, indicating a higher degree of certainty of the information.

The frequency domain statistics spread the statistical uncertainty across different

frequency components. The frequency components with significant magnitude values tend to have a lower COV, combined with the fact that the load profiles can be represented with the small set of frequency components with significant magnitude values. Due to this, it is easier to extract load profile characteristics information from such statistic datasets.

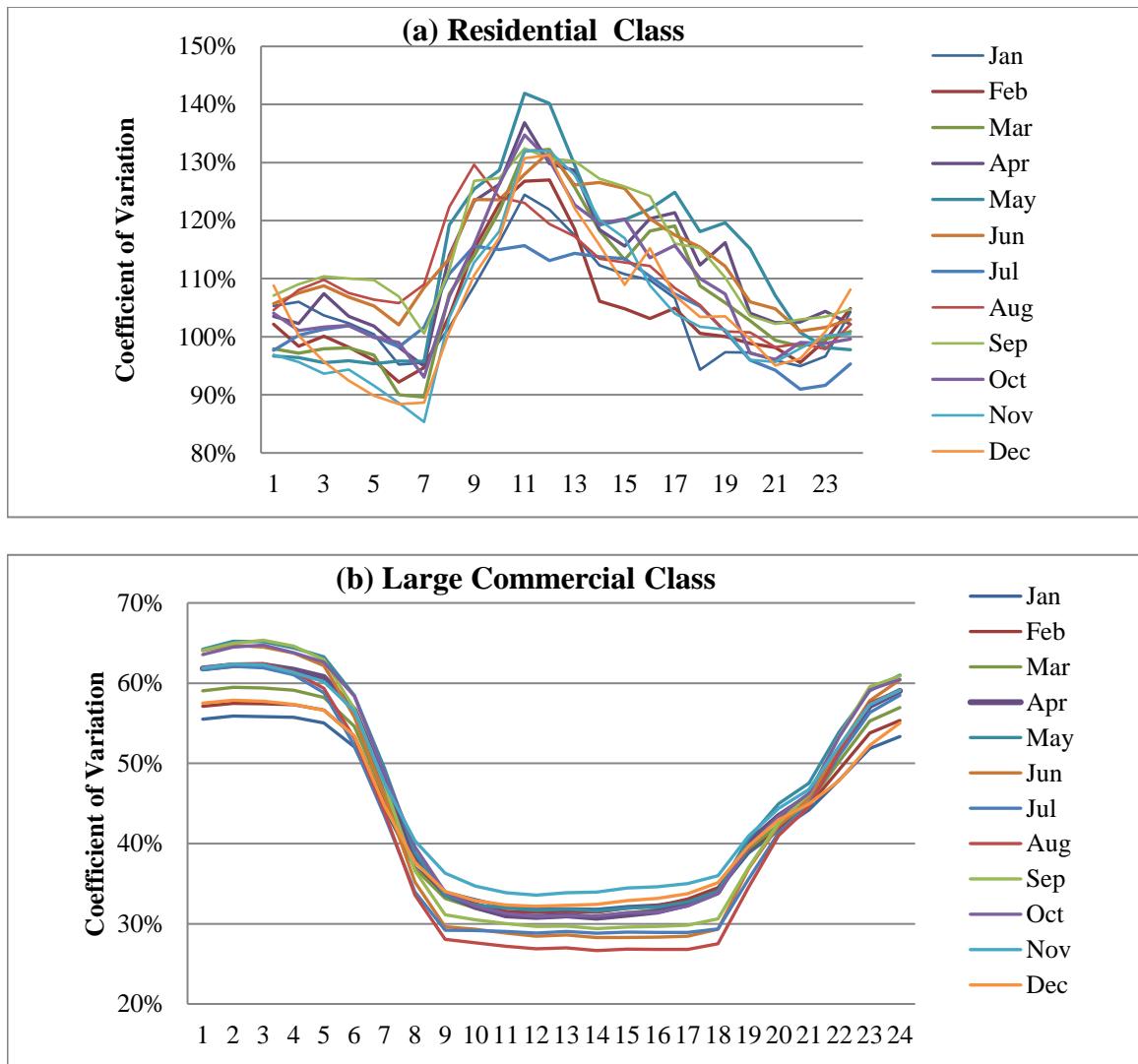


Figure 2-5: Plots of coefficient of variation for the average daily load profiles for two classes of customers of Utility B

2.3.3 Similar Patterns for Historical Data at Different Utilities

In Figure 2-6 the first column shows the multi-year (05-07 for utility A and 02-05 for utility B) magnitudes and phase comparison charts for the large commercial class load profiles' first three frequency components. The second and third columns show the same type of historical data for the residential classes from the same utilities during the same time period. Each individual curve represents 12 monthly frequency components' magnitude/phase angle values.

The charts in Figure 2-6 show that 1) From year to year the harmonics tend to have a very similar overall trend in terms of both magnitudes and phase angles, and therefore the predictability for such frequency components is high. 2) There exist certain seasonal patterns in the harmonic components. For example, the residential customer class's 1st harmonics magnitude has a higher value than the 2nd harmonic magnitude during the summer (June to September); otherwise, these two harmonic magnitudes are much alike during the other seasons.

During the summer season the residential class's 2nd harmonic phase angles are smaller than the 1st harmonic phase angles. The 2nd harmonic phase angle tends to be consistent throughout the year, while 1st harmonic phase angle tends to rise up during the winter (December to March). The consistency of the 2nd harmonic's phase angle seems to correlate with the consistency of the typical household activity (residential customer consistently has a higher usage during the morning and evening hours).

These observations indicate the potential to use the historical load profile frequency domain characterization statistics to forecast the future load profile's frequency domain

representation, and to reconstruct its time domain load profile using the forecasted frequency components.

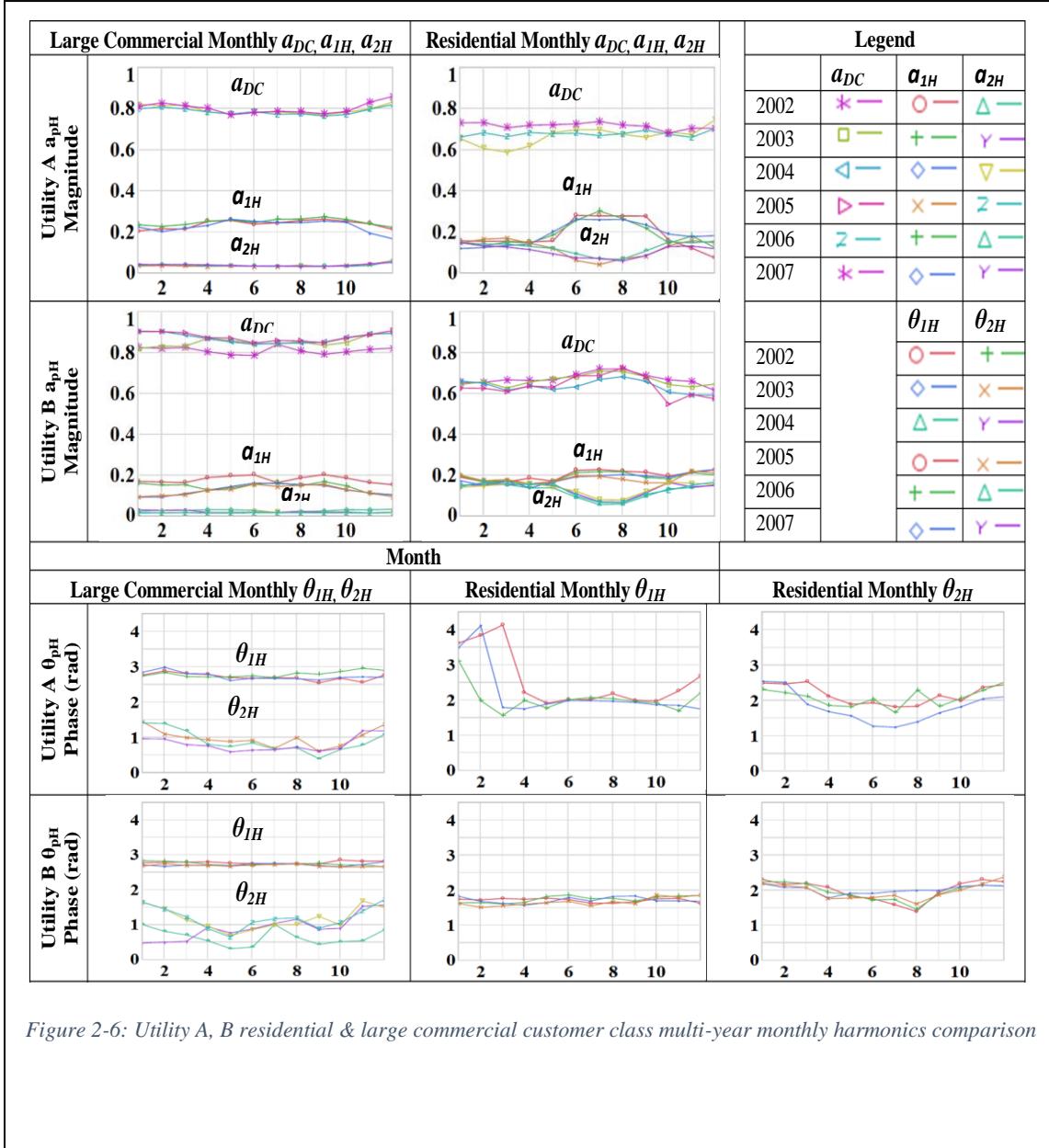


Figure 2-6: Utility A, B residential & large commercial customer class multi-year monthly harmonics comparison

2.4 Summary

This chapter has accomplished the following: First, it has formalized the methodology for characterizing and analyzing load profiles in the frequency domain using the DFT. It has shown that the customer class load profile data satisfies two important but often overlooked prerequisite conditions for such applications: 1) the hourly sampling rate is adequate to represent the aggregated customer class group load profile; 2) the group load profile is band-limited. Using real customer load data from three different utilities, it has been demonstrated that the time domain load profile data can be transformed to the frequency domain as a set of independent frequency components.

Second, this chapter has demonstrated that the different customer classes' load profiles can be represented with a small set of frequency components (the highest magnitude set) without sacrificing major load profile characteristics. Section II showed that the customer class load profiles from 3 utilities can be accurately reconstructed using only frequency components while satisfying the requirements specified.

Third, this chapter proposed that load profiles can be described using the shape descriptor presented in section II. The descriptor is compact (includes only a small subset of all frequency components) and flexible (can incorporate more or less components as needed). Each component in the string can be independently analyzed. This shape descriptor string is well suited for automated machine processing and analysis.

And last, this chapter presents several findings from analyzing the frequency domain statistics from monthly customer class load profiles for three utilities, which include: 1) consistent frequency component patterns in both residential and commercial customer classes for different years. Such evidential predictability of frequency

components of group load profile can potentially be used in load forecasting in the frequency domain; 2) low uncertainty among frequency components with high magnitude; 3) similar customer classes from different utilities possesses similar annual frequency domain component patterns.

In this chapter the load profile data for different customer classes do not represent any typical load profile (TLP). Although the proposed approach can be applied to characterize and analyze TLPs after they have been developed, further research is needed to determine how the proposed approach can be applied in the development of TLPs.

Chapter 3: Hierarchical Classification of Load Profiles Based on Their Characteristic Attributes in Frequency Domain

Utilities generally group their customers into residential, commercial and industrial classes, and respective subclasses. There is no systematic framework that is used by utilities to automatically characterize different classes and subclasses based on their load profile pattern characteristics. The work presented in this chapter attempts to formulate the theoretical framework for customer classification based on their frequency domain characteristic attributes.

The goal of load profile classification is to find a model for predicting values of customer class variables from predictor variables. Conventionally, the predictor variables in load profile classifications are based on the load profile's time domain representation. This is often done using a Typical Load Profile (TLP) within a time window (daily, monthly...). Recent research suggests describing the load profile characteristics using time domain parameters, such as Base Load, Peak Load, Rise Time, Fall Time, and High-Load Duration [1].

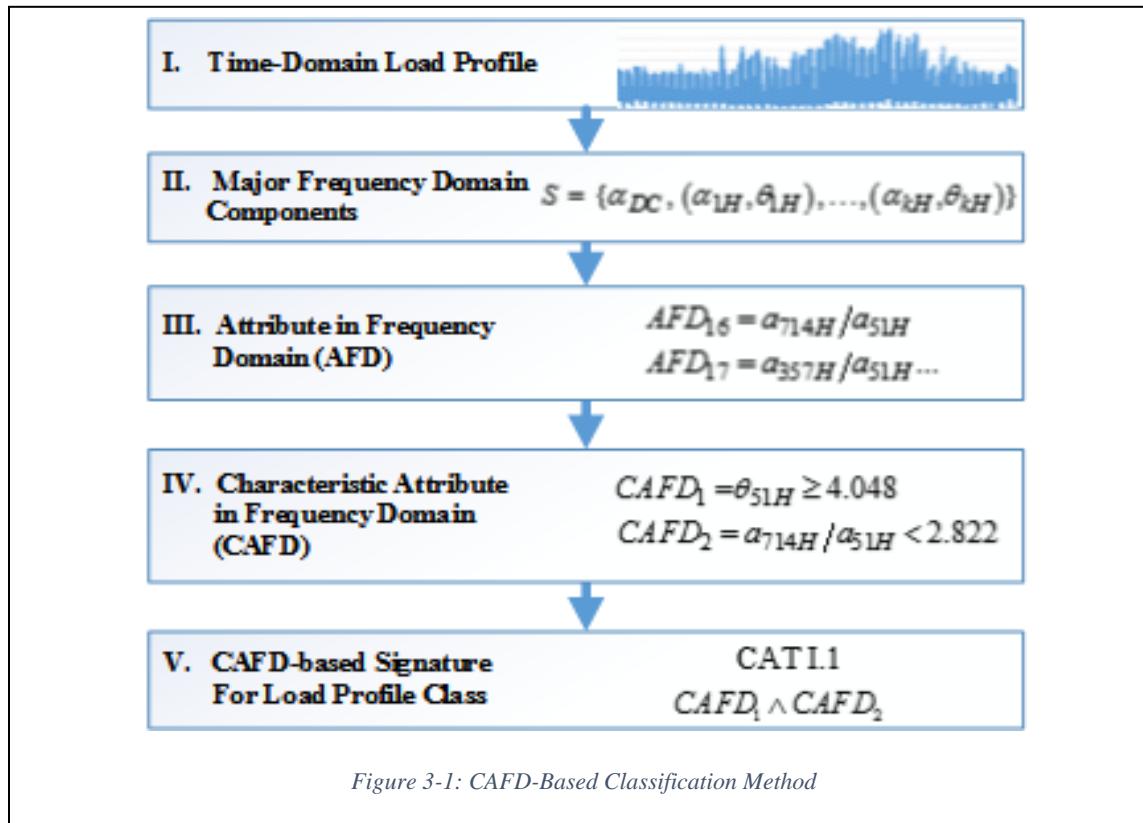
After establishing the time domain predictor variables, various models, such as linear discriminant analysis, nearest neighbor classification, k-means, fuzzy-statistic, neural network and support vector machines [1-2,18-23], are used to classify load profiles.

In Figure 1-4, there are two normalized annual hourly load profile samples (residential and commercial) and two magnified portions of the residential profile (one in

Feb, one in July). It is difficult to use aforementioned classification methods to classify such large, complex, load behavior, which requires a large number of predictor variables to model.

In order to classify load profiles based on their annual load profile pattern characteristics, a hierarchical classification method based on the load profiles' Characteristic Attributes in the Frequency Domain (CAFD), is investigated here. Using the annual load data as the starting point, the proposed method can identify the CAFD-based signatures for different classes and subclasses of load profiles by using the steps outlined in Figure 3-1.

While the work presented in this chapter is built upon the work discussed in chapter 2 [10], the two chapters are significantly different. The work in chapter 2 shows the benefits of using frequency domain descriptors to characterize and analyze load profiles. However,



different customer classes can be characterized by frequency domain descriptors that have the same members, but some or all of the members exist in different proportions. The work presented in this chapter demonstrates how to extract a load profile's CAFDs from the descriptors, and how to use these CAFDs to formulate a hierarchy of load profiles that can be used as a systematic framework for customer load classification. As illustrated shortly, some of the CAFDs are derived from the descriptors, but are not descriptors. As signatures for customer classes and subclasses, the CAFDs are obtained by using a data mining method called CART (Classification and Regression Tree) [35]. The input predictor variables are obtained from the analyses of the frequency domain descriptors.

While the work presented in chapter 2 is based on daily load data, the work presented in this chapter is based on annual load profiles, which is more difficult to characterize. The work presented in chapter 2 represents Step I to Step II in Figure 3-1, while the work presented in this chapter covers from Step II to Step V. The results presented in chapter 2 do not demonstrate any hierarchical structure, but the work presented in this chapter demonstrates hierarchy in both the method (Section III.2) and the results (Section III.3).

The proposed classification framework could serve as the foundation of a standard and universal load classification system. This hierarchical classification method using the concepts of CAFD and CART. And such method can classify all the load profiles to its respective place in the CAFD-based classification hierarchy. The research result will demonstrate that the CAFD-based signatures can be used as the definition for different load profile classes and subclasses that correspond well with real-life load profiles.

3.1 Hierarchical Classification Tree

The classification tool used in this work is based on the CART concept developed by Breiman, et al [36]. There are many applications for CART in a lot of different fields. For example, some researchers used it to classify the housing prices using environmental/social factors (Boston Housing data [37]). In the medical field, researchers use it to classify whether a disease has progressed based on various factors [38].

For classification, CART constructs a categorical predication model from a finite number of unordered input variables. CART approximates the parameter space by a piecewise constant function. CART will first split the entire data set using a predictor variable condition that produces the smallest impurity score. This recursive partitioning process (the next split only happens in the parent's data space) will continue until the leaf nodes contain the minimum number of samples or are all homogeneous.

Each partition step must be performed in a hierarchical order and must apply locally within the boundary of a previous partitioned space, as a recursive greedy algorithm should, so that the partitioned spaces are disjointed, with no overlap. The order of condition in the hierarchical classification rules indicate their significant ranking in the classification process: the condition used in the root, node 1, has the most significant impact on classifying the data.

The final number of leaf nodes (classes) can be controlled by experimenting with different values of the complexity parameter (CP, a tuning parameter: 0 means full model and ∞ means no split). When CP=0, which has the least impurity requirement for each split, this setting will produce a tree with the largest number of leaf nodes with homogenous characteristic attributes. By experimenting with a larger CP value (less restriction on the

purity of each split), the classification tree can be pruned back to have fewer, less homogenous leaf nodes/classes. The statistical package, *rpart* [38] was used in this work to construct the classification tree and extract the classification rules for the annual load profiles.

3.2 Load Profile Characteristic Attributes in the Frequency Domain

653 annual hourly load profile data (recorded at different years for different types of customers) were collected from 5 utilities' websites [39-43]. In order to capture the load profile periodical pattern characteristics, all annual hourly load profiles used in this paper need to be placed in a fixed time window, which is defined by the day of the week (consistent and periodical) for 51 consecutive weeks. The first time point in the 51-week window is set to 12 am of the 1st Monday of the year. The last time point is set to 11 pm of the 51st Sunday of the year.

After the DFT transformation of the time-domain load profiles (stage I), the major frequency domain components are identified to form the Frequency Domain Load Profile Descriptor [10] (S at stage II), which is introduced in chapter I..

$$S = \{\alpha_{DC}, (\alpha_{1H}, \theta_{1H}), \dots, (\alpha_{kH}, \theta_{kH})\} \quad (3.1)$$

where α_{dc} and α_{kH} represent the magnitudes of DC and the k th harmonic respectively

θ_{kH} is the phase angle of the k th harmonics

Among the many characteristic attributes in the annual hourly load profiles, the daily, weekly and seasonal periodical pattern attributes are the primary focus of the classification process used here. Such load profile characteristics are naturally related to

some of the modular and orthogonal frequency domain components illustrated in Table 3-1. The characteristics and their related frequency components normally have the similar period. The DC component represents the average of the whole profile, which is similar to load factor in value [27]. Even though the DC component does not relate to any periodical pattern in the time domain, its value does impact every harmonic magnitude value because it normally has the largest magnitude value among all load profile harmonics.

Table 3-1: Common Load profile patterns

Harmonic Index	Period (Hours)	Pattern Type
DC		Constant
357	24	Daily Pattern
714	12	Twice-a-Day Pattern
51	168	Weekly Pattern
1	8568	Annual Pattern
2	4284	Twice-a-Year Pattern
3	2856	Thrice-a-Year Pattern

Not all major frequency components in S can serve as predictor variables in a classification tree because each S' member only represent one periodical component in the load profile. Individually, each member will not be able to represent some complex load profile characteristic attributes of interest. For example, the periodical load drop between weekdays and weekends, will require combination of a number of frequency components in S to model.

The shortcomings of S lead to the development of a hierarchical load profile classification method. The major steps of this method are presented in Figure 3-1. Based on preliminary analysis results, the following major harmonics are proposed to be used as the **Attributes in the Frequency Domain, AFD**,

$$AFD_{1-15} = \{\alpha_{DC}, \alpha_{1H}, \theta_{1H}, \alpha_{2H}, \theta_{2H}, \alpha_{3H}, \theta_{3H}, \alpha_{51H}, \theta_{51H}, \alpha_{102H}, \theta_{102H}, \alpha_{357H}, \theta_{357H}, \alpha_{714H}, \theta_{714H}\} \quad (3.2)$$

The 15 harmonics of the AFD are not capable of modeling some of the complex characteristics of load profiles, such as daily double peaks and/or differences in weekday and weekend patterns. To incorporate such complex characteristics, two new attributes to be incorporated into the AFD are now introduced. The first proposed attribute is designed to distinguish commercial and residential load profiles which have the following characteristics.

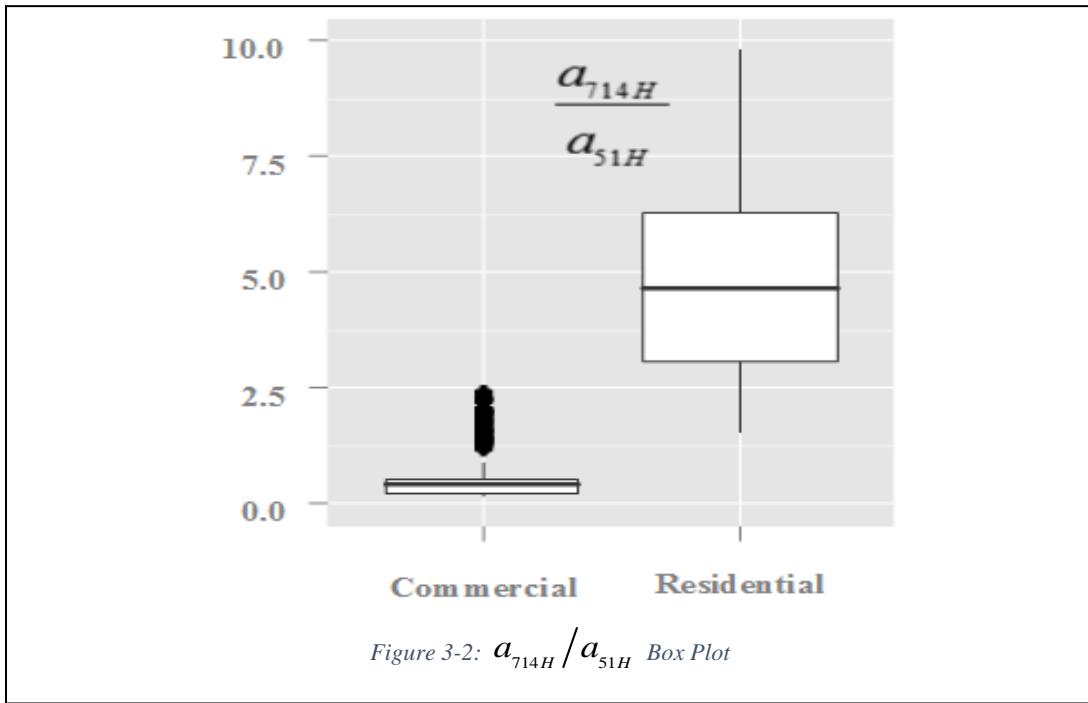
- 1) Residential double-peak daily pattern vs. commercial single-peak daily pattern.
- 2) The residential daily load pattern is more consistent, while the commercial load profile has a load drop/rise pattern between weekdays and weekends.

In order to captures the essence of these two characteristics. The following new attribute is used to supplement the load profile characteristic modeling in the AFD

$$AFD_{16} = a_{714H} / a_{51H} \quad (3.3)$$

For residential load profiles with significant double-peak daily patterns (large a_{714H}) and no weekday-weekend pattern (small a_{51H}), its AFD_{16} will have a higher value than its commercial counterpart's AFD_{16} because the commercial customer has a large difference between the weekday-weekend patterns and no double-daily pattern.

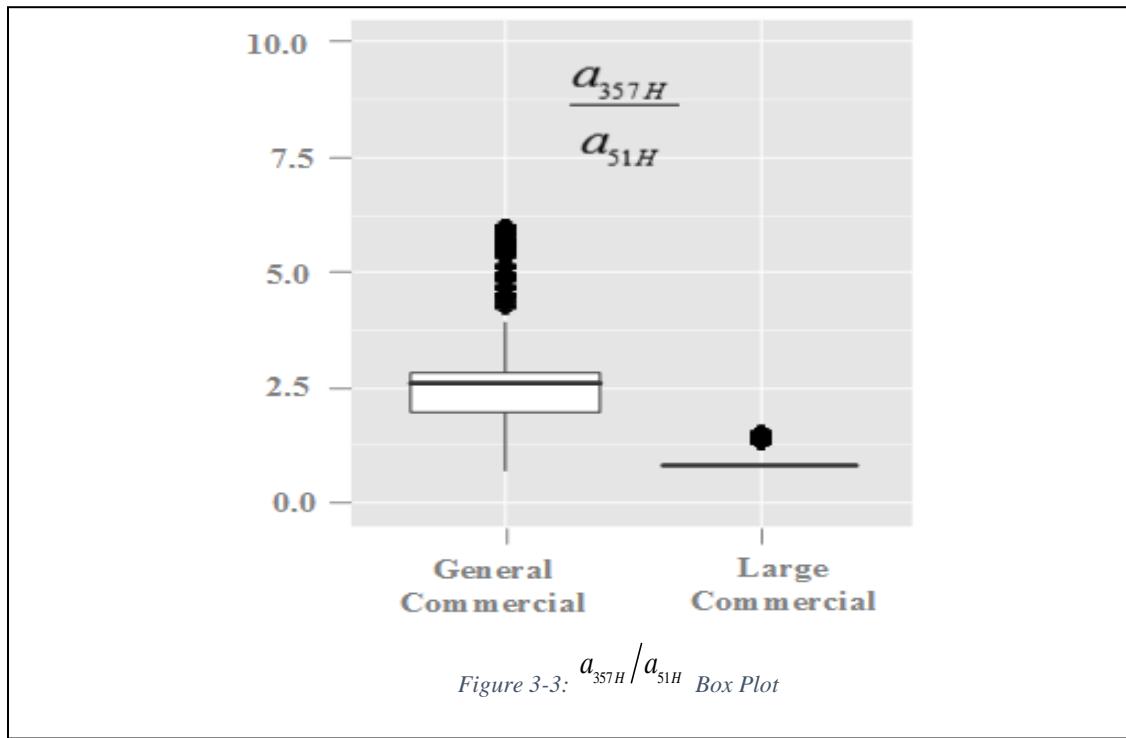
This attribute is presented as a box-plot in Figure 3-2. For commercial load profiles, AFD_{16} has a very small and consistent distribution. For residential load profile, AFD_{16} has a significantly higher value and a wider variation. There is a significant separation between the residential and the commercial classes for their AFD_{16} values. It will be beneficial to include AFD_{16} to construct the classification tree.



The second AFD is proposed to model the common load-drop pattern that occurs between weekdays and weekends in commercial load profiles. This characteristic is related to the magnitude of the weekly harmonic component, a_{51H} , and the magnitude of the daily harmonic component, a_{357H} . The following attribute AFD_{17} is used to model this characteristic:

$$AFD_{17} = a_{357H} / a_{51H} \quad (3.4)$$

For example, the large commercial load profiles normally have a consistent daily pattern and not much load drop between the weekday and weekend load, which leads to the AFD_{17} value being close to 1. For commercial load profiles that have a higher weekday load level than the weekend load level (high a_{357H} and low a_{51H}), their AFD_{17} value will be larger than 1. In Figure 3-3 there is a clear separation between the large commercial and other commercial classes' AFD_{17} distribution, which indicates that this attribute is a good classification input parameter to classify large commercial and other types of commercial load profiles.



These 17 AFDs can be used to describe the aforementioned load profile characteristics. Traditional statistical methods are poorly suited for comparing such large numbers of variables in a classification process [35]. On the other hand, the proposed classification method constructs a hierarchical classification tree to identify load classes with homogenous characteristics. Each class signature can be defined by a set of CAFDs.

For example, the residential load profile in Figure 1-4 can be defined by the two CAFDs in stage IV of Figure 3-1. The load profile's CAFD-based signature is presented in stage V. This type of signature is easy to process by both human and machine [44].

3.3 Classification Procedure

After processing the annual load profile data into 51-week load profiles, the following procedures were implemented to construct the hierarchical classification trees in the frequency domain:

- 1) Transform the 51-week, normalized hourly load profile data into its frequency domain representation, an 8568-point DFT.
- 2) Identify the periodic patterns in the load profile and their related harmonics and frequency domain attributes.
- 3) Construct a classification tree using identified frequency domain attributes as predictor variables. Experiment with different CP values to fine-tune the number of final terminal leaves (classes) of the tree.
- 4) Summarize classification rules of the leaf node on the classification tree as CAFD, which will serve as the definition of load profile classes for classification.
- 5) Each terminal leave can be further divided into a set of subclasses using steps 1-4, depending on the expectation of the classification (such as the different targeted characteristic attributes). This will construct a multi-level hierarchical classification tree for classifying different load profile classes and their respective subclasses.

Figure 3-4 is the two-level hierarchical classification tree constructed with AFD₁₋₁₇ as the predictor input variables. The level 1 portion of tree is used to classify commercial and residential types of load profiles. The level 2 portion is used to further classify the commercial load profiles (CAT I.1) into several more homogenous sub-classes.

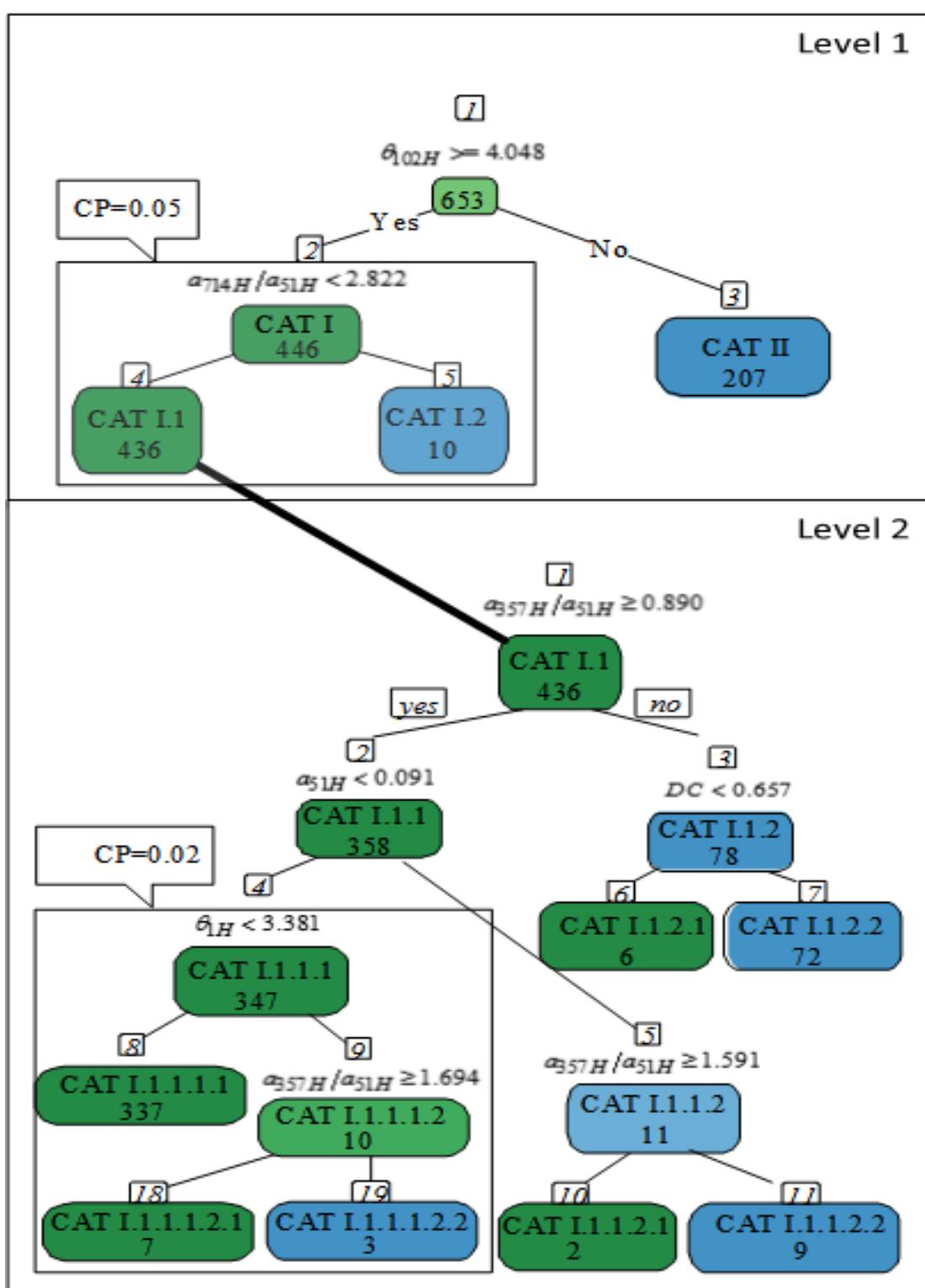


Figure 3-4: Hierarchical Classification Tree

3.4 Hierarchical Frequency Domain Load Profile Classification

There is no established way to label load profile classes based on their characteristic attributes. In this thesis, the load profile classes are labeled hierarchically by roman numerals and numerical indexes. Roman numerals are used to represent the partitioned root level data spaces. The dot notation is used to signify the hierarchical relationship, where each dot indicates one more data space partitioning operation and adds one more CAFD.

In Figure 3-4 the number at each node represents the number of load profiles that satisfy the conditional path from the root. The left branch of each internal node represents the TRUE path from evaluating if each load profile at current node satisfies the CAFD (displayed on the top of each internal node), while the right branch represents the FALSE path.

The level 1 portion of the classification tree has at most 3 load profile classes (terminal leaves) when CP=0.0. The *rpart* software produces a table that lists the CP value for each possible branch pruning, which suggests that if CP is relaxed to 0.05, the CAT I.1, CAT I.2 in the box can be pruned back into one category as CAT I. After evaluating the purity level at each split for all 16 input variables, the classification tree at level 1 has identified 2 CAFDs, ($(\theta_{102H} < 4.048)$ and $(a_{714H}/a_{51H} < 2.822)$), that can be used to specify the boundary conditions for the 3 parameter spaces with the highest purity score.

Figure 3-5 illustrates how the 653 load profile data space is partitioned by these two CAFDs at level 1 of the classification tree. The first partitioning operation occurs along the $\theta_{102H} = 4.048$ reference line (thick dashed line). $CAFID_1(\theta_{102H} \geq 4.048)$ is the most important CAFD in this partition process because it is the first CAFD chosen by the CART and carves out the largest data space with high purity. The CART then chooses AFD₁₆ (used to

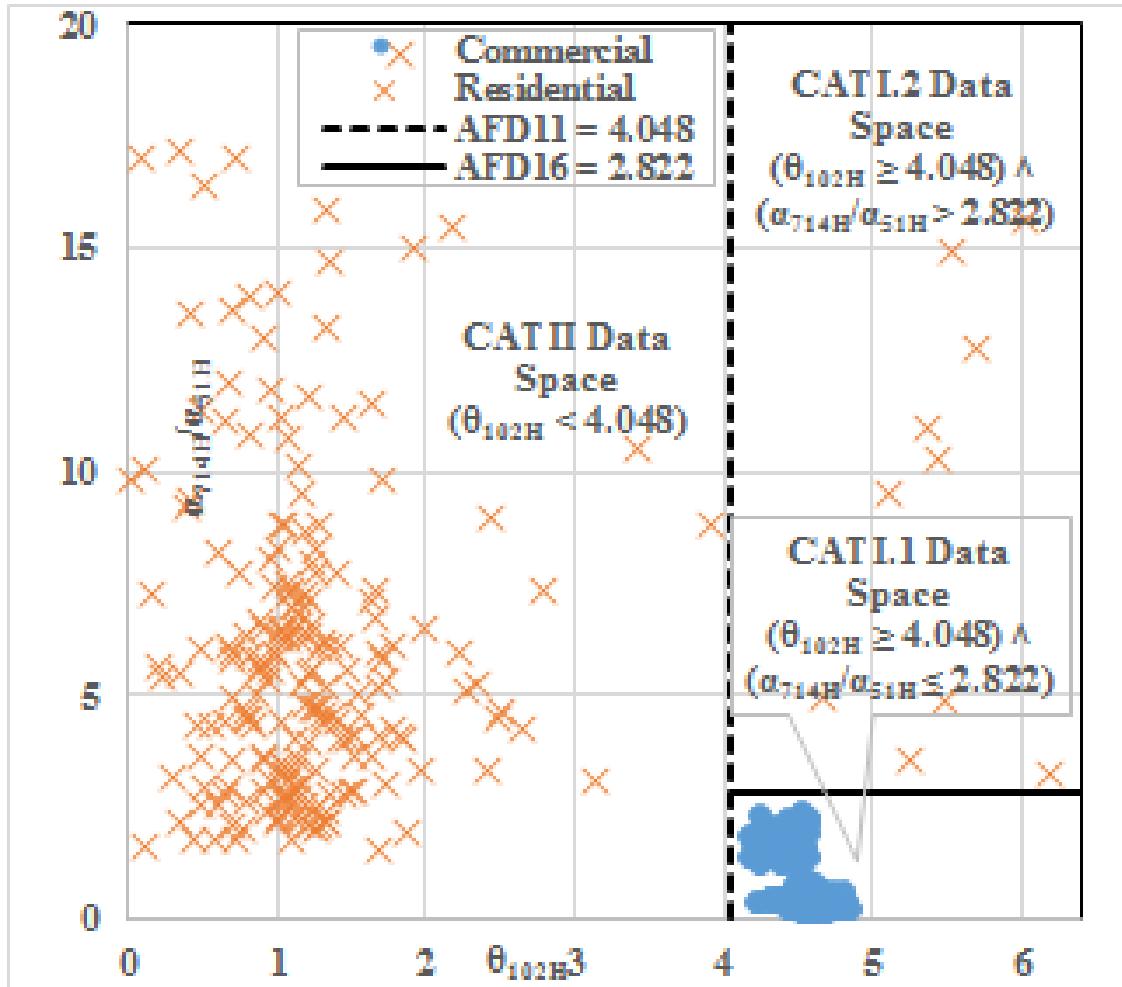


Figure 3-5: Partition of the feasible space created by θ_{102H} and a_{714H}/a_{51H}

describe the double-peak pattern) to partition CAT I data space along the $a_{714H}/a_{51H} = 2.822$ reference line t. It divides CAT I's data space into CAT I.2 and CAT I.1 data spaces. The commercial load profile's data space (CAT I) is quite compact because the commercial type of load profiles' load patterns is more consistent. Residential load profiles have more variations in term of their pattern characteristics, which leads to a broader data space.

The level 2 of the classification tree in Figure 3-4 will further divide the load profiles in CAT I.1 into several subclasses. The subclasses will all inherit CAT I as prefix to indicate that level 2 load profiles are finer partitioned space inside CAT I.1's data space. This portion of the classification tree has 7 classes listed in Table 3-3 (page 67) when the

CP=0.0. When using *rpart* suggested next pruning value CP=0.02, CAT I.1.1.1.1, CAT I.1.1.1.2.1 and CAT I.1.1.1.2.2 will be pruned back into CAT I.1.1.1 as node 4 at the level 2 of the classification tree.

This classification tree has identified 6 CAFDs based on four predictor variables: a_{357H}/a_{51H} , a_{51H} , θ_{1H} and DC. Each variable represents a load characteristic attribute. The a_{357H}/a_{51H} can be used to describe the weekday weekend load relationship. The a_{51H} can be used to describe the weekly pattern. The θ_{1H} is related to the seasonal pattern and DC is related to the load factor of the load profile.

At level 2 of the classification tree, $CAFID_{17} (a_{357H}/a_{51H} \geq 0.890)$ is the most important condition used by the tree to partition the CAT I.1 data space. The load profiles in node 2 have smaller CAFD₁₇ values than the load profiles in node 3, Comparing with load profiles in node 3, the load profiles in node 2 have more dominant daily pattern than the weekend-load-drop pattern.

Figure 3-6 presents the CAFD-base signatures for all the classes in Figure 3-4. The CAFD-based signature is a joined set of the CAFDs along the path from the root to each class (leaf) node. The \wedge symbol represents the Conjunction (AND operation) and the \neg symbol represents the Negation (NOT operation). For example, the CAT I.1's signature can be defined as $CAFID_1 \wedge CAFID_2$. The CAT I.1.1.1 can be defined as $CAFID_1 \wedge CAFID_2 \wedge CAFID_3 \wedge CAFID_4$ because it's a subset of CAT I.1 and naturally inherits its CAFD-based signature: $CAFID_1 \wedge CAFID_2$.

This notation is significant because it not only defines the load profile using general quantitated parameters but also defines the relationship between different load classes.

CAFD	
CAFD₁ : $\theta_{51H} \geq 4.048$	CAFD₅ : $\theta_{1H} < 3.381$
CAFD₂ : $a_{714H}/a_{51H} < 2.822$	CAFD₆ : $a_{357H}/a_{51H} \geq 1.694$
CAFD₃ : $a_{357H}/a_{51H} \geq 0.890$	CAFD₇ : $a_{357H}/a_{51H} \geq 1.591$
CAFD₄ : $a_{51H} < 0.097$	CAFD₈ : DC < 0.657

Load Class	CAFD-based Signature
CAT I	CAFD ₁
CAT I.1	CAFD ₁ \wedge CAFD ₂
CAT I.1.1	CAFD ₁ \wedge CAFD ₂ \wedge CAFD ₃
CAT I.1.1.1	CAFD ₁ \wedge CAFD ₂ \wedge CAFD ₃ \wedge CAFD ₄ \wedge CAFD ₅
CAT I.1.1.1.2	CAFD ₁ \wedge CAFD ₂ \wedge CAFD ₃ \wedge CAFD ₄ \wedge \neg CAFD ₅
CAT I.1.1.1.2.1	CAFD ₁ \wedge CAFD ₂ \wedge CAFD ₃ \wedge CAFD ₄ \wedge \neg CAFD ₅ \wedge CAFD ₆
CAT I.1.1.1.2.2	CAFD ₁ \wedge CAFD ₂ \wedge CAFD ₃ \wedge CAFD ₄ \wedge \neg CAFD ₅ \wedge \neg CAFD ₆
CAT I.1.1.2	CAFD ₁ \wedge CAFD ₂ \wedge CAFD ₃ \wedge \neg CAFD ₄
CAT I.1.1.2.1	CAFD ₁ \wedge CAFD ₂ \wedge CAFD ₃ \wedge \neg CAFD ₄ \wedge CAFD ₇
CAT I.1.1.2.2	CAFD ₁ \wedge CAFD ₂ \wedge CAFD ₃ \wedge \neg CAFD ₄ \wedge \neg CAFD ₇
CAT I.1.2	CAFD ₁ \wedge CAFD ₂ \wedge \neg CAFD ₃
CAT I.1.2.1	CAFD ₁ \wedge CAFD ₂ \wedge \neg CAFD ₃ \wedge CAFD ₈
CAT I.1.2.2	CAFD ₁ \wedge CAFD ₂ \wedge \neg CAFD ₃ \wedge \neg CAFD ₈
CAT I.2	CAFD ₁ \wedge \neg CAFD ₂
CAT II	\neg CAFD ₁

Figure 3-6: CAFD-Based Signatures

From CAT I.1.1.'s definition, it can be interpreted that this load profile class not only inherits the CAT I.1.'s characteristic attributes but also contains two additional attributes that set it apart from the other classes in CAT I.1.

One potential application for this signature is to encode it as XML (Extensible Markup Language, an open standard that has been widely adapted for data exchange in many industries [45]) document. It will provide a standard and efficient way for different systems/utilities to exchange the load profile class definitions. For example, the CAT I.1.'s definition in Figure 3-1 can be formatted as the XML document shown in Figure 3-7. The

```
<?xml version="1.0"?>
<LoadClass>
  <Name>CATI.1</Name>
  <AFDSet>
    <AFD Name="CAFD1" Definition="Phase102"/>
    <AFD Name="CAFD2" Definition="Mag714H/Mag51H"/>
  </AFDSet>
  <ClassDefinition>
    <CAFD AFDName="CAFD1" Operator=">=" Value="4.048">
      <CAFD AFDName="CAFD2" Operator="<" Value="2.822"/>
    </CAFD>
  </ClassDefinition>
</LoadClass>
</xml>
```

Figure 3-7: XML document for CAFD-based Signature

nested formation of CAFD nodes (inside **<ClassDefinition>** node) naturally reflects the hierarchical structure of CAFD-based signature. This type of signature can form a key to index each load class, therefore making it possible to build a load profile database with searchable keys based on CAFD-based signatures.

3.5 Comparison between Frequency Domain Hierarchical Load Profile Classification and Utilities' Current Practices

The CAFD-based load profile classes are compared with the conventional utility load profile classes in Table 3-2 and 3-2. Classes in Table 3-1 are from level 1 of the classification tree and classes in Table 3-2 are from level 2.

Table 3-2 shows that the CAT I.1's load profiles can all be mapped to the commercial type of load classes defined by 5 different utilities (highlighted by green). The CAT I.2 and CAT II's load profiles are all mapped to residential type of load profiles by 5 utilities (highlighted by yellow).

There are several important observations from Table 3-2:

Chapter 1: Much commonality exists between different utilities classification: CAT I.1 have commercial load profiles from 5 different utilities.

Chapter 2: There is more variation among residential classes from different utilities. Some load profiles from same load class by utilities are assigned with different CAFD-based signatures: such as RESLOWR from ERCOT is assigned to CAT I.2 and CAT II.

Chapter 3: CAT I.2's load profiles are defined as residential by utilities, even though it shares the same CAFD₁ with CAT I.1, whose members are all commercial classes. It indicates certain residential classes' load profiles share some similar patterns with other utilities' commercial class load profiles. In the end, the most pronounced residential load pattern (double peak) modeled by CAFD₂ is identified and the load profiles are assigned to the right class.

Table 3-2: Comparison of CAFD-based classes and Utility Commercial & Residential Load Classes

CP Value		Utility Load Class	Utility
0.05	0		
Load Class			
CAT I	CAT I.1	A1	PGE
		A10	PGE
		A6	PGE
		BUSHILF	ERCOT
		BUSIDRRQ	ERCOT
		BUSLOLF	ERCOT
		BUSMEDLF	ERCOT
		BUSNODEM	ERCOT
		GLP	PSEG
		LPL	PSEG
		GS-1	SCE
		GS-2	SCE
		SC2DEM	NG
		SC2ND	NG
	CAT I.2	RESLOWR	ERCOT
		RHS	PSEG
		RLM	PSEG
		SC1C	NG
CAT II	CAT II	DOM-S/M	SCE
		E1	PGE
		E7	PGE
		RESHIWR	ERCOT
		RESLOWR	ERCOT
		RHS	PSEG
		RLM	PSEG
		RS	PSEG
		SC1C	NG
		SC1STD	NG
Commercial		Residential	

The conventional utility classification methods do not utilize the load profiles' frequency domain pattern characteristics. As shown in the comparison between CAFD-based classes and the utilities' classes, the conventional method can assign many load profiles with similar pattern characteristics into different classes. They may also assign the load profiles with different pattern characteristics into one class.

Comparing with the more pronounced characteristic attributes variation between the commercial and the residential load profiles at level 1 of the classification tree, the load pattern differences among commercial load profiles are subtle as it requires more CAFDs to define the boundaries for these more closely resembled load profiles.

4 CAFDs are used in level 2 of the classification tree to define the categories' parameter spaces. The first CAFD, $CAF D_3 (a_{357H}/a_{51H} \geq 0.89)$ divides the CAT I.1 load profiles into the following two groups:

Group 1: The CAT I.1.1 and CAT I.1.2 all have $CAF D_3$. The CAFD's high value indicates the CAT I.1.1-2 load profiles have more consistent weekly pattern, less variation between weekday and weekend load.

Group 2: The CAT I.1.2.1 and CAT I.1.2.2's load profiles share a common $-CAF D_3$, which indicates a larger difference between their weekend and weekday usage. The load profiles in CAT I.1.2.2 (BUSIDRRQ: large commercial by ERCOT definition [35]) are less dynamic than CAT I.1.2.1 because of their higher DC values in $CAF D_8$, which also indicates that the CAT I.1.2.2 load profiles' overall patterns are more consistent than the CAT I.1.2.1 load profiles (resembling more the traditional large commercial load profile pattern).

There are percentages associated with each load class in Table 3-3. It represents the percentage of the utility load class' samples belonging to the mapped CAFD-based class. Table 3-3 shows that the largest sub-category CAT I.1.1.1 is defined as $CAF D_3 \wedge CAF D_4 \wedge CAF D_5$ which consists of load profiles from 4 utilities: 2 utilities' general commercial load profiles satisfy this rule 100%. The other 2 utilities' general commercial also show relatively high matching rate (the lowest is 71%).

Table 3-3: Comparison of CAFD-based Classes and Utility Commercial Load Classes

CP Value		Utility Load Class	Utility
0.02	0.00		
CAT I.1.1.1	CAT I.1.1.1.1	A1 100%	PGE
		A10 100%	PGE
		A6 100%	PGE
		BUSHILF 100%	ERCOT
		BUSLOLF 100%	ERCOT
		BUSMEDLF 100%	ERCOT
		BUSNODEM 92%	ERCOT
		GLP 100%	PSEG
		GS-1 100%	SEC
		GS-2 100%	SEC
CAT I.1.1.1.2.1	SC2DEM	SC2DEM 71%	NG
		SC2DEM 29%	NG
	SC2ND	SC2ND 71%	NG
CAT I.1.1.1.2.2		LPL 25%	PSEG
CAT I.1.1.2.1	SC2ND	SC2ND 29%	NG
CAT I.1.1.2.2	LPL	75%	PSEG
CAT I.1.2.1	BUSNODEM	8%	ERCOT
CAT I.1.2.2	BUSIDRRQ	100%	ERCOT
Commercial		Large Commercial	

BUSHILF (average monthly load factor > 0.6), BUSMEDLF (average monthly load factor between 0.4 and 0.6) and BUSLOLF (average monthly load factor < 0.4) load profiles from ERCOT are classified as CAT I.1.1.1.1. Their load factor value is very similar to the DC component in frequency domain which normally has the largest magnitude value. DC's large magnitude value doesn't mean DC by default is the most important predictor. CART evaluates all AFD input parameters equally and identifies AFD_{17} and AFD_8 are more effective than DC in partitioning the CAT I.1.'s data space into disjoint spaces with highest purity. These 3 ERCOT classes are similarity in term of these AFDs and therefore are assigned to the same class.

The CAT I.1.1.1.2.1 is defined as $CAFID_3 \wedge CAFID_4 \wedge \neg CAFID_5 \wedge CAFID_6$ which has same first 2 CAFDs as CAT I.1.1.1. The CAT I.1.1.1.2.2 has same first 3 CAFDs in its definition

as CAT I.1.1.1.2.1, with a different $-CAF_{D_6}$ term. CAF_{D_6} 's phase angle value is related to the position of annual peak in time domain.

The CAT I.1.1.2.2 is another category that consists of load profiles from a utility (PSEG) classified large commercial class. These two sets of load profiles both are labeled as large commercial types by their utilities while CAFD classification puts them in different categories. This example shows the ambiguity of the current practice of load profile classification: the large commercial load profiles from different utilities don't necessarily share the same characteristic attributes.

In Table 3-3, 12 LPL load profiles are divided into CAT I.1.1.2.2 (3 LPL load profiles with 25% probability) and CAT I.1.2.1 (9 LPL load profiles with 75% probability). Both categories share the highest ranked hierarchical CAFD condition A. Their differences start at the node 2 with B condition, which is the phase angle of 51st harmonic that is related to the weekly pattern timing.

The lack of a universal load profile classification methods can lead to inconsistent load class definition across different utilities. The examples show that LPL load profiles' weekly pattern may vary from year to year. It should not be assumed that the load profiles from a sampling group will have same characteristics year after year. It also shows that the LPL from PSEG doesn't have the consistency of BUSIDRRQ from ERCOT in term of their CAFDs.

3.6 Load Profile Classification

A load profile classification test has been developed based on the proposed approach. This test extracts 7 sets of AFD₁₋₁₇ from 7 new annual load profiles [46] (not from the original 5 utilities' 653 load profiles). These load profile's AFD sets are then fed into the decision tree in Figure 3-7 to be classified using CAFD signatures.

As shown in Table 3-4, every profile successfully finds its respective place in the CAFD-based classification hierarchy by matching their AFD component values to each class's CAFD signature. The RES load profile is classified as CAT II residential class. All 6 GS load profiles are part of the CAT I.1 commercial class hierarchy: the GS1 profile is classified as a CAT I.1.2.1. The GS2 to GS3 profiles share similar CAFD signature and are classified as CAT I.1.1.2.1. The GS4 profile is classified as CAT I.1.2.2 large commercial subclass.

Table 3-4: Load Profile Classification Test

Load Class	AEP	ERCOT	NG	PSEG	PGE	SCE
I.1.2.1	GS1	BUSNODEM				
I.1.1.2.1	GS2O		SC2ND			
I.1.1.2.1	GS2U		SC2ND			
I.1.1.2.1	GS3O		SC2ND			
I.1.1.2.1	GS3U		SC2ND			
I.1.2.2	GS4	BUSIDRRQ				
II	RES	RESHIWR RESLOWWR	SC1C	RHS, RS RLM	E1 E7	DOM
Residential		Commercial		Large Commercial		

The efficacy of the proposed approach has been established by the success of this test. This is significant improvement over current practices that provide mostly qualitative labeling.

3.7 Summary

This chapter has accomplished the following: First, the concept of CAFD for load profile is introduced. The CAFD used in classification process can elegantly describe the load profile's characteristics. It has the potential to improve knowledge sharing among different entities since now a load profile can be described in a standard, modular, coherent and portable format that is independent from each utility's load class terminologies.

Using the hierarchical classification tree with CAFDs can efficiently and effectively classify the long and complex annual load profiles from different utilities. The hierarchical classification tree is presented to show that this classification method is very efficient and effective. As shown in Figure 3-4 and Table 3-2, two CAFD ($\theta_{102H} < 4.048$) and ($a_{714H}/a_{51H} < 2.822$) can separate commercial from residential load profiles. In Figure 3-4 and Table 3-3, it shows that two CAFDs ($a_{357H}/a_{51H} \geq 0.89$) and ($DC > 0.657$) can separate large commercial from other commercial load profiles.

Second, the proposed approach introduces an innovative way to classify load profiles. This approach is universal applicable to all utilities' load profiles because it utilizes a set of universal components (AFDs) that are directly derived from the profile data. New incoming annual load profile can be successfully classified into one of the identified load classes by matching its AFD component values to the load classes' CAFD signatures.

Third, using the proposed approach, each load class can be characterized by a CAFD-based signature. Taken together, the quantitative information contained in these signatures could deepen the understanding of load profiles and may lead to the development of new approaches in demand-side management and demand response.

Fourth, the CAFD-based signature can be easily interpreted to show the differences or the similarities in load profiles' characteristic attributes among different class. It can be coded as a self-describing XML document that is both human and machine readable. Such load profile XML document can be easily exchanged through internet and efficiently processed using standard software tools.

Fifth, this chapter has shown that the current utility classification is not precise, and is limited to a small number of characteristic attributes of the load profile patterns in time domain. On the other hand, the hierarchical CAFD classification process can precisely target the characteristic attributes that each user wants to focus on. The number of characteristic attributes is scalable and can incorporate a large number of CAFDs depending on the goal of the classification.

At last, this classification method is a systematic and efficient way to classify load profiles. It has high adaptability potential because the CAFD classification rule is easy for human and machine to interpret and well suited for automation. With additional research, this method has the potential to be used to construct a universal comprehensive CAFD-based load profile database that can be used to study load profiles from all sources.

Chapter 4: Wavelet based Load Model for AMI Data

After introducing the framework of frequency domain load modeling and classification, this part of the thesis will focus on estimating and forecasting the daily and annual load profiles.

As more and more AMI data comes on-line, utilities are accumulating tremendous amounts of load profile data. Power flow analysis, as well as many other power system analysis applications, can benefit from the load data collected from Advanced Metering Infrastructure (AMI). Planning, forecasting, automated customer type identification and classification, real-time analysis, and even real-time control can benefit from AMI load data or information derived from AMI load data.

Some previous efforts in using customer load measurements in power system analysis have used load research statistics to generate 8760 hourly statistical load models for classes of customers [47]. The AMI implementation gives a utility the ability to collect load data from each customer every 15 to 60 minutes. Thus, each customer can have 8,760 to 35,040 time-stamped load measurements annually. For a small circuit with 100 customers, an annual AMI load data table in a database can contain from 876,000 hourly load data rows to 3,540,000 rows.

The sheer size of the AMI load data makes the direct AMI data integration somewhat impractical for large scale system analysis. For example, using double precision numbers a utility with 1 million customers would require $1,000,000 \times 8 \text{ bytes} \times 8,760 \text{ hours/year} \approx 65\text{G bytes}$ of RAM to store the annual hourly AMI load data in memory.

There is a need for load models that have the following characteristics: 1) the load model can be used in the field without network connections, 2) the load model needs to be compact so that it will fit into laptops commonly used in the field, and 3) the load model needs to be efficient and in-memory because of time constraints of some analysis, where querying the AMI data warehouse constantly is impractical.

Currently there is lack of efficient, and compact AMI hourly load models that can be integrated with a large-scale system model in computer memory and used in analysis such as a power flow time-series analysis. The work here focuses on developing such load models using wavelet technology.

The first load model considered in the work here is the wavelet load model, which uses the Discrete Wavelet Transformation (DWT) [49] to transform the original load profile from the time domain to the wavelet domain. With the wavelet load model, each individual customer's AMI data is compressed, and load models are maintained for each individual customer.

The second load model considered here is the classified wavelet load model. With the classified wavelet model a single load model is used for many customers that exhibit similar load behavior. With wavelet based load models, the determination of which class a given customer should be assigned to, based on having similar time varying behavior, can be automated.

Conventionally, load profile classification is based on the load profile's time domain representation [2] or frequency domain descriptors [27]. Time domain parameters used to describe load profile characteristics include Base Load, Peak Load, Rise Time, Fall Time, and High-Load Duration. After a set of variables is established to describe load

profiles, various models, such as linear discriminant analysis, nearest neighbor classification, k-means, fuzzy-statistics, neural networks, and support vector machines are used to classify load profiles [18-23, 35].

A major challenge in classifying load profiles is that AMI load data for an individual customer may contain extreme data changes, such as the load stepping to zero due to a missed reading in the AMI system. Sometimes these extreme changes are errors, but not always. Figure 4-1 illustrates representative AMI hourly load samples for a commercial customer. The aforementioned classification methods, where relatively small variable sets are employed, have trouble in accurately capturing the pattern characteristics of Figure 4-1.

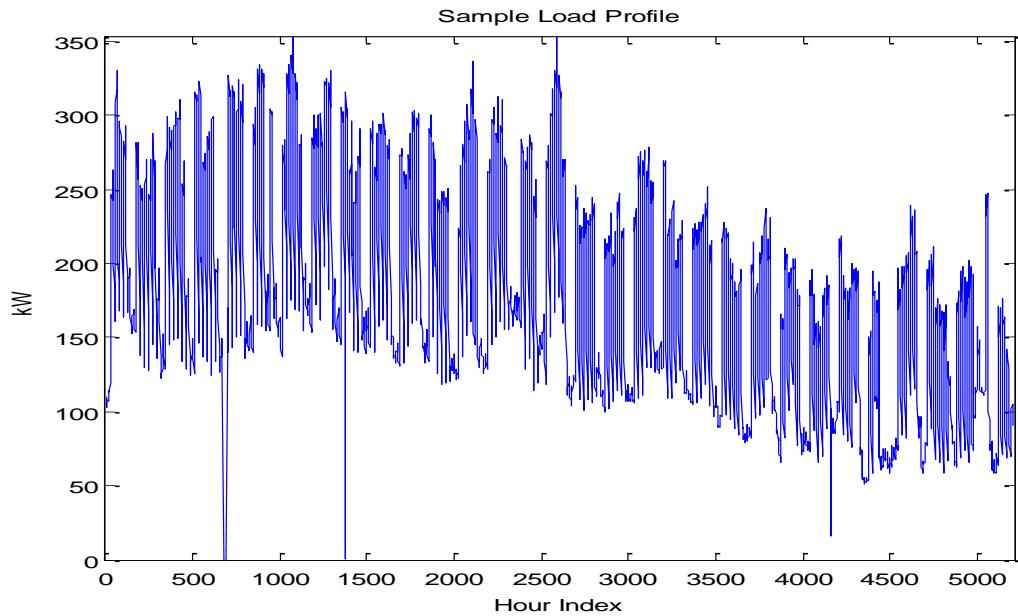


Figure 4-1: Sample AMI Load Profile Data

In order to better visualize load profile pattern characteristics, a 2 dimensional, or 2D, wavelet load profile representation is introduced. It will be illustrated that the significant load patterns in an individual customer's AMI load data can be modeled by the wavelet components from the load profile's lower 2D DWT transformation. Furthermore,

the 2D DWT transformation can be used in an unsupervised clustering process to identify load classes. The classified wavelet load models are derived from the load classes identified by the 2D DWT based clustering algorithm.

4.1 Discrete Wavelet Transformation

The development of wavelet transformations began with Alfréd Haar's work in the early 20th century [48]. The wavelet related research accelerated after the ground-breaking works from Ingrid Daubechies [48] and Stéphane Mallat [49] in the 1980s. There are many applications of wavelets, including digital signal processing and image processing [50].

In the DWT process, a time-domain discrete signal $s(n)$ can be decomposed into a set of Approximation Coefficients $A_{j_0,k}$ (equation 4.1) and Detail Coefficients $D_{j,k}$ (equation 4.2) with a predetermined discrete scaling function $\varphi_{j,k}(n)$ (equation 4.3) and wavelet function $\psi_{j,k}(n)$ (equation 4.4). The $s(n)$ can be reconstructed using its $A_{j_0,k}$ and $D_{j,k}$ in the inverse DWT (IDWT) process (equation 4.5).

$$A_{j_0,k} = M^{-\frac{1}{2}j} \sum_n S(n) \varphi_{j_0,k}(n) \quad (4.1)$$

$$D_{j,k} = M^{-\frac{1}{2}j} \sum_n S(n) \psi_{j,k}(n) \quad (4.2)$$

$$\varphi_{j,k}(n) = 2^{j/2} \varphi(2^j n - k) \quad (4.3)$$

$$\psi_{j,k}(n) = 2^{j/2} \psi(2^j n - k) \quad (4.4)$$

$$S(n) = M^{-\frac{1}{2}j} \sum_k A_{j_0,k} \varphi_{j_0,k}(n) + \sum_{j=0}^{\infty} \sum_k D_{j,k} \psi_{j,k}(n) \quad (4.5)$$

where $n = 0, 1, \dots, M - 1$ and $M = 2^J$

decomposing level $j = 1, \dots, J - 1$, $j_0 = 0$

coefficient index $k = 0, 1, \dots, 2^{j-1}$

$\varphi_{j_0,k}(n)$ is the discrete scale function

$\psi_{j,k}(n)$ is the discrete wavelet function

4.1.1 Multi-Resolution DWT

Mallat introduced an efficient multi-resolution DWT/Inverse DWT (IDWT) algorithm in 1989 [49] which make DWT/IDWT implementation practical by taking advantage of the family of orthogonal, compact support wavelets introduced by Daubechies [48].

Mallat's pyramid DWT/IDWT algorithm used a set of discrete Quadrature Mirror Filters (QMF) ($h_\psi, h_{\psi'}, h_\phi, h_{\phi'}$) to decompose or reconstruct a discrete signal. The coefficients for the filters are pre-determined. The simplest Daubechies (DB) wavelet filters, DB1 (Haar wavelet), is used in the AMI load data DWT here. The DB1 scale and wavelet discrete filter coefficients are specified in Table 4-1.

Table 4-1: DB1 QMF Coefficients

Wavelet	$h_\psi = \left[-1/\sqrt{2}, 1/\sqrt{2} \right]$ $h_\psi = \left[1/\sqrt{2}, -1/\sqrt{2} \right]$
Scale	$h_\varphi = \left[1/\sqrt{2}, 1/\sqrt{2} \right]$ $h_\varphi = \left[1/\sqrt{2}, 1/\sqrt{2} \right]$

In the work that follows, $S[n]$ will represent the AMI data set for a single customer. Figure 4-2 illustrates a 3-level DWT process. The discrete signal $S[n]$ convolutes with the

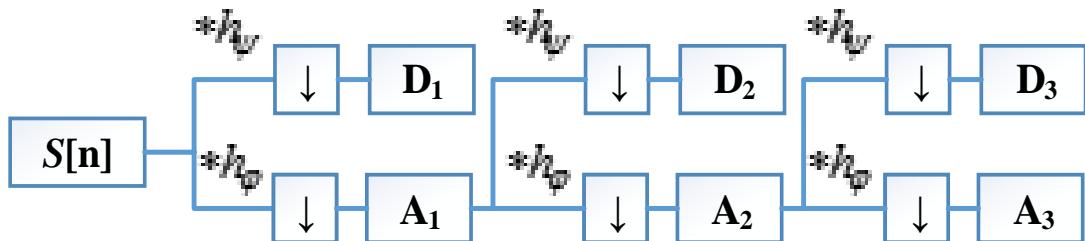


Figure 4-2: Multi-resolution DWT

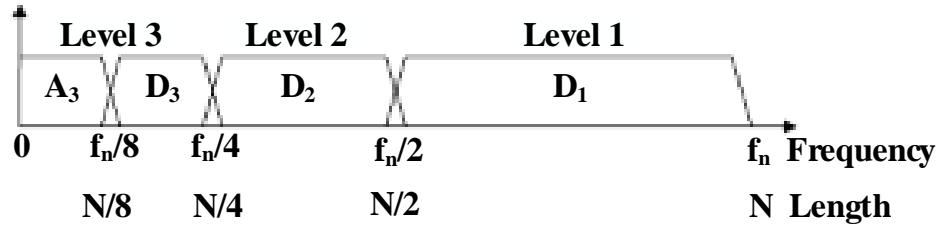


Figure 4-2: Multi-resolution DWT Frequency Band

low band-pass filter h_φ (covers 0 to $f_n/2$ frequency band as in Figure 4-3) which generates n number of values. Half of this set of numbers is redundant because it represents only the lower half of the frequency information in $S[n]$. Therefore, it can be down-sampled by 2 to get the Approximation Coefficients A_j . The down-sampling (decimation) process involves removing every other coefficient from the A_j approximation coefficients.

$S[n]$ also convolutes with the high band-pass filter h_ψ (covers $f_n/2$ to f_n frequency band) and is then down-sampled by 2 to calculate the Detail Coefficients D_j . The approximation coefficients A_j are then passed to the next level to repeat the same transformation process to generate the $j+1$ level A_j and D_j as specified in equations 4.6 and 4.7.

$$D_{j+1}[n] = \sum_{k=0}^N h_\psi[k] A_j[2n-k] \quad (4.6)$$

$$A_{j+1}[n] = \sum_{k=0}^N h_\varphi[k] A_j[2n-k] \quad (4.7)$$

N is the length of Finite Impulse Response Filter h_φ, h_ψ

Figure 4-4 illustrates a 3-level IDWT process which reverses the DWT process to reconstruct the $S[n]$ using A_j and D_j .

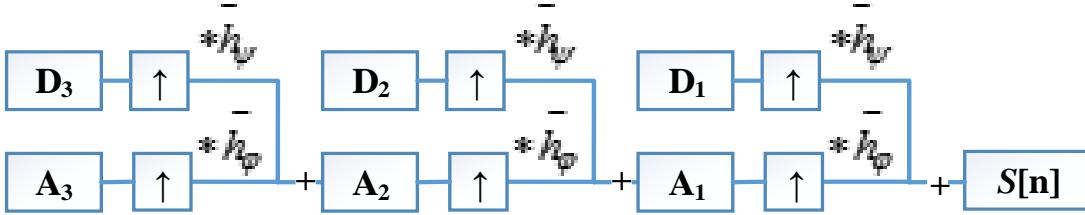


Figure 4-3: Multi-Resolution IDWT

During the IDWT process, the coefficients A_j and D_j at level j , are up-sampled by 2 with zero (expansion) and the up-sampled coefficients convolute with the mirror discrete filter h_ϕ and h_ψ respectively. The two convolution products are added together to generate coefficients A_{j-1} at level $j-1$ as in equation 4.8. These steps are repeated until the $S(n)$ is fully reconstructed.

$$A_{j-1}[n] = \sum_{k=0}^N h_\phi[k] A_j[2n-k] + \sum_{k=0}^N h_\psi[k] D_j[2n-k] \quad (4.8)$$

where N is the number of coefficients in h_ϕ, h_ψ

This DWT/IDWT algorithm described by equations (4.6-4.8) is efficient and has a linear computational complexity $O(n)$ that needs n operations to decompose a discrete load profile (n number of measurements) into the wavelet domain.

The flow chart in Figure 4-5 illustrate the process of multi resolution wavelet decomposition. The Table 4-2 presents a step-by-step numerical example for a 24-hour daily load profile's 3 level DWT. Figure 4-6 presents the synthesized load profiles at different resolutions. The left column contains the synthesized load profiles using

approximation coefficients. The top plot is the fully reconstructed load profile (same as original). The right column contains the synthesized load profiles using detail coefficients.

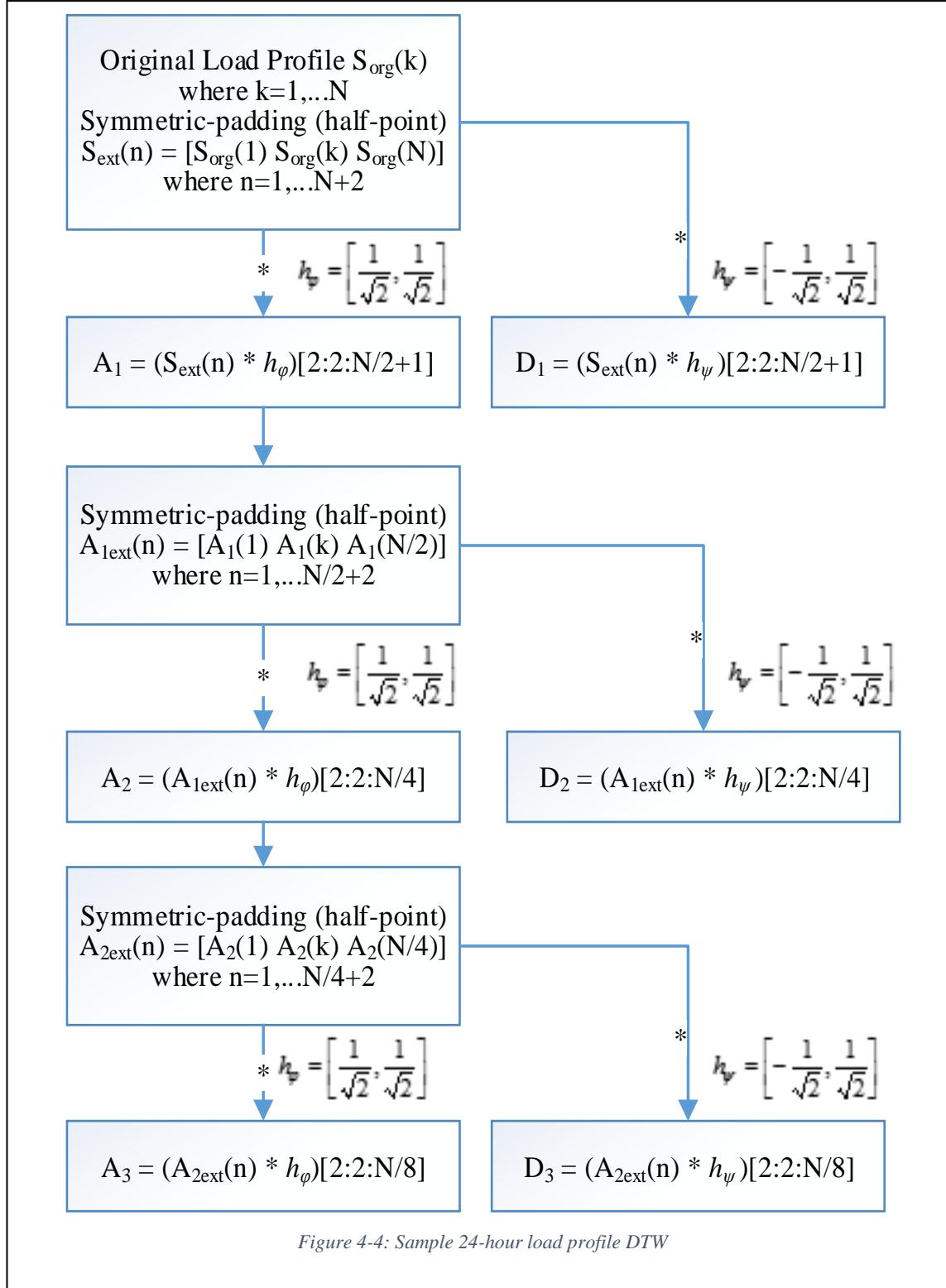


Table 4-2: Sample 24hour load profile DTW

$S_{ext}(n)$												
1	2	3	4	5	6	7	8	9	10	11	12	13
10.1	10.1	10.2	10.4	10.1	9.9	9.1	10.1	11.1	15.1	16.1	12.1	12.6
14	15	16	17	18	19	20	21	22	23	24	25	26
12.3	12.1	12.6	12.3	13.0	19.9	20.1	21.1	22.1	20.1	17.0	10.1	10.1
$(S_{ext} * h_\varphi) \downarrow 2$												
1	2	3	4	5	6	7	8	9	10	11	12	13
14.2	14.3	14.5	14.4	14.1	13.4	13.5	15.0	18.5	22.0	19.9	17.4	17.5
14	15	16	17	18	19	20	21	22	23	24	25	26
17.2	17.4	17.5	17.9	23.3	28.2	29.1	30.5	29.8	26.2	19.1	14.3	
$A_{j=1}$												
k	1	2	3	4	5	6	7	8	9	10	11	12
$A_{1,k}$	14.3	14.4	13.4	15.0	22.0	17.4	17.2	17.5	23.3	29.1	29.8	19.1
$(S_{ext} * h_\psi) \downarrow 2$												
t	2	3	4	5	6	7	8	9	10	11	12	13
0.00	-0.08	-0.13	0.21	0.14	0.57	-0.73	-0.67	-2.84	-0.72	2.84	-0.35	0.22
14	15	16	17	18	19	20	21	22	23	24	25	26
0.13	-0.35	0.22	-0.55	-4.83	-0.14	-0.72	-0.69	1.39	2.20	4.87	0.00	
$D_{j=1}$												
k	1	2	3	4	5	6	7	8	9	10	11	12
$D_{1,k}$	-0.1	0.2	0.6	-0.7	-0.7	-0.4	0.1	0.2	-4.8	-0.7	1.4	4.9
$A_{1ext}(n)$												
1	2	3	4	5	6	7	8	9	10	11	12	13
14.29	14.29	14.43	13.38	14.95	22.02	17.41						
8	9	10	11	12	13	14						
17.19	17.54	23.25	29.09	29.81	19.15	19.15						
$(A_{1ext} * h_\varphi) \downarrow 2$												
1	2	3	4	5	6	7	8	9	10	11	12	13
20.21	20.31	19.67	20.03	26.14	27.88	24.47						
8	9	10	11	12	13							
24.56	28.85	37.02	41.65	34.62	27.08							
$A_{j=2}$												
k	1	2	3	4	5	6	7	8	9	10	11	12
$A_{2,k}$	20.31	20.034	27.88	24.56	37.015	34.617						
$(A_{1ext} * h_\psi) \downarrow 2$												
1	2	3	4	5	6	7	8	9	10	11	12	13
0.0	-0.1	0.7	-1.1	-5.0	3.3	0.2						
8	9	10	11	12	13							
-0.2	-4.0	-4.1	-0.5	7.5	0.0							
$D_{j=2}$												
k	1	2	3	4	5	6	7	8	9	10	11	12
$D_{2,k}$	-0.1	-1.1	3.3	-0.2	-4.1	7.5						
$A_{2ext}(n)$												
1	2	3	4	5	6	7	8	9	10	11	12	13
20.3	20.3	20.0	27.9									
5	6	7	8									
24.6	37.0	34.6	34.6									
$(A_{2ext} * h_\varphi) \downarrow 2$												
28.7	28.5	33.9	37.1	43.5	50.7	49.0						
$A_{j=3}$												
k	1	2	3	4	5	6	7	8	9	10	11	12
$A_{3,k}$	28.5	37.1	50.7									
$(A_{2ext} * h_\psi) \downarrow 2$												
0.0	0.2	-5.5	2.3	-8.8	1.7	0.0						
$D_{j=3}$												
k	1	2	3	4	5	6	7	8	9	10	11	12
$D_{3,k}$	0.2	2.3	2.3	1.7	1.7	0.0						

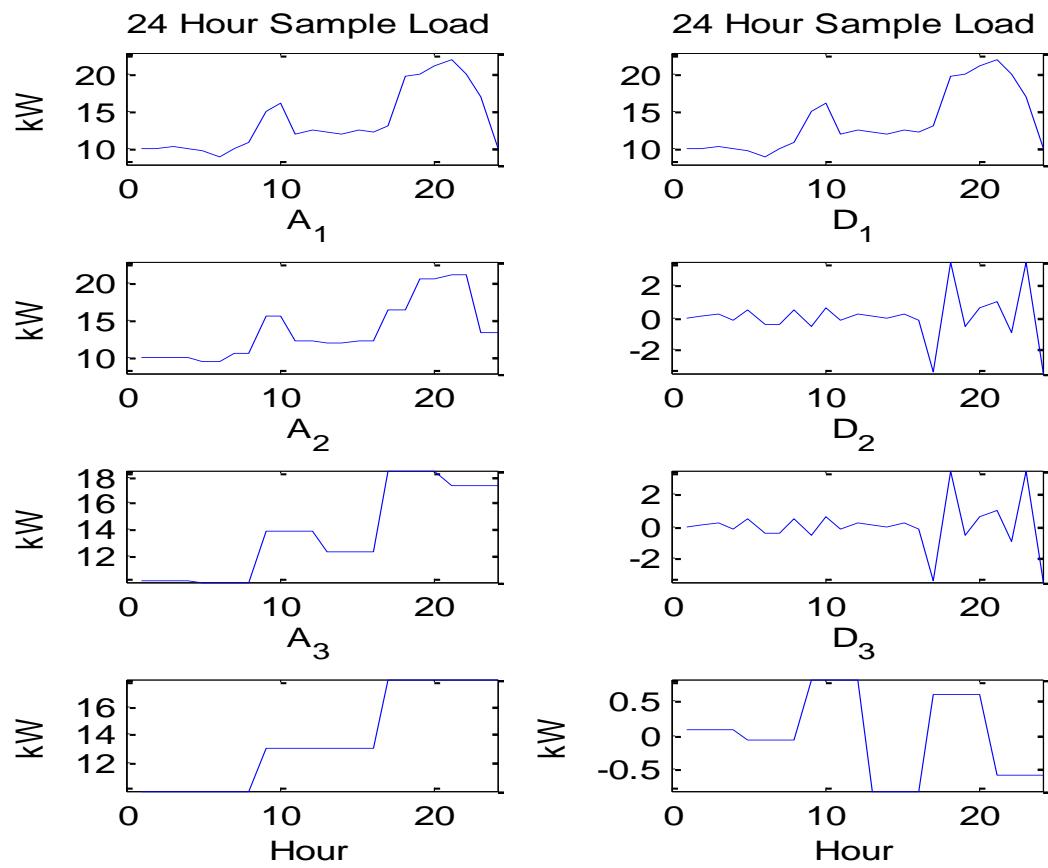


Figure 4-5: Synthesized 24hour load profile at different resolution

Figure 4-7 presents 3 level synthesized load profiles for a sample AMI load profile (with 5,208 measurements). The plots in the left column are the synthesized load profiles $S_{A_j}(n)$, which is reconstructed (equation 4.9-4.10) by using only the approximation coefficients A_j at level j (the D_j coefficients at all level j , are set to 0).

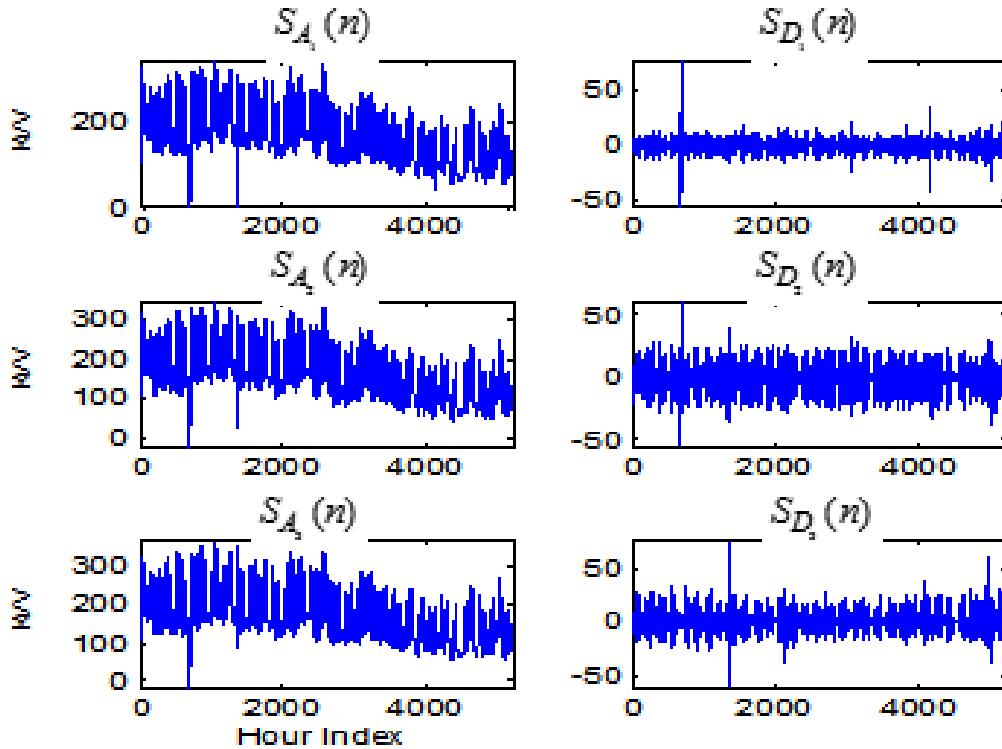


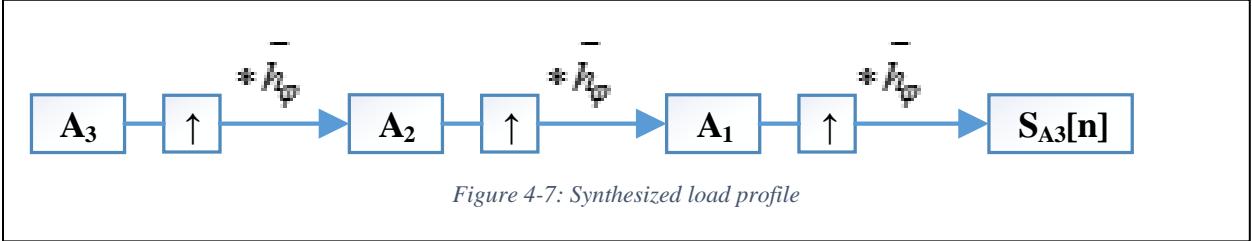
Figure 4-6: 3 level synthesized load profile

$$S_{A_j}[n] = \sum_{k=0}^N h_\varphi[k] A_j[2n-k] \quad (4.9)$$

$$A_{j-1}[n] = \sum_{k=0}^N h_\varphi[k] A_j[2n-k] \quad (4.10)$$

where N is the number of coefficients in h_φ
 $j \in [2, J]$ where J is the largest decomposition level.

This process is illustrated in Figure 4-8 which shows that detail coefficients are eliminated in the IDWT. These synthesized load profiles cover the lower frequency sub-band of the original AMI load profile.



The second column plots are the synthesized $s_{D_j(n)}$, which is reconstructed (illustrated in Figure 4-9) using only the detail coefficient at level j . The coefficients are not included in the synthesized process. These synthesized load profiles represent the higher frequency sub-band information in the original AMI load profile

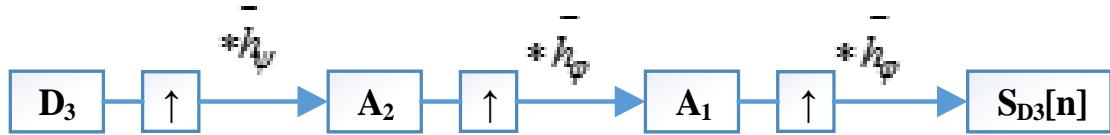


Figure 4-8: Synthesized load profile

The synthesized level j approximation load profile $s_{A_j}[n]$ needs twice as many coefficients as the level $j+1$ synthesized load profile $s_{A_{j+1}}[n]$. In Figure 4-7, the $s_{A_1}[n]$ load profile is modeled by $5,208 / 2^1 = 2,604$ coefficients, the $s_{A_2}[n]$ load profile is modeled by $5,208 / 2^2 = 1,302$ coefficients, and the $s_{A_3}[n]$ load profile is modeled by $5,208 / 2^3 = 651$ coefficients, which is 1/8 the size of the original load profile.

From visual inspection, the level 3 synthesized load profile is very similar to the original load profile presented in Figure 4-1. In section 4 it will be illustrated that the energy content of the signal is maintained by $s_{A_3}[n]$.

Wavelet coefficients can also be used in frequency-time localization analysis, such as the detection of large sudden changes in a load profile. In Figure 4-10 a small sample load profile and its 2 level DWT detail coefficients are presented. There are two sudden and significant load drops in the load profile: the first one takes place from the 51st to the 52nd hour, and the second one occurs at the 73rd hour. The 73rd hour drop has a higher frequency since its period is shorter. Once the load profile is transformed into the wavelet domain with 2-levels of resolution, the two load drops can be detected by analyzing the two detail coefficients sets (plots 2 and 3 in Figure 4-10). The D1 coefficient from the level 1 DWT covers the higher half of the frequency band ($f_n/2$ to f_n). The higher frequency 73rd hour load drop can be observed when the D1 magnitudes suddenly drops below -100 at the 74th hour. The lower frequency 51st to 52nd hour load drop can be detected by analyzing the D2 coefficients, which covers the lower half of the frequency band. In this case the D2 coefficient spikes to 100 at the 51st hour when the load drop occurs. In this

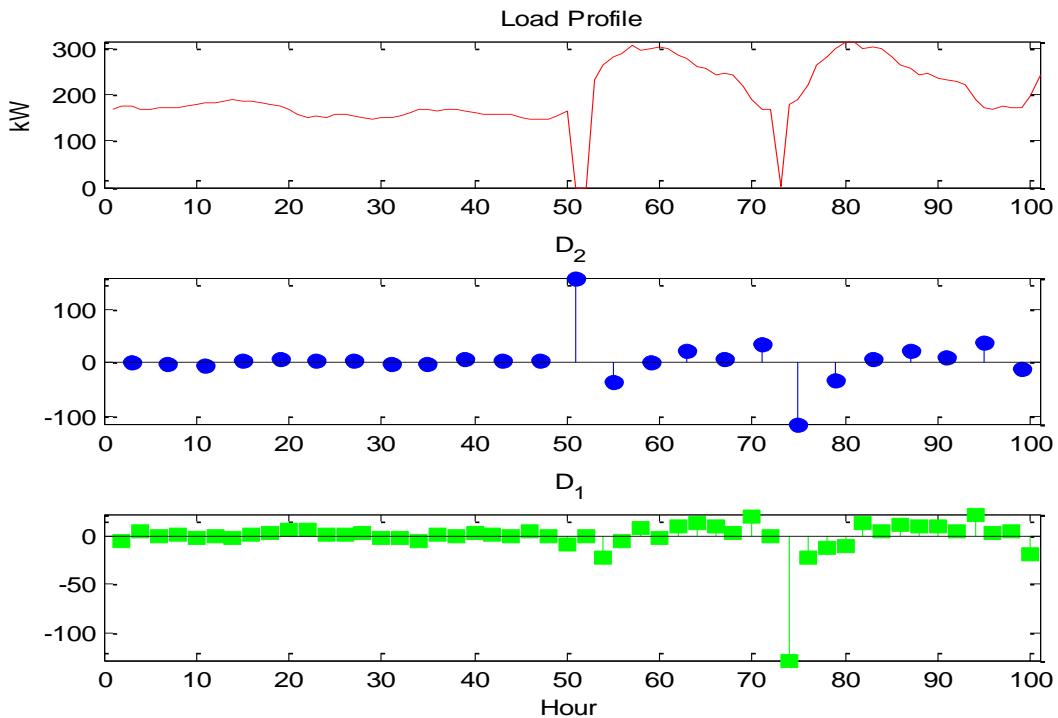


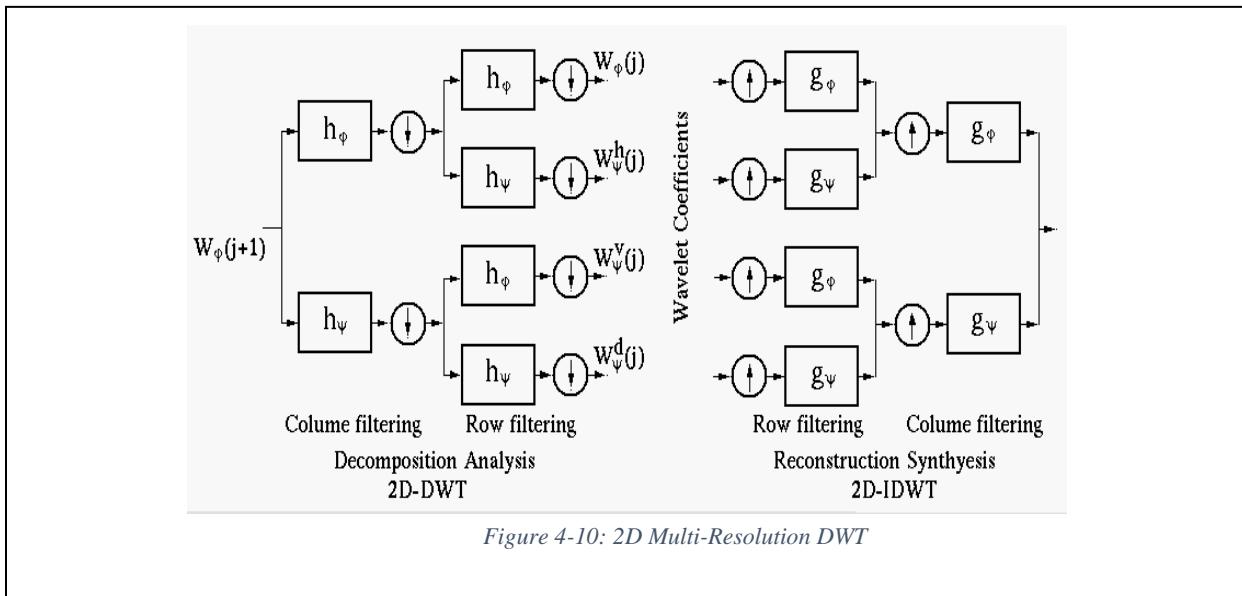
Figure 4-9: Load Profile Detail Coefficient from 2-Level DWT

example, the approximation coefficients are left out because the sudden load changes are high frequency signals that should not be visible in the lower frequency band where the approximation coefficients reside.

4.1.2 2D Multi-Resolution Wavelet Transformation

The 2D wavelet decomposition process is similar to the 1D DWT decomposition as the signal also goes through a series of discrete wavelet filters h_ψ, h_ϕ . The 2D discrete wavelet decomposition first decomposes the signal in columns, followed by a decomposition into rows, as illustrated in Figure 4-11. At each level the 2D wavelet transformation will generate one approximation coefficient $W_\phi(j)$ and three different detail coefficient sets: the horizontal $w_\psi^h(j)$, the vertical $w_\psi^v(j)$ and the diagonal coefficients $w_\psi^d(j)$.

The reconstruction process reverses the decomposition process which reconstructs the data by performing the IDWT along the row direction, followed by performing the IDWT along the column direction.



4.2 AMI Load Data Modeling using Wavelet

The real-world circuit used in this research serves 323 customers whose hourly AMI load measurements were recorded from May 2013 to Dec 2013. These customers represent a diverse group of load patterns. For example, there are customers with the double-peak residential load type, businesses with the typical 9am to 5pm commercial load type, loads with on-off type behavior (similar to street light load pattern), and others. Identifying the various load patterns for the circuit will be considered shortly.

In the analysis, AMI measurements from 2013-05-27 00:00 to 2013-12-29 23:00 were used, representing a 31-week time window. Each customer had 5208 hourly kW measurements, for a total of 1,682,184 rows of AMI measurements for the 323 customers served by the circuit.

4.2.1 Multi-Resolution Wavelet Load Model

For time-series analysis involving AMI data, SCADA data, and others, it is desirable to load the data into computer memory for the most rapid analysis. However, for models consisting of thousands of circuits with similar AMI coverage it is not be practical to load all of the AMI data directly into computer memory. The goal of a compressed AMI load data model is to construct a load model that accurately approximates the original load profile's energy and significant pattern characteristics with a much smaller data set.

Typical load profile patterns tend to be relatively slow changing, indicating that most load profile energy is located in the lower frequency band, which can be represented by the approximation coefficients A_j introduced earlier. The sudden, abnormal changes in a load profile have much higher frequencies which are modeled by the D_j detail coefficients. This kind of high frequency behavior modeled by the detail coefficients generally represents a small portion of the original time domain load profile's energy. From figure 3, it may be noted that the $s_{D_j}(n)$ load profiles have much smaller magnitudes than the corresponding $s_{A_j}(n)$ load profiles.

In this section the $s_{A_3}(n)$ load profile is used as the wavelet load model. This model has a coefficient set that is 1/8 of the original AMI data size. The model retains most of the energy residing in the original load profile and captures much of the original load profile's characteristics. However, it should be noted that it is possible to model the AMI load data at either higher (more coefficients) or lower (less coefficients) resolutions in the wavelet domain. But in this section characteristics of using $s_{A_3}(n)$ will be investigated.

In order to evaluate the performance of the load model, the following 3 parameters are proposed. First, the SE in equation 4.11 is the synthesized load profile's total energy in terms of percentage of the original load profile's total energy.

$$SE = \left[\sum_{n=1}^N S_{A_i}(n)^2 \middle/ \sum_{n=1}^N S(n)^2 \right] \times 100\% \quad (4.11)$$

Table 4-3 presents the SE values for all 323 level 3 synthesized $S_{A_3}[n]$ load profiles. It shows that only 3 of the $S_{A_3}(n)$ load profiles have a total energy less than 80% of their original load profile total energy. Seventeen of the $S_{A_3}(n)$ load profiles have a total energy that is between 80 to 90% of their original load profile total energy. The majority of the load profiles, 303, have an $S_{A_3}(n)$ that captures 90% or more of their original AMI data load profile total energy.

Table 4-3: $S_{A_3}(n)$ vs. Original Load Profile Energy

Number of $S_{A_3}(n)$ Load Profiles	SE
3 load profiles (0.93%)	Less than 80.000%
17 load profiles (5.26%)	80.000% to 90.000%
303 load profiles (93.81%)	90.000% to 99.998%

AMI load data contains sharp load changes that are similar to the spikes previously discussed for figure 8, and the energy contained in these spikes is modeled with the detail coefficients and not the approximation coefficients of $S_{A_3}(n)$. The energy contained in these spikes is neglected when just $S_{A_3}(n)$ is used to synthesize the load profile. Thus, the energy errors shown in Table 4-3 are an indication of the variability of the loads, where three of the loads (i.e., those with an energy match less than 80%) have a much higher variability than the other 320 loads.

The second parameter to be employed in evaluating the performance of the wavelet load model is the hourly error, $E(n)$, which is defined as the difference between the synthesized load profile value and the original load profile value at each respective time point n, as shown in equation 4.12:

$$E(n) = |S_{A_3}(n) - S(n)| \quad (4.12)$$

Figure 4-12 presents the histogram for $E(n)$, where the distribution fit is shown by the red line. These plots show that the distribution of $E(n)$ is approximately normal with a mean that is centered close to zero with a small variance. This indicates that the error between the wavelet based load model and the original load model is small and predictable.

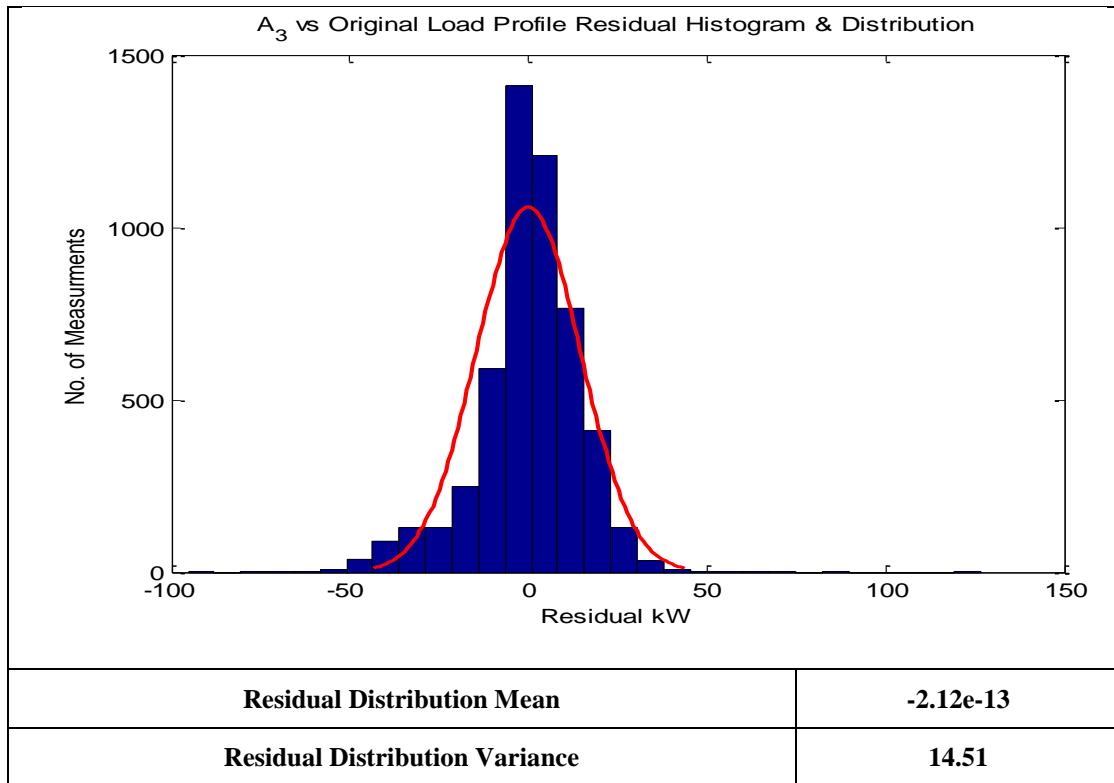


Figure 4-11: Error of Synthesized Load Profile

The third parameter to be employed in evaluating the performance of the wavelet load model is the hourly percentage error, $PE(n)$ which is defined in equation 4.13.

$$PE(n) = \left| \frac{S_{A_3}(n) - S(n)}{S(n)} \right| \times 100\% \quad (4.13)$$

$PE(n)$ can be used to indicate how close a synthesized load profiles relates to its original load profile. Figure 4-13 presents the histogram for the level 3 wavelet load model's hourly percentage error. The result shows that 74% of the $S_{A_3}(n)$ hourly loads are within 10% of their original AMI load measurements.

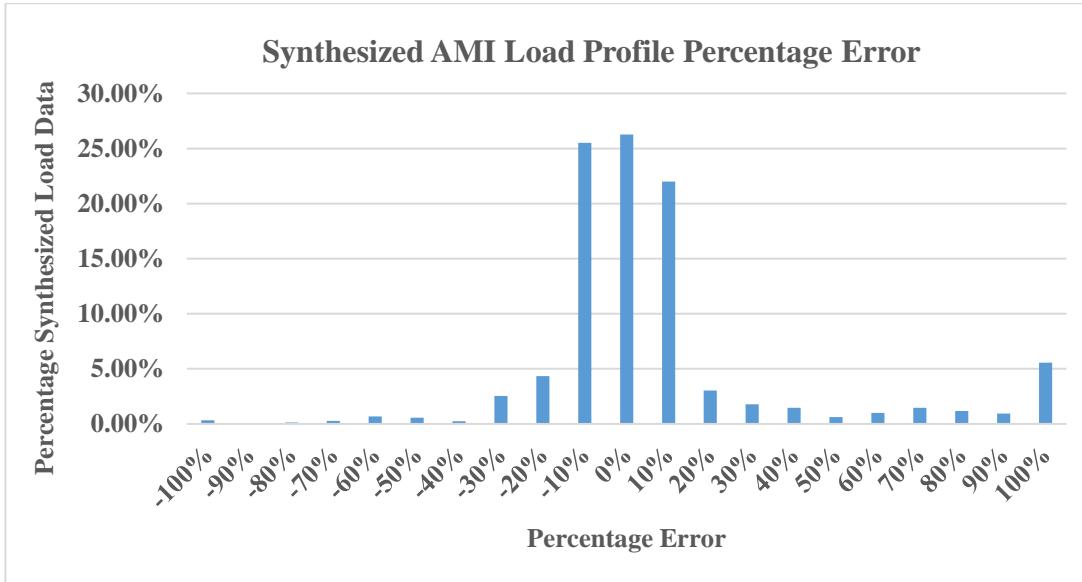


Figure 4-12: Synthesized Load Profile Percentage Error

4.2.2 Classified Wavelet Load Model

The second proposed load data model utilizes the Typical Load Profile (TLP) of the normalized load class to model the AMI load data. The individual load profiles are grouped together using their wavelet coefficients. The AMI load data used in the research here comes from various types of customers. Conventionally, the utility categorizes customers by rates instead of load patterns. In order to classify loads with patterns, the 2D DWT load profile pattern model is used.

Thus, using our AMI load data set, each of the 323 load profiles is normalized by its annual peak and converted into a matrix form where each row of the matrix represents a week of hourly load measurements: from Monday 00 hour to Sunday 23rd hour, which has 168 hourly readings. Thus, a 31-week hourly AMI load profile is represented as a matrix with 31rows and168 columns.

The matrix representation of the load profiles can be visualized as a set of color images, as illustrated in the second column of Figure 4-14. The plots in the first column represent the original, normalized AMI load profiles. The third column is the level 3 approximation coefficient image, which is a lower resolution of the images of the second column. The fourth and fifth columns represent the visual images of the Vertical Coefficients of 2D DWT at levels 3 and 4, respectively. All visual images in Figure 4-14 share the same color map.

The edges of a significant load pattern in a load profile normally associate with rapid load rises and/or drops. This is why some research suggests using terms such as Rise Time and Fall Time [2] to describe load patterns in the time domain. These sudden changes in load profiles can be detected using the 2D DWT coefficients. For example, the first load profile of Figure 4-14 is relatively consistent without significant hourly load variations. Its level 3 Vertical Coefficient, V_3 shows a very consistent pattern without significant color variation in column 3.

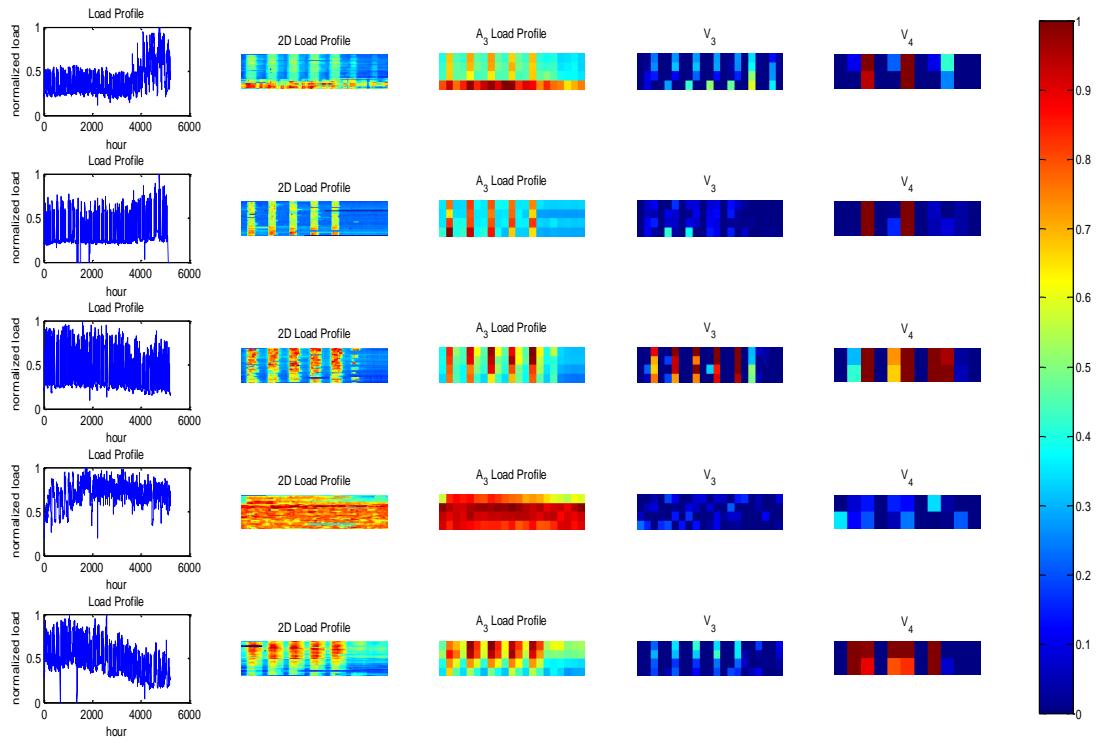


Figure 4-13: 5 AMI load profiles' pattern visualization

The other four load profiles in Figure 4-14 all show significant daily pattern variations. The significant daily pattern differences in load profiles is normally correlated with the changes in the vertical coefficients, as may be seen from Figure 4-10. There are 7 very distinctive color bands in the level 3 vertical coefficients image. The color differences of these vertical bands indicate the difference between their daily load profiles.

The 2D DWT can generate multiple representations of the original data at different resolution levels. The fifth column in figure 9 contains the level 4 vertical coefficients, V4. This can be viewed as a “zoom-out” of V3 at a lower resolution, but still provides the significant patterns of the original load profile. Instead of using all 5,208 hourly load data measurements, the more compact form of the DWT representations of the load pattern is used in the classification process. The V3 matrix size is 4 by 21 and the V4 matrix size is 2 by 11.

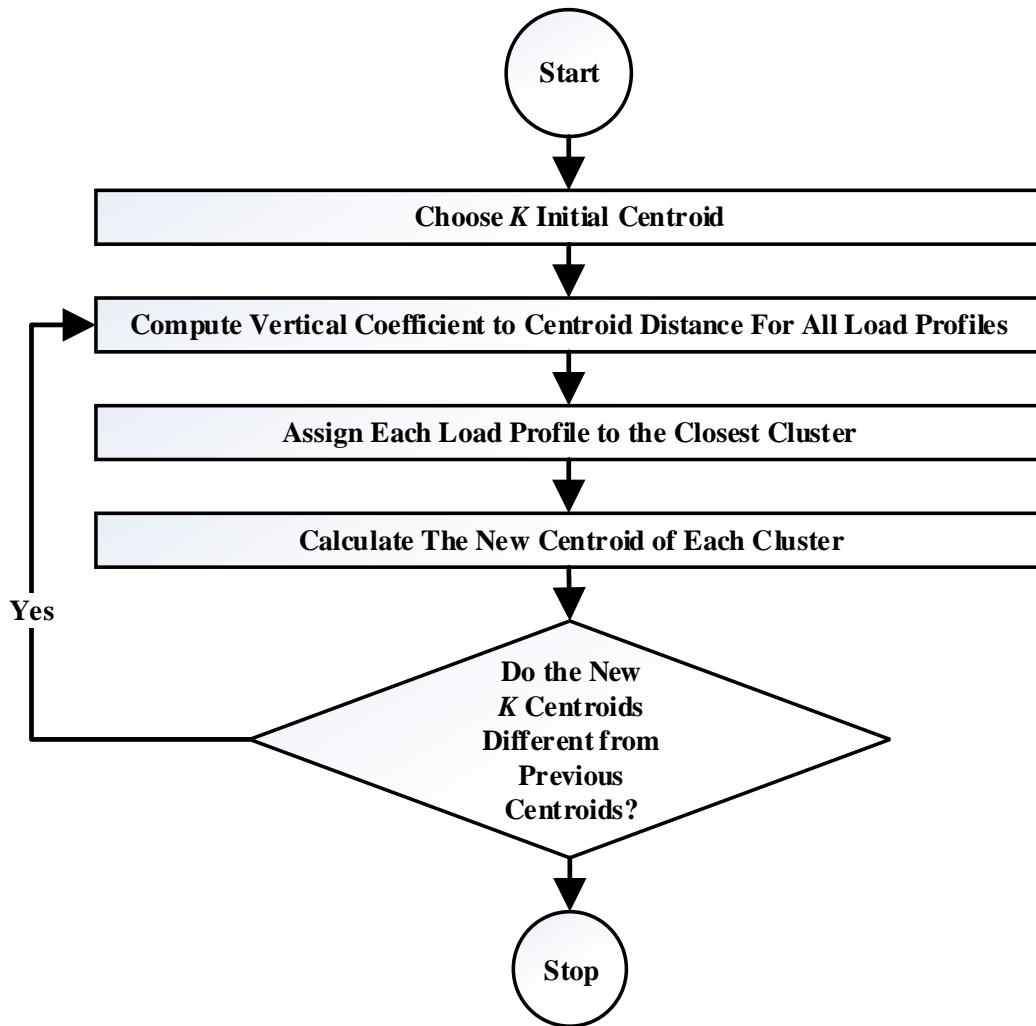


Figure 4-14: K-Mean Clustering Algorithm

This classified load model uses a K-mean clustering algorithm [51] illustrated in Figure 4-15. The algorithm minimizes the mean of the Squared Euclidean distance between the load profile's Vertical Coefficients and the load class k centroid (equation 4.14) for each class, as indicated by the following equation.

$$D_{V_i, C_k} = \sqrt{\sum_{n=1}^N (V_{i,n} - C_{k,n})^2} \quad (4.14)$$

where i is the load profile index

k is the load class index

n is the data index

The K-mean algorithm is used to group individual customer load profiles into k number of load classes (k has to be determined by the user prior to the clustering process). The class k's average hourly load profile $TLP_k(t)$ can be calculated from the load profiles of the class using equation 4.15:

$$TLP_k(t) = \frac{\sum_{n=0}^N S_{k,i}(t)}{N} \quad (4.15)$$

where k is the load class index,

i is the index for a load profile in load class k.

Combining the individual load profile annual peak load $P_{k,i}$ (equation 4.16), average load scale factor $M_{k,i}$ (equation 4.17) ,and the TLP of the normalized load profile of the class, $TLP_k(t)$, the method can construct a synthesized load profile, $S'_{k,i}(t)$, for each customer (equation 4.18) to replace the original AMI data as given by

$$P_{k,i} = \max_{t \in [0,T]} S_{k,i}(t) \quad (4.16)$$

$$M_{k,i} = \frac{\sum_{t=0}^T TLP_k(t) \times P_{k,i}}{\sum_{t=0}^T S_{k,i}(t)} \quad (4.17)$$

$$S'_{k,i}(t) = \frac{TLP_k(t) \times P_{k,i}}{M_{k,i}} \quad (4.18)$$

where k is the load class index

i is the load profile index in load class k

T is the largest time point index

Table 4-4 presents the 5 load profile classes and their sizes. This load model can reduce the size of the original AMI load data set even further, as the model only stores two measurements ($P_{k,i}$ and $M_{k,i}$) for each customer and k number of TLP data sets, $TLP_k(t)$, for the load classes.

Table 4-4: Load profile classification

classes id	Number of Load Profiles
1	89
2	166
3	39
4	14
5	15

For the classified wavelet load model, 234 (72% of 323) of the synthesized load profiles have a total energy value from $E(n)$ (equation 4.12) that is within 10% of their original AMI load profile total energy. The other 28% of the synthesized load profiles have 60% to 80% of the original load profile energy.

Table 4-5 shows several ranges of the classified wavelet load model's percentage error values, $PE(n)$ (equation 4.13), and the percentage of all classified wavelet load model data that is within each respective $PE(n)$ range.

Table 4-5: Synthesized Load Profile Evaluations with $PE(n)$

$PE(n)$	% synthesized load profiles
10% Less	22%
10% to 20%	24%
20% to 30%	43%
over 30%	11%

The classified wavelet load model is not as accurate as the wavelet load model because the granularity of the classes is not as fine-grained in terms of their patterns. The goal of this chapter is to show that the AMI load profile coarse-resolution representation can be used in load profile classification. It should be noted that the example given here has not been optimized. Given some optimization criteria, such as energy ranges, peak values, and others, and using the classification technique presented here, a set of classes could be sought to meet the criteria.

4.3 Load Estimation with Wavelet Based Load Model

In this section three load models for the 323 customers are used with a model of a real-world circuit: 1) the original AMI load data, which is referred to as the reference data model, 2) the wavelet load model with level 3 approximation coefficients and 3) the classified wavelet load model. Three hourly load estimations are performed at the beginning of the circuit for a period of 5,028 hours from 5/27/2013 00:00 to 12/29/2013 13:00.

The wavelet load model stores 651 level 3 wavelet approximation coefficients for each of the 323 customers. The classified wavelet load model stores 2 values for each customer plus 5 different hourly TLPs (each has 5,028 data points). Table 4-6 presents the data reduction that is achieved by each wavelet-based load model. Both data models can reduce the AMI data size by over 87%.

Table 4-6: Data Reduction

	No. Records	AMI Records	Reduction
Wavelet Load Model	210,273	1,682,184	-87.50%
Wavelet Classified Load Model	26,686	1,682,184	-98.41%

The second comparison uses the hourly estimation percentage error histogram (Figure 4-16) to show how the proposed wavelet-based load models match up with the original AMI data model in terms of hourly load estimates. In this case 71% of the hourly load estimates from the wavelet load model are within 10% of the loads from the original AMI data model. With the classified wavelet load model, 80% of the load estimates are within 10% of the loads from the original AMI data.

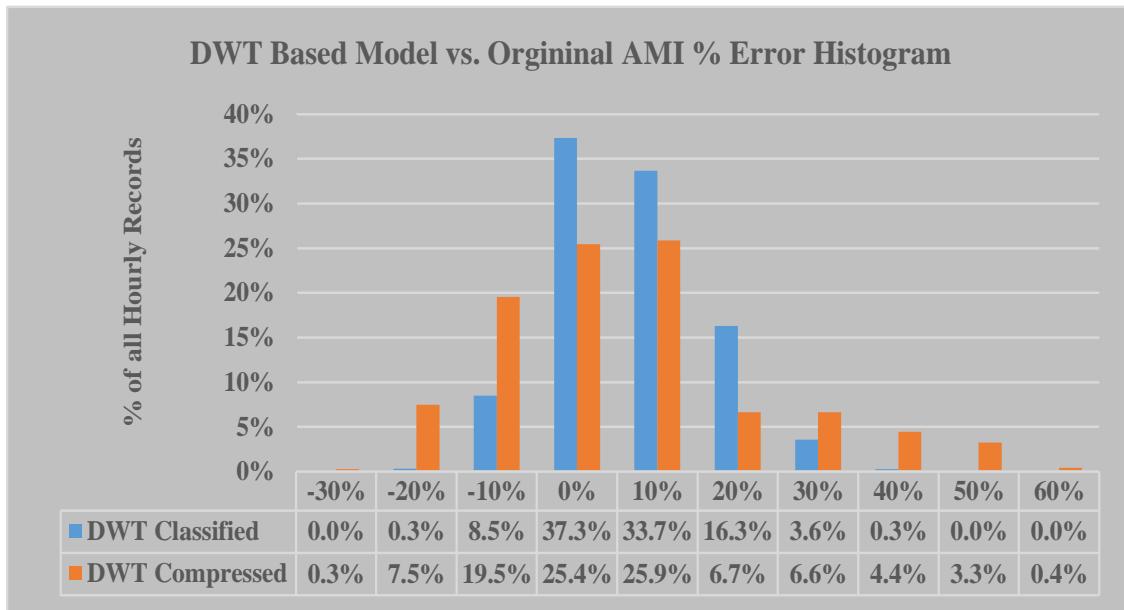


Figure 4-15: Hourly Load Estimation % Error Histogram

Table 4-7 presents the system peak estimation comparison: the original AMI data indicates the peak load time occurred at 5PM on 7/18/2013 for A and C phases, and at 4PM on 7/18/2013 for phase B. Both wavelet-based load model estimates the peak load occurred at 5PM on 7/18/2013 for all phases. There is a 1 hour difference in the phase B peak load estimate. The peak load estimates from the wavelet-based load models are within 12% of the peak estimates using the original AMI load data.

Table 4-7: System Peak Load Estimation

Phase	AMI Peak (kW)	Peak Time	Wavelet Load Model Peak (kW)	Peak Time	Error %
A	878.20	7/18 5 PM	968.71	7/18 5 PM	10.3%
B	1132.82	7/18 4 PM	1262.76	7/18 5 PM	11.5%
C	569.73	7/18 5 PM	629.88	7/18 5 PM	10.6%
Phase	AMI Peak (kW)	Classified Wavelet Load Model Peak (kW)			
A	878.20	868.75		7/18 5 PM	
B	1132.82	1010.68		7/18 5 PM	
C	569.73	552.38		7/18 5 PM	

Table 4-8 shows that the total energy estimates from the wavelet-based load models are within 1% of the energy of the original AMI data

Table 4-8: System Total Energy Estimation

Phase	AMI (kWH)	Wavelet Load Model (kWH)	% Error
A	2013192.71	2011098.14	-0.10%
B	2321937.76	2321775.12	-0.01%
C	1306861.05	1306961.94	0.01%
	Classified Wavelet Load Model		
A	2013192.71	2007990.47	-0.26%
B	2321937.76	2322710.13	0.03%
C	1306861.05	1304125.65	-0.21%

4.4 Summary

The major challenge to using big AMI data in large scale system power flow analysis is the size of the data. The proposed two wavelet-based load data models can reduce the size of the AMI load data by 88% (multi-resolution wavelet load model) and 98% (classified wavelet load model), respectively.

Individual customer load profiles can be synthesized by using the lower resolution coefficients of its multi-resolution DWT. This is because most of the energy of the original AMI load data profile is located in the lower frequency band. In the multi-resolution DWT the energy of the lower frequency bands can be captured by a set of approximation coefficients at level k , with a data size that is of the original AMI data length.

It has been shown that the wavelet representation of sudden/abnormal load changes (i.e., load changes lasting for a short period, such as occur from metering errors) can be detected in the higher frequency band of the original AMI load profile. By using the lower resolution wavelet approximation coefficients to model the load profile, these abnormal load changes are filtered out.

A matrix derived from a 2-dimensional DWT has been used to represent individual load profiles. The vertical coefficients from the matrix can be used to detect periodic load patterns, such as daily or weekly patterns. A data mining algorithm and the vertical coefficients can be used to classify customer load profiles into groups that have similar load behavior.

The load class hourly Typical Load Profile (TLP) identified by the DWT vertical coefficients can be used to model the load pattern of an individual customer belonging to the class. Using the load class TLP along with the individual customer's peak and average load measurements, a synthesized load profile can be constructed for each customer.

For the example circuit considered, power flow time series analysis was used to compare the total energy from the three models. The total energy calculated using the wavelet load model was within 0.1% of the original AMI load data, and the total energy calculated using the classified wavelet load model was within 0.26% of the original AMI data. The peak load estimates from the wavelet load models were within 11% of the peak estimates from the AMI data, where the wavelet load model is conservative, overestimating the peak.

Wavelet-based load models have the potential to be a standard way to process and store AMI load data to be used in analysis. The wavelet models provide a new approach to improving the classification of customers into groups that display similar load behavior.

Chapter 5: Stochastic Load Model in Frequency Domain

In order to efficiently and reliably operate a smart grid an accurate and efficient load estimation model is needed. Load estimation models can generally be categorized into two groups: 1) Traditional time series and statistical methods, such as regression and ARMA models [2-6, 8, 24-26, 52], and fuzzy approaches [9-11]. 2) Machine learning methods involving neural network and support vector machine [12-13, 24, 26, 52-53].

The majority of load modeling literature focuses on short (from minutes to several days) and long term (from a year to a decade) load estimation at the system or substation level, where the load generally ranges from several MW to several GW [24, 26, 52]. These conventional load models are generally constructed using time-domain load data. In the time-domain it is difficult to describe and quantify the diverse and complicated load pattern characteristics that exists in AMI load data [27, 54].

The load patterns for distribution transformers and feeders are diverse and strongly influenced by local weather conditions. For these load types, there is limited research on developing medium term (from a day to a year) load estimation models using statistics from spectral analysis of hourly AMI load data and the corresponding weather data.

In this chapter, the proposed load model is constructed using frequency-domain statistics derived from AMI load data spectral analysis [27]. The statistics are used to estimate the hourly load profile for transformers and feeders under different weather conditions. It has the following original features:

1. Load model is driven by non-temporal parameters.

2. The model is flexible and can incorporate many partitioning parameters (weather, customer class, spatial...).
3. Load Estimation Envelope.
 - a. Detect and filter error/outlier electric load data.
 - b. Equipment Sizing.
4. Load Data Interpolation, back-cast or filling data gap of various size with known weather data.
5. Load Data forecast with forecasted weather data.

5.1 Frequency Domain Stochastic Load model

The load model investigated in this chapter is referred to as the Frequency Domain Weather Sensitive Stochastic (FDWSSM) load model. The FDWSSM is derived from AMI load and weather data collected for a set of transformers and feeders. The model construction process consists of three major steps, which are presented in the following three subsections.

A. AMI Data Partition by Daily Weather Index

In the US, Heating-Cooling Degrees (HCDD) [18] is a commonly used measure to evaluate a utility customer's energy usage dependency on the average daily temperature $\overline{\text{Temp}}_D$. HCD_D is defined in equation 2, where the temperature is in Fahrenheit degree ($^{\circ}\text{F}$).

$$\overline{\text{Temp}}_D = \frac{\text{Max}(\text{Temp}_D(n)) + \text{Min}(\text{Temp}_D(n))}{2} \quad (1)$$

$$HCD_D = \overline{\text{Temp}}_D - 65^{\circ}\text{F} \quad (2)$$

where subscript D is the day index, hour index $n \in [0, 23]$

For the majority of customers, the largest portion of their electricity usage is from the heating, ventilating, and air conditioning (HVAC) equipment, whose operation is sensitive to the HCD_D value. In order to model the weather sensitivity of daily load profiles and calculate the frequency domain statistics of a set of loads under certain weather conditions, a HCD_D based Weather Index (WI in equation 3) is used to partition the daily AMI load data into sets.

$$WI = \text{Round}(HCD_D / 10.0) \times 10 \quad (3)$$

where $\text{Round}(x)$ rounds value x to the nearest integer

subscript D is the current day index

WI represents a range of HCD_D values. In this paper, each WI has a 10 HCD_D degree range.

Using this degree range results in all AMI data sets used in the statistical analysis having at least

30 daily samples. In this paper, this minimal sample size is due to the limit of the current AMI data coverage. The WI is an integer number used as a data index in the load database to categorize a set of days whose daily heating-cooling degree values are within the WI's range.

Table 5-1 lists the values of WI used in this paper and the corresponding daily average temperature (\overline{Temp}_D) range in both $^{\circ}\text{C}$ and $^{\circ}\text{F}$.

		\overline{Temp}_D		\overline{Temp}_D	Range for WI
WI		$^{\circ}\text{F}$		$^{\circ}\text{C}$	
-40	20	30	-7	-1	
-30	30	40	-1	4	
-20	40	50	4	10	
-10	50	60	10	16	
0	60	70	16	21	
10	70	80	21	27	
20	80	90	27	32	

For different WI, the daily load profiles of a particular transformer/feeder exhibit different load pattern characteristics. The AMI data used in this paper were collected in an area where the majority of the electrical load is from residential customers and small-to-medium sized commercial customers. The weather condition has a strong influence on the daily load profiles of the transformers and feeders that serve these customer types.

Figure 5-1 presents a set of normalized daily load profile (by the daily peak) plots for a transformer supplying commercial customers (first, second plots), and a transformer supplying residential customers (third, fourth plots). Each line in the plot represents a daily profile with the same WI value. There are two WI values in the sample plots: the winter WI = -20 is between 40°F and 50°F, and summer WI = 10 is between 70°F and 80°F. The transformer that supplies the

commercial customers has a more consistent single-peak, daily load pattern for both WI values. During winter the transformer's hourly load variation is smaller than the hourly load variation observed during the summer. The transformer that supplies residential customers has different load patterns for winter and summer. In the winter the transformer exhibits a double-peak pattern (one peak occurs in the morning and the other peak occurs in the evening). In the summer the transformer's daily peak region moves to a later hour of the day.

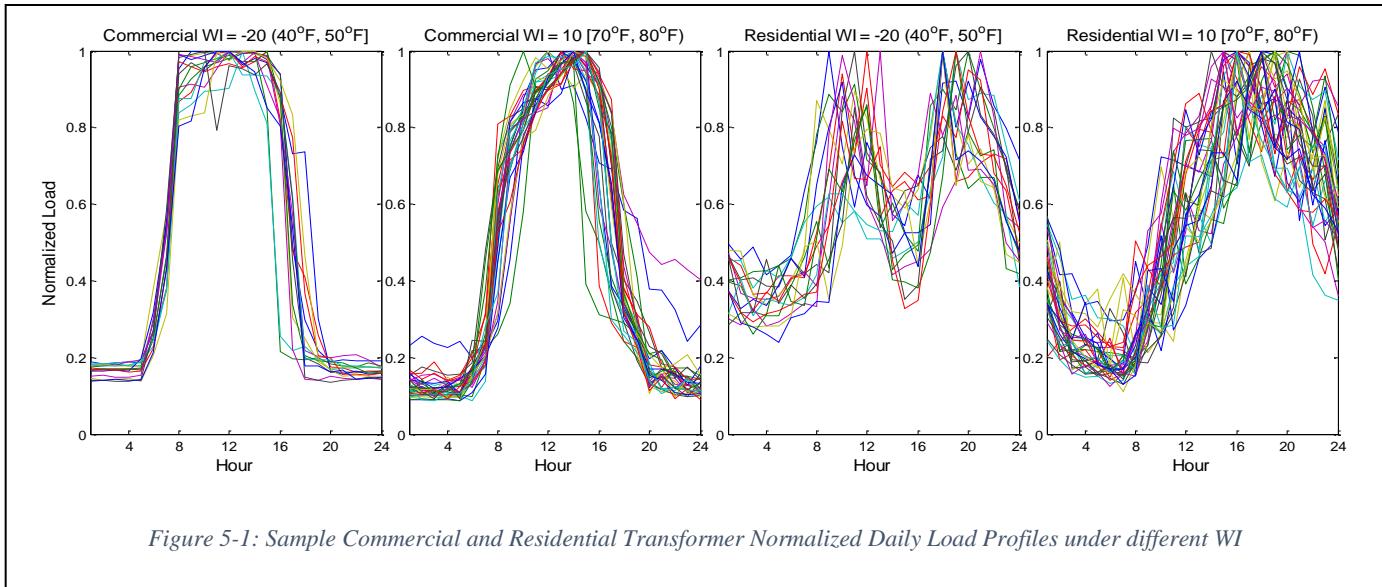


Figure 5-1: Sample Commercial and Residential Transformer Normalized Daily Load Profiles under different WI

B. AMI Load Data Frequency Domain Representation

Time domain AMI daily load profiles are expressed as a set of time-series load data points.

These data points have an auto-correlated relationship [55] which lacks orthogonality. Because of the lack of orthogonality, the time domain hourly data is not suitable to be modeled and analyzed independently. Because of this it becomes more difficult to model stochastic load pattern characteristics using time domain load profile representations [28].

Using the technique introduced in [27], the original AMI load profile can be transformed to a set of independent, orthogonal, frequency domain components (defined by a magnitude value and a phase value) that can be modeled and analyzed independently. Once the daily AMI load data is partitioned by WI, the frequency domain statistics can be extracted and the stochastic models constructed for each of the AMI daily load profile frequency domain components.

From [17] the daily AMI load profile $x_m(t)$ is normalized by its daily peak x_{peak} as

$$x_n(t) = x_m(t) / x_{peak} \quad (3)$$

The normalized load curve $x_n(t)$ is then transformed into a frequency domain representation $X(f)$ using the Discrete Fourier Transform (DFT) as

$$X(f) = \sum_{k=0}^{23} X_k \delta(f - k\Delta f) \quad (4)$$

X_k is the frequency spectrum indexed by k,

Δf is the frequency resolution

Each member of the frequency spectrum, X_k , is expressed by equation 5. X_k is a complex number with magnitude a_k and phase angle θ_k .

$$X_k = \sum_{n=0}^{23} x_n e^{-\frac{j2\pi kn}{24}} \quad (5)$$

Using the Inverse Discrete Fourier Transform (IDFT) the time domain load profile $x^r(t)$ can be reconstructed using some or all of the frequency components as

$$x^r(t) = \sum_{k=0}^{23} X_k^r \delta(t - n\Delta t) \quad (6)$$

$$x_n^r(t) = \frac{1}{24} \sum_{k=0}^{23} X_k e^{\frac{j2\pi kn}{24}}, \quad n = 0, 1, 2, \dots, 23 \quad (7)$$

Using Euler's identity to combine exponentials into cosine functions, $x^r(t)$ can be expressed as a sum of frequency components as

$$x^r(t) = x_{DC}(t) + x_{1H}(t) + \dots + x_{kH}(t) + \dots + x_{11}(t) \quad (8)$$

$x_{DC}(t) = a_{DC} = \frac{a_0 + a_{12}}{24}$ is the DC component

$x_{kH}(t) = \frac{a_k}{12} \cos\left(\frac{\pi n \Delta t}{12} + \theta_k\right)$ is the k th harmonic

with amplitude $a_{kH} = \frac{a_k}{12}$ and phase angle θ_k

Now, the original time domain daily load profile data can be represented by the frequency domain daily load profile descriptor, S , which is given by

$$S = [x_{peak}, (a_{1H}, \theta_{1H}), \dots, (a_{kH}, \theta_{kH})] \quad (9)$$

S is defined as a set of frequency domain parameters which consist of the daily peak x_{peak} and 12 set of harmonic components (magnitude value a_{kH} and phase value θ_{kH} indexed by k).

C. Frequency Domain Weather Dependent Stochastic Load Model

Figure 5-1 illustrates that a transformer's daily load profiles under the same WI can share a similar pattern with certain levels of variations in both the vertical direction (along the load magnitude axis) and the horizontal direction (the time axis). It is difficult to model these directional variations independently in the time domain. However, in the frequency domain once the daily load profile is transformed into a series of harmonics, these directional variations can be modeled independently by harmonic magnitude a_{kH} (vertical variation) and phase θ_{kH} (horizontal variation) [27].

As discussed in section B, each daily load profile can be represented in the frequency domain using S defined in equation 9. For each WI, there is a set of daily load profiles whose S parameter variations which can be modeled as a stochastic distribution. The null hypotheses (H0) tests are used here to test if the daily load peak x_{peak} , the harmonic magnitude a_{kH} , and the phase angle θ_{kH} are from normal distributions, respectively.

Table 5-2 presents the H0 test results for x_{peak} , a_{kH} and θ_{kH} . For x_{peak} , a_{kH} , the Chi-Square Test is used. For θ_{kH} the Waston's U2 [34] test is used because of its circular nature. It

Normal Distribution Test Method		Chi-square				Watson's U2	
Load Model Data Set		x_{peak}		a_{kH}		θ_{kH}	
Null Hypothesis (H0) Test Result		Accept	Reject	Accept	Reject	Accept	Reject
WI	-40	82%	18%	80%	20%	85%	15%
	-30	93%	7%	93%	7%	91%	9%
	-20	92%	8%	95%	5%	96%	4%
	-10	96%	4%	96%	4%	97%	3%
	0	100%	0%	100%	0%	100%	0%
	10	100%	0%	100%	0%	100%	0%
	20	95%	5%	95%	5%	94%	6%

should be noted that the number of records for the case WI = -40 ranged from 5 to 20, and hence represented a small number of samples. For the other weather indices, the results of the H0 tests indicate that it is reasonable to consider x_{peak} , a_{kH} and θ_{kH} as normally distributed.

The proposed frequency domain load model consists of a normalized, stochastic, daily peak, load profile model and a frequency domain shape descriptor, stochastic model. Using the sample residential load profiles illustrated in figure 1 as an example, the daily peak values under different

Table 5-3: Residential Transformer Daily Peak Partitioned by WI		
WI	Daily Peak Mean (kW)	Daily Peak Standard Deviation (kW)
-40	12.145	1.284
-30	13.109	1.851
-20	12.435	1.022
-10	12.097	1.934
0	14.719	1.427
10	27.329	2.988
20	39.630	3.792

WI can be modeled as normal distributions using the mean and standard deviation values shown in Table 5-3.

For a given WI, the normalized, daily load profile's frequency domain shape descriptor components, defined in equations 10 and 11, are $x_{WI}^r(t)$, and may be represented as a set of stochastic distributions with magnitudes $N_{WI}(a_{kH})$ and phase angles $VM_{WI}(\theta_{kH})$. The phase angle values are always within a circular range (0 to 2π). Here the $VM_{WI}(\theta_{kH})$ is modeled as a circular, normal random number function based on the Von Mises distribution [34].

$$x_{WI}^r(n) = x_{DC}^{WI}(n) + x_{1H}^{WI}(n) + \dots + x_{kH}^{WI}(n) \quad (10)$$

$$x_{kH}^{WI}(n) = \frac{N_{WI}(a_{kH})}{12} \cos\left(\frac{p\pi n \Delta t}{12} + VM_{WI}(\theta_{kH})\right) \quad (11)$$

where WI is the weather index for the daily load profile
 harmonic index $k \in [1, 13]$
 hour index $n \in [1, 24]$
 Δt is the sampling time
 a_{kH}, θ_{kH} are k th harmonic magnitude and phase angle
 $N_{WI}(a_{kH})$ is normal random number function with
 k th harmonic's magnitudes mean $\mu(a_{kH})$ and
 standard deviation $\sigma(a_{kH})$ under WI
 $VM_{WI}(\theta_{kH})$ is *Von Mises* random number function
 with k th harmonic's mean direction $\mu(a_{kH})$ and
 concentration parameter $k(\theta_{kH})$ under WI

Table V-4 lists the coefficient values in the functions $N_{WI}(a_{kH})$ and $VM_{WI}(\theta_{kH})$ for modeling a sample transformer's daily 24-hour AMI data (plotted in Figure 1), partitioned by the WI values. The magnitude values of lower frequency components (from DC to 5th Harmonic) are much larger than the higher frequency components. This indicates that most of the energy of the transformer's load profile is in the lower frequency components. The overall pattern of the load profile is a slowly changing trend, and the hour-to-hour variation is relatively small.

Table 5-4: Residential Transformer Frequency Domain Daily Normalized Load Profile Coefficients Partitioned by WI

WI	DC	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12
Harmonic Magnitude Mean $\mu(a_p)$													
-40	0.641	0.180	0.144	0.046	0.071	0.068	0.042	0.022	0.028	0.029	0.019	0.014	0.007
-30	0.585	0.198	0.137	0.065	0.056	0.043	0.031	0.030	0.029	0.026	0.024	0.029	0.010
-20	0.592	0.178	0.137	0.067	0.045	0.047	0.031	0.037	0.035	0.030	0.024	0.024	0.014
-10	0.596	0.204	0.126	0.054	0.050	0.037	0.039	0.039	0.037	0.034	0.031	0.028	0.013
0	0.558	0.223	0.111	0.059	0.060	0.054	0.042	0.038	0.037	0.033	0.034	0.034	0.017
10	0.548	0.350	0.057	0.050	0.053	0.038	0.039	0.026	0.029	0.024	0.025	0.017	0.010
20	0.605	0.371	0.050	0.023	0.037	0.021	0.017	0.019	0.014	0.015	0.012	0.008	0.008
Harmonic Magnitude Standard Deviation $\sigma(a_p)$													
-40	0.018	0.134	0.021	0.014	0.024	0.008	0.013	0.009	0.013	0.015	0.011	0.015	0.003
-30	0.037	0.055	0.038	0.028	0.026	0.017	0.020	0.015	0.018	0.011	0.020	0.014	0.007
-20	0.061	0.053	0.046	0.032	0.024	0.021	0.018	0.021	0.015	0.016	0.013	0.015	0.010
-10	0.061	0.070	0.049	0.021	0.023	0.024	0.019	0.025	0.021	0.014	0.015	0.015	0.010
0	0.066	0.077	0.049	0.033	0.026	0.028	0.023	0.021	0.021	0.017	0.018	0.018	0.011
10	0.045	0.044	0.026	0.025	0.022	0.021	0.021	0.014	0.021	0.016	0.011	0.010	0.007
20	0.025	0.025	0.025	0.003	0.013	0.019	0.014	0.009	0.010	0.012	0.010	0.007	0.006
Harmonic Phase Mean Direction (Radian) $\mu(a_p)$													
-40	0.000	1.701	1.847	-1.443	2.491	2.674	-1.060	-2.526	2.607	-0.347	2.561	-0.146	0.000
-30	0.000	2.046	2.037	-1.913	1.997	-2.963	-1.447	-0.991	1.824	-1.243	0.426	-0.411	0.000
-20	0.000	2.118	1.987	-1.637	1.896	2.578	-1.744	-0.356	2.040	-2.343	1.119	-2.351	0.000
-10	0.000	2.205	1.876	-1.895	2.423	1.916	-1.248	0.896	0.120	-2.409	-0.205	-1.165	0.000
0	0.000	1.818	1.375	1.980	2.523	2.040	3.108	-2.376	2.226	-2.805	-0.978	-2.745	0.000
10	0.000	1.913	0.717	1.791	2.137	1.759	2.153	-2.596	-2.914	-2.420	-1.778	-0.648	0.000
20	0.000	2.063	0.393	2.551	2.667	1.687	1.961	-2.972	-1.619	-2.803	-1.966	-1.253	0.000
Harmonic Phase Concentration Parameter $k(\theta_p)$													
-40	0.000	1.175	1.655	0.730	0.778	-1.062	1.880	1.479	0.925	1.299	0.415	0.883	0.000
-30	0.000	0.475	0.081	1.588	1.189	1.270	1.276	0.244	1.096	0.690	0.337	0.758	0.000
-20	0.000	1.003	1.439	1.005	0.954	0.987	0.727	0.120	0.663	0.816	0.373	0.120	0.000
-10	0.000	-1.457	-1.503	0.526	1.044	0.448	0.299	0.261	0.378	0.397	0.757	0.450	0.000
0	0.000	0.714	1.355	0.259	0.873	0.557	0.255	0.464	0.270	0.272	0.198	0.270	0.000
10	0.000	-1.189	0.880	0.266	1.139	0.464	0.559	0.386	0.431	0.297	0.876	0.132	0.000
20	0.000	1.072	-0.251	0.795	-1.981	1.567	0.771	0.306	1.052	0.743	0.480	-1.099	0.000

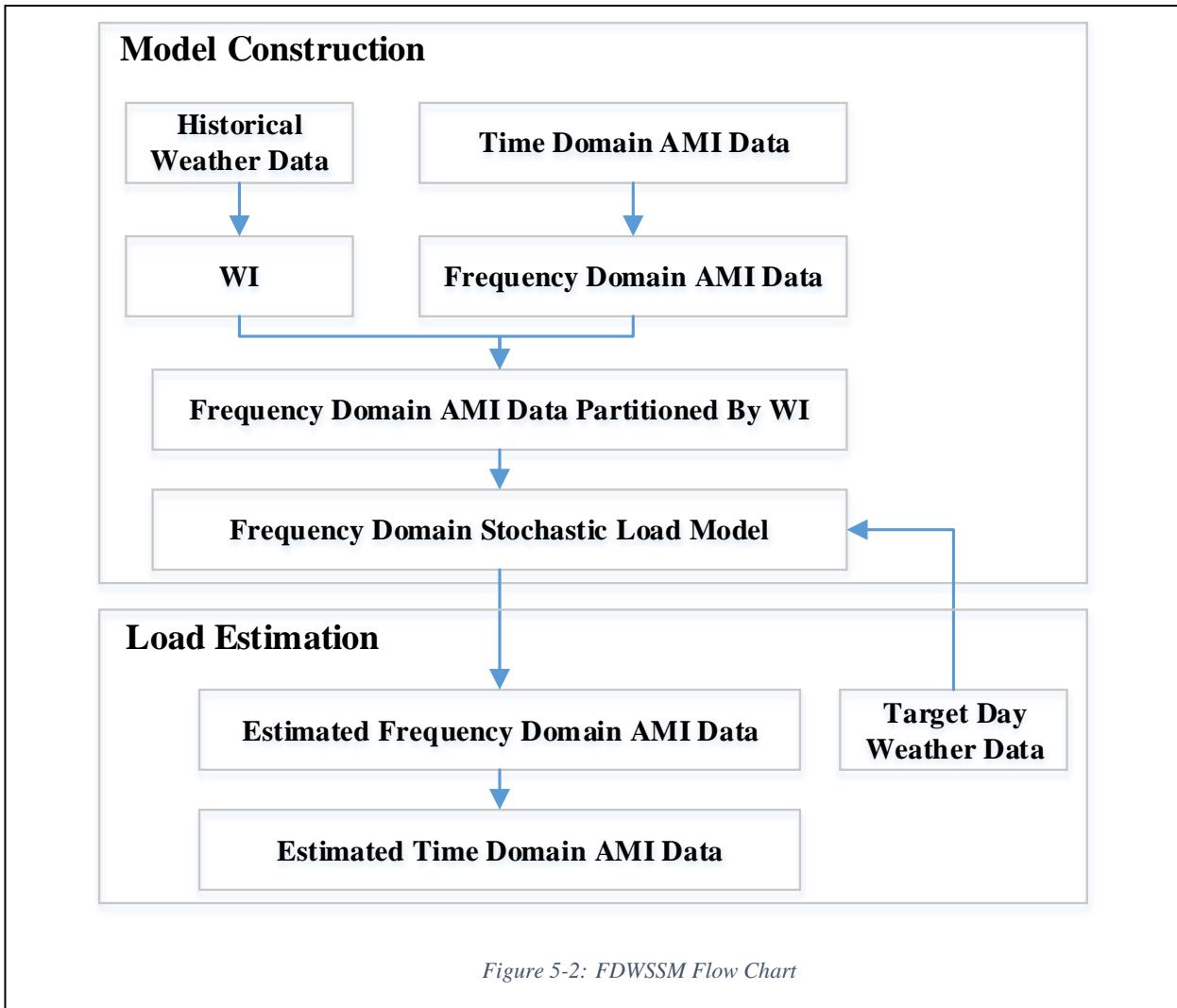
The data of Table 5-4 also shows that the variations among the phase angle values are small. This indicates that the overall shape of the load profile doesn't shift that much along the timeline under the same WI.

The proposed FDWSSM is a compact model for the AMI load data. The load model for each customer requires 416 coefficients (for 13 sets of 4 harmonics coefficients and for 8 different WI values) for the frequency domain hourly load model. The size of FDWSSM set is approximately 4% of the original AMI data size, where in the original AMI load data sets analyzed there are 8760 hourly AMI data points for each customer. Because most of the load profile energy is located in the lower frequency components there is the potential to further reduce the FDWSSM model size by excluding some of the higher frequency harmonics from the model [27].

One advantage of the FDWSSM model is the scalability. Once the model is constructed it only depends upon the weather condition, and not the time, to perform the estimation or forecasting. As the system accumulates more AMI data and corresponding weather data, the proposed model only needs to update the parameters for the weather conditions corresponding to the new AMI load data, which will not cause the model size to increase. If the behavior of the customer does not change, only the weather conditions that have not been encountered before will give rise to new sets of coefficients. For each new weather condition WI, 52 new parameters are needed in the current implementation of the FDWSSM model.

5.2 Load Estimation Experiment

Figure 5-2 illustrates the FDWSSM model construction and load estimation process. The WI calculation, time-to-frequency domain AMI data transformation, and AMI daily load profile frequency domain statistic calculation (partitioned by WI) are implemented in the model



construction process. Once the model is created, the WI of the target day is incorporated into the estimation process to estimate the load profile.

The hourly AMI load data used here was collected in the suburbs of a major Mid-Atlantic metropolitan area during 2014. In the experiment the data in February and August are reserved for the FDWSSM model performance evaluation tests, and the rest of the 10 months of data are used to construct the FDWSSM.

The data set consists of hourly AMI load data from 420 transformers and 56 feeders (The 420 transformers are not part of the 56 feeders). The types of customers served by these transformers and feeders range from small residential customers to large commercial customers.

The load distributions for different average hourly load ranges are presented in Table 5-5. The data shows a diverse group of electricity loads. The average hourly loads for the transformers range from 0.9kW to 188kW. The average hourly loads for the feeders range from 325 kW to 3849 kW. Twenty-nine feeders have average hourly loads larger than 1,000kW.

Table 5-5: The Number of Feeders/Transformers within Different Average Hourly Load Ranges

Average Hourly Load Range		Number of Transformers	Number of Feeders
Min kW	Max kW		
0	10	66	
10	40	292	
40	80	48	
80	160	14	
300	500		14
500	1000		13
1000	4000		29

The local 2014 hourly weather data used in the experiments was downloaded from National Weather Service on-line weather data repository [56].

After the FDWSSM models were constructed for all transformers and feeders, 200 estimation runs were executed with the models to estimate the 2014 February and August daily 24-hour load values for each feeder and transformer. The hourly average of the 200 estimated hourly load data sets are then compared with the corresponding AMI hourly load data measurements taken during the same time window. This comparison for feeders is shown in Figure 3 for two days in February, the 1st and 3rd, and for two days in August, the 1st and 2nd. In Figure 3 the blue points represent FDWSSM model calculated values and the grey points represent AMI load measurements. Note that the average estimated daily load profiles (blue points) closely resembles the original AMI daily load profiles (grey points). Also shown in Figure 3 are estimation envelopes to be discussed in the next section.

5.3 Load Estimation Envelopes

The estimation envelope is constructed using the hourly maximum value and hourly minimum value of the FDWSSM model estimated loads. Thus, each plot in Figure 5-3 contains estimated hourly load values for the maximum, minimum, and average, and the actual AMI load measurements. For four different days, February 1st and 3rd and August 1st and 2nd, the actual AMI load profile curves are situated within the boundaries of the estimation envelopes, whose boundaries are defined by the maximum forecasted load curve (in orange color) and the minimum forecasted load (in yellow color).



Figure 5-3: FDWSSM Estimation Envelope Plots

The estimation envelope can be used to determine if a load data point is an erroneous reading or an outlier, which can be filtered out. The estimation envelope can also be used to

determine the equipment rating/size. Since this envelope is a detailed hourly data set, the equipment sizing can be evaluated thoroughly at all time points, not just at the peak.

The data in Table 5-6 shows that over 94% of the original AMI hourly load data for the feeders and transformers is contained within the boundaries of the estimation envelopes generated by the FDWSSM load models. These extreme load estimates can be used in applications such as equipment sizing or system planning.

Table 5-6: Estimation Envelope vs. Original AMI Data

	Number of Original AMI Hourly Load in Feb. and Aug. 2014		
	Outside Estimation Envelope	Within Estimation Envelope	Total
Feeder	2,378 (3%)	76,918 (97%)	79,296
Transformer	35,683 (6%)	559,037 (94%)	594,720

5.4 Load Model Performance Evaluation

5.4.1 Model Performance Metrics

Commonly used performance metrics in load modeling are Mean Absolute Percentage Error (MAPE in equation 12), and Coefficient of Variation (CV in equation 13).

$$MAPE = \frac{1}{T} \times \sum_{n=0}^T \left| \frac{x(n) - \hat{x}(n)}{x(n)} \right| \times 100 \% \quad (12)$$

$$CV = \frac{\sqrt{\frac{1}{T} \times \sum_{n=1}^T (x(n) - \hat{x}(n))^2}}{\frac{1}{T} \times \sum_{n=1}^T x(n)} \times 100 \% \quad (13)$$

where T is number of hours, n is the hour index

$x(n)$: the AMI hourly load

$\hat{x}(n)$: the estimated hourly load

In electric load estimation, the time and magnitude of a daily peak is important. Two additional metrics are used in this research to evaluate the FDWSSM model's performance. The first one is the Mean Absolute Percentage Error for the Daily Peak (Daily Peak MAPE in equation 14), which is used to evaluate accuracy of the estimated daily peak.

$$\text{Daily Peak MAPE} = \frac{\sum_{k=1}^D \left| \frac{\text{Max}(x_k(n)) - \text{Max}(\hat{x}_k(n))}{\text{Max}(x_k(n))} \right|}{D} \times 100\% \quad (14)$$

where D is the number of days, k is the day index

$\text{Max}(x_k(n))$: the k th day AMI peak

$\text{Max}(\hat{x}_k(n))$: the k th day estimated peak

The second metric is the Mean Absolute Peak Time Error (MAPTE in equation 15), which is used to evaluate the accuracy of the estimated time at which the daily peak occurs.

$$\text{MAPTE} = \frac{1}{D} \times \sum_{k=1}^D \left| t_{\text{Max}(x_k(n))} - t_{\text{Max}(\hat{x}_k(n))} \right| \quad (15)$$

whereas D is number of days, k is the day index

$t_{\text{Max}(x_k(n))}$: hour index of the k^{th} day AMI peak

$t_{\text{Max}(\hat{x}_k(n))}$: hour index of the k^{th} day estimated peak

The last performance metric used in this research is the Pearson Correlation Coefficient (PCC), defined in equation 16. PCC ranges from -1 to 1, where 1 means total positive correlation and 0 means no correlation. This metric is used to evaluate the strength and direction of a linear relationship between the original AMI hourly load data and the FDWSSM model's estimated hourly load data.

$$PCC = \frac{\sum_{n=0}^T (x(n) - \bar{x}(n)) \cdot (\hat{x}(n) - \bar{\hat{x}}(n))}{\sqrt{\sum_{n=0}^T (x(n) - \bar{x}(n))^2 \times \sum_{n=0}^T (\hat{x}(n) - \bar{\hat{x}}(n))^2}} \quad (16)$$

where $x(n)$ is the original AMI hourly load.

$\hat{x}(n)$ is the estimated hourly load.

5.4.2 Load Model Performance

The estimated hourly load data MAPE and Daily Peak MAPE values for the transformers and feeders are presented in Figures 5-4 plot A to D. Table 5-7 summaries these two performance metric distributions in different percentage error ranges, showing the absolute number of

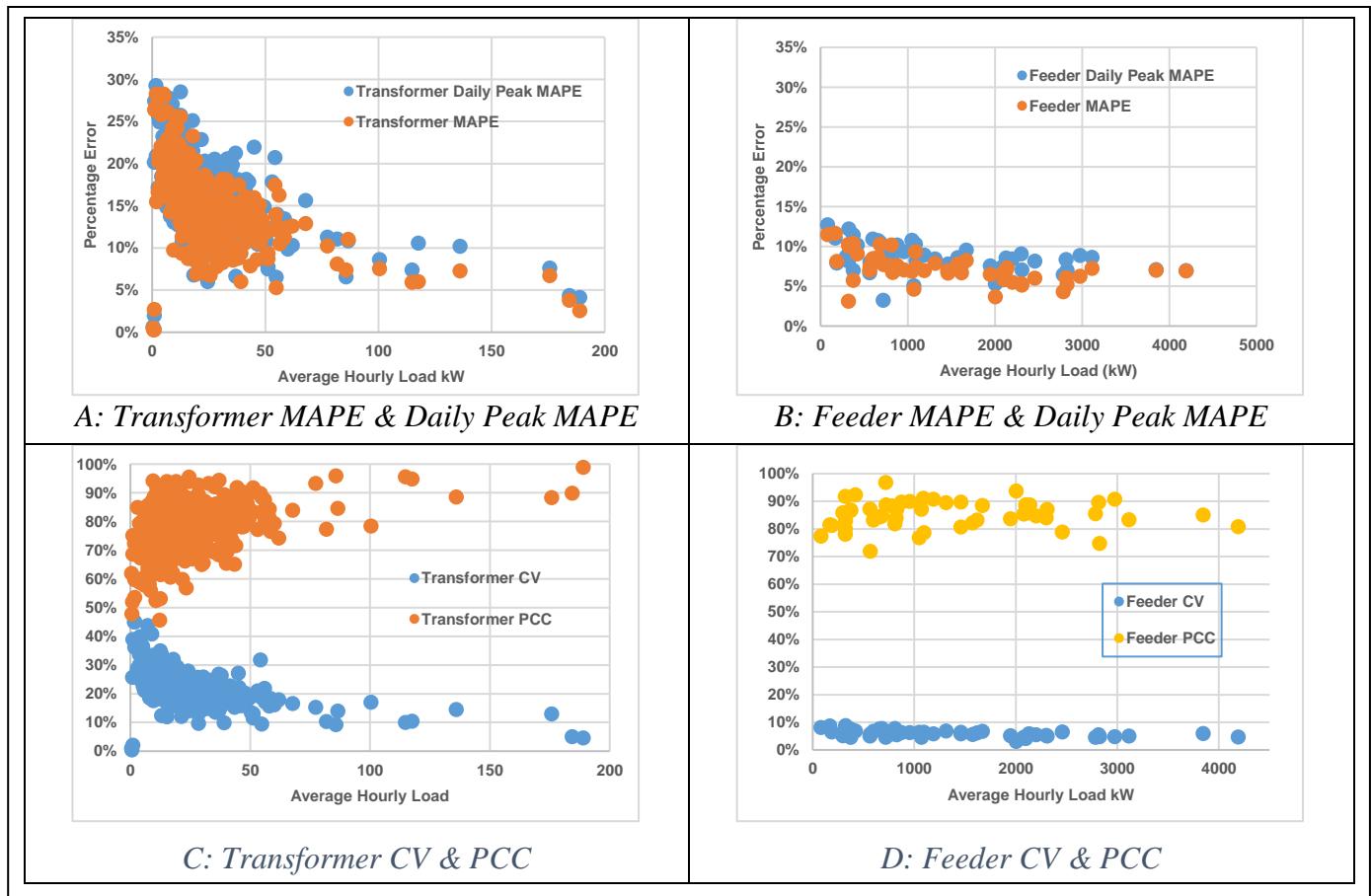


Figure 5-3: Model Performance Metrics Comparison

Table 5-7: Error Range Distributions for MAPE and Daily Peak MAPE

Error Range	MAPE	Daily Peak MAPE
Number of Transformers		
0% to 10%	47 (11.2%)	28 (6.7%)
10% to 13%	178 (42.4%)	176 (41.9%)
13% to 20%	151 (36.0%)	156 (37.1%)
20% to 30%	44 (10.5%)	60 (14.3%)
Number of Feeders		
% to 10%	48 (85.7%)	41 (73.2%)
10% to 15%	8 (14.3%)	15 (26.8%)

transformers or feeders in each percentage range along with the percent of the transformers or feeders in the percentage range. The data shows that more than 85% of the estimated transformer hourly load data has MAPE and Daily Peak MAPE values that are less than 20%. The data also shows that 85.7% of the estimated feeder hourly load data has MAPE values that are less than 10%, and that 73.2% of the estimated feeder load data has Daily Peak MAPE values that are less than 10%.

The CV and PCC values of the estimated load data are presented in Figures 5-4 plot C (transformers) and plot D (feeders). For the plot C, there are more than 70% of the estimated transformer load values have PCC values that are more than 0.8. Also, about 60% of the estimated transformer load values have CV values that less than 20%. All feeder estimated load values have CV values that are less than 10%, and 88% of the estimated feeder load values have PCC values that are more than 0.8.

The MAPTE values are summarized in Table 5-8. All feeder daily peak time estimates are within 2 hours of the original AMI daily load profile peak time. Only about 9% of the transformer daily peak time estimates are 3 hours from their corresponding AMI data measured peak time.

The performance metric data shows that in general the FDWSSM estimation performs better on the feeders/transformers that have larger loads. The MAPE and Daily Peak MAPE values

Table 5-8: Error Distributions for MAPTE

MAPTE	Under 2 Hour	3 Hour
Number of Feeders	56 (100%)	
Number of Transformers	381 (91%)	39 (9%)

have very similar trends. For feeders or transformers with smaller aggregated loads, the FDWSSM tends to have larger MAPE and Daily Peak MAPE values. This indicates a higher level of volatility of the aggregated load profiles, which leads to lower levels of predictability and higher estimation errors.

The CV and PCC values of the model also improve as the load gets bigger, which indicates that larger loads tend to have more stable load profiles. The larger feeder and transformer PCC values approach 1, which indicates the estimated load profiles have a strong and linear correlation with the corresponding original AMI hourly load profile.

5.4.3 Comparison with Other Models

There are a limited number of medium term load models based on AMI load data for transformers and feeders reported in the literature. In [26] the short-term load estimation mean MAPE for residential customers is 29%. Reference [26] also estimates the MAPE errors ranges from 24% to 45% for residential customers, whose hourly energy usage ranges from 1kWh to 4kWh. In [5] the medium term MAPE of the daily peak forecast error is about 38%.

The performance metric data in Table 5-7 shows that the FDWSSM model consistently performs better than the aforementioned models in term of MAPE values. The Daily Peak MAPE, CV and PCC data in Figure 5-5 also indicate the FDWSSM model also performs better, not only in term of estimating daily peak magnitude and time, but also in terms of the overall shape of the hourly load pattern.

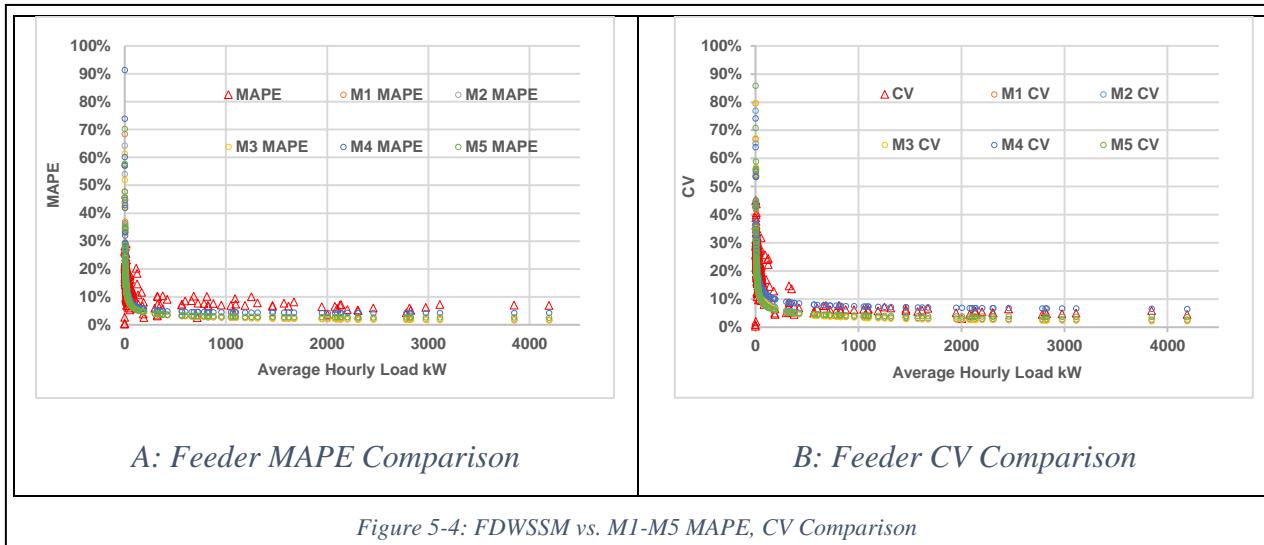


Figure 5-4: FDWSSM vs. M1-M5 MAPE, CV Comparison

Figures 5-5 plot A and B show comparisons between the FDWSSM model MAPE and CV values and the ideal empirical MAPE and CV values of five different algorithms proposed in [26], referred to as M1, M2, M3, M4, and M5. M1, M2 and M3 are Seasonal ARMA models, M4 is a Support Vector Regression model, and M5 is a Feed Forward Neural Network model.

When plotted against the average hourly load, the MAPE and CV values of the FDWSSM model exhibits a very similar pattern to all of the M1-M5 models. On average the MAPE and CV values of the FDWSSM model are about 40% different than the M1-M5 population average performance metric values, which are calculated using equation 5 and 6, along with the Scaling Law Fit parameters listed in tables 2 and 3, of reference [26].

There are two contributors to these differences: 1) The FDWSSM model deals with load patterns over a longer time frame than the short-term forecasting methods of [26]. 2) The limited coverage of the AMI data used in the FDWSSM model. The model was built using only 10 months' worth of AMI load data for a limited number of transformers and feeders. 3) The ideal empirical MAPE and CV curves were generated using data from Pacific Gas & Electric, which serves a region where weather diversity, mild summer and winter, is low. The AMI data used here was collected from the Mid-Atlantic region, which generally has a hot, stormy, summer season and a

cold winter season. Thus, the seasonal weather pattern of the Mid-Atlantic region is in general more diverse than that of California, leading to larger load variations that are more difficult to forecast.

D. Computational Performance

The construction of the FDWSSM model, the daily 24-hour load profile estimation, and FDWSSM model performance evaluations were all performed on a 64-bit, Window 7, Pro Laptop equipped with 16G RAM, 500GB hard drive, and an Intel I7 CPU. All related software modules are developed using Microsoft C# 4.0 and the T-SQL language. The original hourly AMI load data, the hourly weather data, the load model parameters, and the estimated hourly load data are all stored in a Microsoft SQL Server 2014 database on the same laptop.

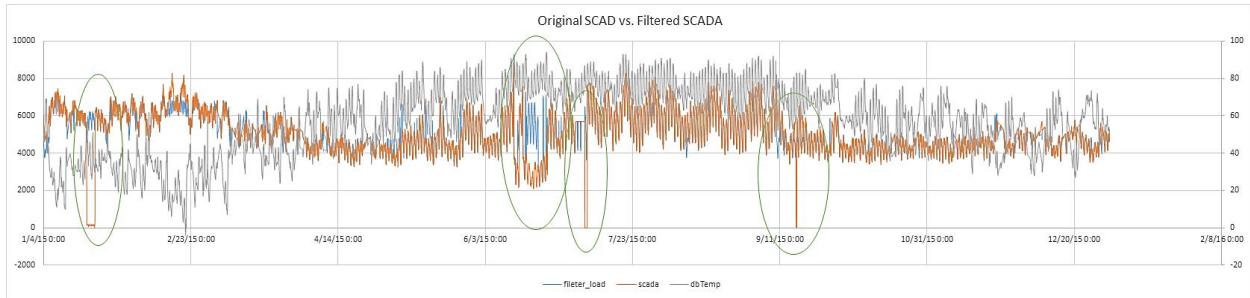
As discussed in section II, the FDWSSM model is constructed with a set of orthogonal frequency domain components which can be analyzed in parallel. The software modules used to calculate the frequency domain load model coefficients are implemented utilizing the parallel processing capability in the SQL server and C# language. The data in Table 5-9 shows that the time complexity of the load model construction and load estimation algorithm approximates the desirable linear $O(n)$.

Table 5-9: Execution Time for FDWSSM Model Construction & Load Estimation

	Execution Time (second)	
	56 Feeder	420 Transformer
Model Construction	33	275
Load Estimation Time (second)		
Single Run	3	26
15 Runs	53	401
30 Runs	106	850
60 Runs	210	1563

5.5 Summary

The Frequency Domain Weather Sensitive Stochastic Model (FDWSSM) provides a flexible, efficient and scalable solution for estimating the load profiles for transformers and feeders. The model is driven by non-temporal parameters and is flexible to incorporate many partitioning parameters (weather, customer class, spatial...). The model can generate a load estimation envelope, which can be used in detecting, filtering error/outlier electric load data. It can also be used to determine the equipment size. Using historical weather data, this load model can be used in data interpolation, back-cast and filling data gap of various size with known weather data as shown in the following plot (the blue line is the estimated line which can replace the erroneous data in original SCADA data in orange). The model can also be used in load forecasting with forecasted weather data.



The proposed load model construction and estimation process can take advantage of the distributed, parallel computing environments because each frequency domain component in the model can be calculated and analyzed independently.

Five performance metrics were used to evaluate the performance of the FDWSSM model, and the calculated metrics were compared with results from other models reported in the literature. From the performance evaluations and comparisons, the FDWSSM model exhibits good performance in terms of medium term forecasting.

In future work, it is planned to investigate the use of the FDWSSM model in load forecasting for large scale, integrated systems with detailed secondary circuits models.

Chapter 6: Contributions and Further Research

6.1 *Contributions*

From the proliferation of AMI load data and other load data monitoring technologies in recent years, the utility industry has accumulated vast amounts of high-resolution load data. There is a need for an analysis framework that can be used to gain insight into load behavior and answer questions such as:

- How does a customer or network's load profile behave under different conditions?
- What kind of patterns exist in the load profiles of various customer types?
- How will the load profiles behave in the future based on their historical performance?

The knowledge extracted from load data can play important roles in load forecasting, demand-side management, resource planning, and other areas. This research has provided contributions to the above questions.

This work has contributed new terminologies and algorithms for modeling load profile data in the frequency domain. It has provided the following three major contributions in the field of load modeling:

1. The load profile frequency domain descriptor introduced in chapter 2. This descriptor is more concise and consistent than the traditional temporal-based representations. It can be used to quantitatively describe a load profile as a set of independent frequency domain components. These components are orthogonal to each other, which makes the computational approach suitable for a parallelized computing environment, which is important to the processing of the vast amounts of load data.

In the frequency domain, it has been demonstrated that the majority of the energy of customer load profiles reside within the lower frequency descriptor components, which is to be expected for normal loads. By keeping these low frequency components, and by eliminating low magnitude, high frequency components (noise, errors, and sudden, unusual load spikes or drops), the length of the load descriptor can be potentially shortened substantially without losing much resolution and accuracy.

2. The load profile characteristic attributes in the frequency domain which are presented in chapter 3. These attributes are used to characterize significant periodic patterns that are contained in a load profile. Currently most utilities still categorize their customers based on economic factors (charge rate). Having the load profile characteristic attributes, a utility will be able to categorize customers based on their load behavior, and by doing so, determine more equitable rates.

In addition, a wavelet technique has been introduced in chapter 4, which can be used to identify and eliminate abnormal/erroneous measurements in load profiles.

3. The load profile, frequency domain, stochastic model introduced in chapter 5. This model is constructed using the frequency domain statistics calculated from the historical load data sets partitioned by meteorological data. The model has the flexibility to incorporate other parameters, such as geographical and sociological data. The stochastic load model can be used in Monte Carlo simulations to generate a load estimation envelope which presents a range of possible values for an estimated load profile. It is suitable for medium to long term load forecasting.

The key contributions of the work here may be summarized as follows:

- A new, generic, frequency domain load profile descriptor, that can quantitatively represent any time-series load profile data, is introduced.
- A new method for extracting the frequency domain pattern characteristic attributes for any time-series, load profile data is introduced. By utilizing these new attributes, a generic, decision tree based, load classification algorithm is presented.
- A new generic, frequency domain, statistical based, load modeling algorithm is introduced.

6.2 Future Research

The proposed framework presents an alternative direction for load modeling. There are many enhancements that can be investigated in future research, including the following:

- 1) Integrate the proposed load model with a detailed, system interconnected model to simulate and forecast the system load behavior.
- 2) Investigate modeling other significant, non-periodical patterns in load profiles.
- 3) Building out a comprehensive and self-upgradable load profile database with data from various sources. Such a database could be shared among utilities to improve their load modeling efforts.
- 4) Investigate incorporating other potentially influential parameters into the new load forecasting technique.
- 5) Incorporate PV and other renewable energy generation into the framework, thus creating a load-generation framework.

Reference

1. P. Price. Methods for analyzing electric load shape and its variability. volume LBNL-3713E, 2010
2. Marcelo Espinoza, Caroline Joye, Ronnie Belmans and Bart De Moor, “Short-Term load forecasting, profile identification, and customer segmentation: a methodology based on periodic time series,” *IEEE Trans. On Power Syst.*, Vol.20, No3, pp.1622-1630, Aug. 2005.
3. J.Nazarko, Z.A.Styczynski, “Application of statistical and neural approaches to the daily load profiles modeling in power distribution systems”, *Transmission and Distribution Conference*,Vol.1, Apr 11-16, 1999, pp320 – 325
4. A. Seppala, “Statistical distribution of customer load profiles”, *Energy Management and Power Delivery*, 1995. Proceedings of EMPD '95, 1995 International Conference on Vol. 2, 21-23, Nov. 1995, pp. 696 – 701
5. Rung-Fang Chang, Rong-Ceng Leou and Chan-Nan Lu, “Distribution transformer load modeling using load research data”, *IEEE Trans. Power Delivery*, Volume 17, Issue 2, pp. 655-661, Apr 2002
6. S.W. Heunis and R. Herman, “A thermal loading guide for residential distribution transformers based on time-variant current load models”, *IEEE Trans. Power Systems*, on Vol. 19, Issue 3, pp1294-1298, Aug. 2004.
7. M.K. Draper, “Modeling weather effects on electric energy sales,” Southeastcon '90. Proceedings vol.1, Apr 1-4, 1990 pp. 133-136.

8. L.G. Dias and M.E. El-Hawary, "Nonlinear Parameter Estimation Experiments for Static Load Modeling in Electric Power Systems". IEEE Proc. Vol. 136. No. 2, Mar. 1989, pp. 68-77.
9. J. Nazarko and W. Zalewski, "An evaluation for an Accuracy of the Fuzzy Regression Analysis in the Electrical Load Estimation", IEEE International Fuzzy Systems Conference Proceeding, Vol. 2, August 22-25 1999 p732-737
10. H.K. Kassael, A. Keyhani, T. Woung and M. Rahman, "a Hybrid Fuzzy, Neural Network Bus Load Modeling and Predication," IEEE Trans. Power Syst, Vol 14, No.2, pp 718-724, May 1999
11. Gheorghe Cartina, Virgil Alexandrescu, Gheorghe Grigoras and Mariana Moshe, "Peak Load Estimation in Distribution Networks by Fuzzy Regression Approach," 10th Mediterranean Eletrotechnical Conference, MeleCon 2000, Vol. III pp. 906-910.
12. P.K.Dash, A.C.Liew and G.Ramakrishna, "Power-demand forecasting using a neural network with an adaptive learning algorithm", IEEE Proceedings Vol. 142, No. 6, Nov. 1995 pp. 560-568.
13. G.Reza, Yousefi, "An application of artificial neural networks in power system load modeling", Power Engineering Society General Meeting, 2005. IEEE 12-16 June 2005, Vol. 1 pp. 308 – 313.
14. Donald B. Percival and Andrew T. Walden, *Spectral Analysis for Physical Applications: Multitaper and conventional Univariate Techniques*. Cambridge University Press, Cambridge, 1993.
15. Petre Stoica and Randolph Moses, *Spectral Analysis of Signals*. Prentice Hall, Upper Saddle River, NJ 2005.

16. Enrico Carpaneto, Gianfranco Chicco, Roberto Napoli and Mircea Scutariu, "Electricity customer classification using frequency-domain load pattern data", *Electrical Power and Energy System*, Vol. 28 pp. 13-20, 2006.
17. Sergio V. Verdu, Mario O. Garcia, Francisco J. B. Franco, Nuria Encinas, Antonio G. Marin, Angel Molina and Emilio G. Lazaro, "Characterizing and Identification of Electrical Customers Through the Use of Self-Organizing Maps and Daily Load Parameters", IEEE PES Power Systems Conference and Exposition 2004, Vol. 2, pp. 899-906, 2004
18. Ramos, S.; Duarte, J.M.; Duarte, F.J.; Vale, Z.; Faria, P., "A data mining framework for electric load profiling," *Innovative Smart Grid Technologies Latin America (ISGT LA), 2013 IEEE PES Conference On* , vol., no., pp.1,6, 15-17 April 2013
19. Kim, Y.; Kang, S.; Ko, J; Choi, S-H, "A study for clustering method to generate Typical Load Profiles for Smart Grid," *Power Electronics and ECCE Asia (ICPE & ECCE), 2011 IEEE 8th International Conference on* , vol., no., pp.1102,1109, May 30 2011-June 3 2011
20. Zhang, T.; Zhang, G.; Lu, J.; Feng, X.; Yang, W., "A New Index and Classification Approach for Load Pattern Analysis of Large Electricity Customers," *Power Systems, IEEE Transactions on* , vol.27, no.1, pp.153,160, Feb. 2012
21. Gerbec, D.; Gasperic, S.; Smon, I.; Gubina, F., "An approach to customers daily load profile determination," *Power Engineering Society Summer Meeting, 2002 IEEE* , vol.1, no., pp.587,591 vol.1, 25-25 July 2002
22. Gavrilas, M.; Sfintes, V.C.; Filimon, M.N., "Identifying typical load profiles using neural-fuzzy models," *Transmission and Distribution Conference and Exposition, 2001 IEEE/PES* , vol.1, no., pp.421,426 vol.1, 2001

23. Gerbec, D.; Gasperic, S.; Gubina, F., "Determination and allocation of typical load profiles to the eligible consumers," Power Tech Conference Proceedings, 2003 IEEE Bologna, vol.1, no., pp.5 pp. Vol.1, 23-26 June 2003
24. Almeshiae, E., Soltan, H. "A Methodology for Electric Power Load Forecasting", Alexandria Engineering Journal, pp137-144, 2011
25. Torkzadeh, R. Anaraki, A.S., Mirzaei, A., Sehhati M.R., Mirjalili, M.M. Behdad, F., "Medium Term Load Forecasting in Distribution Systems Based on Multi Linear Regression & Principal Component Analysis: A Novel Approach", The 19th Electrical Power Distribution Conference (EPDC2014), 6-7 May, 2014, Niroo Research Institute
26. Sevlian R., Rajagopal R., "Short Term Electricity Load Forecasting on Varying Levels of Aggregation", arXiv preprint arXiv:1404.0058
27. Zhong, S.; Tam, K., "A Frequency Domain Approach to Characterize and Analyze Load Profiles," Power Systems, IEEE Transactions on , vol.27, no.2, pp.857,865, May 2012
28. Zhong, S.; Tam, K.-S., "Hierarchical Classification of Load Profiles Based on Their Characteristic Attributes in Frequency Domain," Power Systems, IEEE Transactions on, vol.PP, no.99, pp.1-8
29. M. D. Felice and X. Yao, "Short-term load forecasting with neural network ensembles: A comparative study (application notes) ;" IEEE Computat. Intell. Mag., vol. 6, no. 3, pp. 47–56, Aug. 2011
30. H. Lee Willis, *Power Distribution Planning Reference Book, Second Edition*, Marcel Dekker, Inc. New York, 2004
31. H. Lee Willis, *Spatial Electric Load Forecasting, Second Edition*, Marcel Dekker, Inc. New York, 2004

32. Rodger E. Ziemer, William H. Tranter, D.Ronald Fannin, *Signals & Systems Continuous and Discrete*, Prentice Hall, Upper Saddle River, NJ 1998
33. Soliman A. Soliman, Ahmad M. Al-Kandari, *Electrical Load Forecasting*, Butterworth-Heinemann, 2010
34. N.I. Fisher, *Statistical Analysis of Circular Data*, Cambridge University Press, 1995
35. Roger J. Lewis, “An Introduction to Classification and Regression Tree (CART) Analysis”, Annual Meeting of the Society for Academic Emergency Medicine, 2000, San Francisco, California
36. Breiman, L.; Friedman, J.H.; Olshen, R.A.; Sone, C.J. *Classification and Regression Trees*. Wadsworth, Belmont, CA, 1983
37. Belsley, D. A., Kuh, E., and Welsch, R. E. (1980), *Regression Diagnostics*, New York, NY: John Wiley & Sons, Inc.
38. Therneau, T.M; Atlinson E.J; “An introduction to recursive partitioning using the rpart routines”, 2013, Mayo Foundation
39. <http://www.ercot.com/mktinfo/loadprofile/alp/>
40. http://www.pseg.com/business/energy_choice/third_party/historical.jsp
41. <https://www.sce.com/wps/portal/home/regulatory/load-profiles>
42. http://www.nationalgridus.com/niagaramohawk/business/rates/5_load_profile.asp
43. http://www.pseg.com/business/energy_choice/third_party/load_profiles.jsp
44. Loh, Wei-Yin, “Classification and regression trees”, WIREs Data Mining Knowledge Discovery 2011 1 14–23 DOI: 10.1002/widm.8, 2011 John Wiley & Sons, Inc.

45. Milano, F.; Zhou, M.; Guanji Hou, "Open model for exchanging power system data," Power & Energy Society General Meeting, 2009. PES '09. IEEE, doi: 10.1109 / PES.2009.5275964
46. <https://www.aepohio.com/service/choice/cres/LoadProfiles.aspx>
47. Sargent, A., Broadwater, R. P., Thompson, J., Nazarko, J., "Estimation of Diversity and KWHR-to-Peak-KW Factors from Load Research Data," IEEE Transactions on Power Systems, Vol. 9, No. 3, pp. 1450-1456, August 1994.
48. Ingrid, D. "Ten Lectures on Wavelets", SIAM 1992
49. Mallat, S.G. "A theory for multi-resolution signal decomposition: The wavelet representation." IEEE Trans. on Patt. Anal. Mach. Intell., 11(7): 674-693.
50. Antonini M., Barlaud M, Mathieu P., and Daubechies I., "Imaging coding using vector quantization in the wavelet transform domain," in Proc. 1990 IEEE International Conference on Acoustic, Speech, Signal Processing, Albuquerque, NM, Apr. 3-6, 1990, pp. 2297-2300
51. A. K. Jain, M. N. Murty, and P. J. Flynn. 1999. Data clustering: a review. ACM Comput. Surv. 31, 3 (September 1999), 264-323. DOI=<http://dx.doi.org/10.1145/331499.331504>
52. Hong T. Load Forecasting Case Study, ISPC and NARUC, Jan, 2015.
53. A.J. Al-Shareef, E.A. Mohamed, and E. Al-Judaibi. "One-hour ahead load forecasting using artificial neural network for the western area of saudi arabia", International Journal of Electrical Systems Science and Engineering, (1):35–40, 2008
54. C. Donald Ahrens, *Meteorology today*. West Publishing Company, 104-105, 1985
55. K.Berk, *Modeling and Forecasting Electricity Demand: A Risk Management Perspective*, Springer Spektrum, 2015

56. National Weather Service on-line weather data repository, <http://www.ncdc.noaa.gov/>