ENHANCED CUSTOMER DEMAND LOAD PROFILES ESTIMATION ALGORITHMS

FOR FIELD APPLICATION

by

XIN WANG

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

August 2016

Acknowledgements

I would like to convey my sincere gratitude to my supervisor professor, Dr. Wei-Jen Lee, for his patience, guidance, help, and support throughout the process of my academic career and this dissertation. I want to take this opportunity to express my deep gratitude to Dr. Lee for extending his help and guidance leading me to the end of my dream of achieving a doctoral degree. I would be forever owe him for his unwavering support and sound advice during my academic career.

I would like to thank Dr. William E. Dillon, Dr. Heng Huang, Dr. David A. Wetz, and Dr. Rasool Kenaragui for their instruction, serving on my dissertation committee, valuable suggestions and review.

I also want to express my deep appreciation to all members of Energy Systems Research Center (ESRC) at the University of Texas at Arlington, for all their valuable assistances, discussions, and enjoyable association. Besides, I would like to thank all colleagues at Consolidated Edison Company of New York, Inc. (Con Edison) for their support and comments on this dissertation.

Finally, I would dedicate my dissertation to my families for serving as my continued source of inspiration for the tenure of my doctoral degree. I am very grateful and appreciate them for their love, encouragement, and support.

July 22, 2016

Abstract

ENHANCE CUSTOMER DEMAND LOAD PROFILES ESTIMATION ALGORITHMS FOR FIELD APPLICATION

Xin Wang, PhD

The University of Texas at Arlington, 2016

Supervising Professor: Wei-Jen Lee

Due to the deregulation of the power system, the electric power industry is undergoing a transformation in terms of its planning and operation strategies. Because of the importance in reducing financial and operational risk, improving load forecasting accuracy is paramount. In some load forecasting applications, K-means clustering is used to group customers prior to forecasting. This method has been shown to improve the accuracy of load predictions. However, there are situations where K-means clustering reduces load forecasting accuracy. This dissertation studies the factors that affect the performance of K-means clustering. The data used for validating the proposed strategies associated with the factors is from Consolidated Edison Company of New York, Inc. (Con Edison). The mean absolute percent error (MAPE) and relative mean square error (RMSE) are utilized to evaluate the forecasting results of K-means based least squares support vector machines (LS-SVM) and preprocessed K-means based LS-SVM. Additionally, the outperformance of preprocessed K-means based LS-SVM is demonstrated via the data results.

As the improving of life quality, electricity consumption is also increased. Though elements that affect electricity consumption are various, the overall influence is more important than individual one. This dissertation proposes the load growth factor, which

focuses on the overall pattern of consumption changes. The procedures for load growth factor application is conducted by different groups of customers, and help to improve the load profile estimation accuracies.

Peak demand forecasting accuracy is affected by the temperature, this dissertation comes up with a method that including temperature variables to the estimation process. The related calculation equations are shown in this dissertation.

This dissertation aims to develop the load profile forecast algorithm and increase the forecast accuracy for coincident peak demand.

Table of Contents

List of Illustrations

List of Tables

Chapter 1

INTRODUCTION

1.1 Research Background

Electricity is an essential part of modern society. Most consumers do not think much about their electricity until a power outage, or when they receive a high utility bill. With better understanding of electricity consumption pattern, it will promote behavior change to improve the efficiency at customer side and provide information for utility to have better planning, operation, and management.

Advanced Metering Infrastructure (AMI) is an emerging technology evolving from Automated Meter Reading (AMR). Con Edison is implementing an AMI system to gather and utilize metered data in a more intelligent and cost effective manner. AMI is a one of the most important components for future Smart Electricity Grids.

Algorithms have been developed in the former project to utilize data collected from AMI system for two purposes: 1) to provide accurate customer daily load profiling for load estimation and network demand reconciliation to improve the efficiency and security of the underground network of Con Edison systems, and 2) to perform a nonintrusive load monitoring analysis for discerning individual major appliances from residential customers in the Con Edison's service area. The former project shows promising results on proposed algorithms. This project processes the following validation/enhancement to improve the former development and make them suitable for field applications:

- Vertical validation process: using multi-year metered data to verify the trend and predictability of future demand profile of the metered customers based upon 2012 data.
- Horizontal validation process: using the developed demand profile to forecast the demand profile of metered data that are not used in the demand profile developing process and forecast the demand profile for the customers without smart meters.
- Peak demand estimation: using the smart meter data to refine service demand estimators slope and intercept to estimate peak demand from annual usage.

In some load forecasting applications, k-means clustering is used to group customers prior to forecasting. This method is also applied in this project to potentially improve the load profile forecast algorithm.

Peak demand is the maximum electricity requirements that customers need at a single point of time. Utilities have to meet the peak demand in order to ensure the reliability of the system operation. Since it related to system security and drives infrastructure investment, it is important to have accurate estimation on the peak demand. In Con Edison, the Service Demand Estimator uses a slope and intercept to estimate peak demand. Refining these slopes and intercept values would help improving the accuracy for coincident peak demand forecast.

1.2 The Proposed Method

In this dissertation, a hybrid method is conducted to perform all requisite predictions. A clustering method is used to group customers first, then a forecasting technique is applied to each group for estimation. The K-means clustering method, considered as one of the most effective clustering algorithms, is adopted to cluster customers. Support Vector Machines (SVMs), which consist of Support Vector Regression (SVR) and Support Vector Classification (SVC), are applied to estimate day ahead load forecasting.

Based on K-means clustering, customers are grouped into certain numbered clusters. After they are grouped, LS-SVM is utilized on each cluster for load forecasting. The prediction results from all of the clusters are combined to obtain the final estimation outcome. If the percentage of customers in one cluster is too low in comparison with the whole customer number, then this kind of cluster would be regarded as an outlier. That is to say, it may be studied separately or even excluded.

The final K value is decided based on the highest accuracy for the combined forecasting results. The following flowchart presents the procedures of the K-means based LS-SVM model:

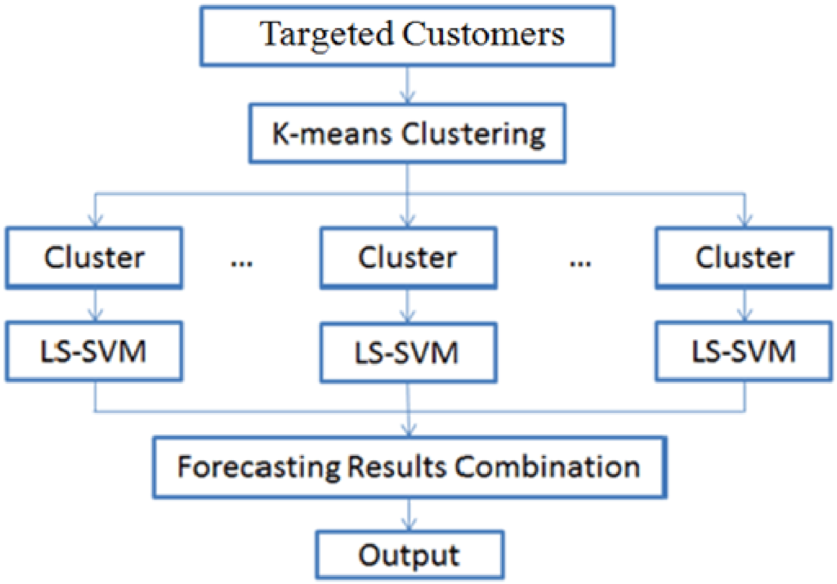


Figure 1-1 Flowchart for Proposed Method

Besides customer behaviors, there are other factors that influence electricity consumption. The load growth factor as well as the variation of temperature impact the load profile estimation and peak demand forecasting significantly.

1.3 Study Objectives

This dissertation shows the comparison results for estimation among stratification groups and cluster groups. The results prove that the clustering technique improves estimation accuracy. In addition to this finding, this study also analyzes other factors that can further improve the prediction precision.

1.4 Synopsis of Chapter

The organizational structure of this dissertation is as follows:

Chapter 1 introduces the general background of load forecasting, illustrates the significance of this research topic, and describes the motive and objective behind this dissertation.

Chapter 2 reviews the importance of load profile forecast and peak demand estimation. It describes the proposed method, which includes the clustering technique and prediction method.

Chapter 3 shows the application of the proposed method in load profile estimation for customer with smart meters.

Chapter 4 describes the application of the proposed method in load profile estimation for customer traditional meters.

Chapter 5 illustrates the process for improving peak demand estimation accuracy.

Chapter 6 presents the conclusions drawn from the research associated with this dissertation and discusses the opportunity for future research.

Chapter 2

LITERATURE REVIEW

2.1 Load Forecast Importance

Load forecast accuracy improvement is always paramount. Energy policy decisions can be made based on accurate load forecasting. High precision load forecasting improve the reliable operation of power systems that lead to higher service continuity to the consumer. An accurate demand forecast can help system operators determine unit commitment schedules more efficiently and economically. Improvement of forecasting accuracy results in the reduction of operating costs and a reliable and economic power system operations.

2.2 Peak Demand Forecast Importance

Understanding the peak demand is necessary for utilities to evaluate and hedge the financial risk. Peak demand forecasting is the basis for the system state estimation and for technical and economic calculations of the generation and distribution system. Operators of dispatching centers for scheduled maintenance or adequacy assessment require a peak load. Peak demand normally occurs in summer due to high temperatures. With the influence of various factors, peak demand tends to increase from year to year. The following figure of historical and forecast peak demand for the Electric Reliability Council of Texas (ERCOT) indicates the trend:

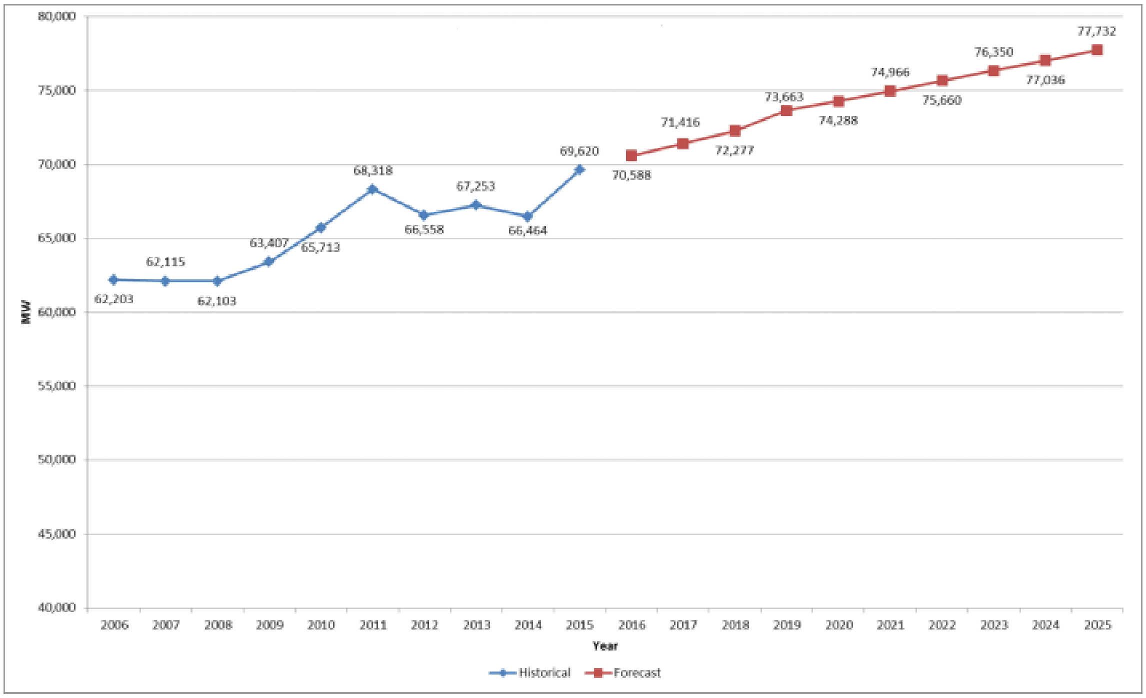


Figure 2-1 ERCOT Historical and Forecast Peak Demand

2.3 Clustering Technique – K-Means Clustering

The purpose of clustering is to discover the natural groupings of a set of points, patterns, objects and individuals by grouping or partitioning the data into clusters containing similar elements. Intuitively, the elements that belong to different clusters are dissimilar [13-16, 25, 26]. There are two large types of clustering: fuzzy clustering and hard clustering. Fuzzy clustering means one object can belong to more than one cluster, while hard clustering means one object belongs to only one cluster. Here, hard clustering method is selected. Hard clustering consists of partitioned clustering and Hierarchical clustering.

Hierarchical is computationally intensive and not often used. Partitioned clustering is more efficient. K-means clustering is one of the most popular partitioned clustering techniques used across various disciplines on a wide variety of data. K-means is widely used due to its fast iterative algorithm that can sort large sets of data.

K-means clustering is the most popular method in partitioned clustering. It partitions the n data points in a set $(\mathrm{N}_1, \dots, \mathrm{N}_\mathrm{n})$ into k subsets $(S_1, \dots, S_\mathrm{k})$. K-means clustering could be described as the following [4-6, 8, 11]:

$$S_i = N_j, i \in \{1, \dots, k\}, j \in \{1, \dots, n\}$$

Once the k clusters are given, the remaining data are allocated to the closest clusters [34-36]. An error function occurs in the K-means clustering. Then, the clusters' elements should be modified by the error function until the error function has little to no change. The error function is shown as:

$$E = \sum_{i=1}^{k} \sum_{N_j \in C_i} \left| N_j - S_i \right|^2$$

Here, C_i is the i^{th} cluster.

In order to solve the local minima problem in K-means, usually the K-means algorithm with various partitions would be selected along with the smallest squared error.

2.4 Forecasting Method –Support Vector Machines (SVM)

Traditional methods are difficult to represent the non-linear nature of the different load patterns of power system. Weak generalization ability is resulted from overfitting and some other disadvantages with artificial neural network, like the number of hidden neurons determination and the sensitivity of the final weights to the initials, cannot be avoided. SVM solves the problems of over-fitting, dimension disaster and local minimum existed in ANN and is currently recognized as one of the standard tools for load forecasting.

Support Vector Machines (SVM) is a learning method developed from Statistical Learning Theory [4, 19, 38]. It deals with regression, classification and other machine learning tasks. LS-SVM is proposed for solving linear equations rather than quadratic programming. In this section, the LS-SVM [3, 18, 20, 27-30, 33] is introduced, which is preferred for large-scale conditions.

Suppose a given observation data set: $\{(x_1, y_1), \dots, (x_i, y_i)\}$, where $x_i \in R^n, y_i \in R$ [3-6, 21]. To achieve the nonlinear regression, a nonlinear mapping $\varphi(\cdot)$ maps the input space to a high dimensional feature space in order to establish a linear regression formula:

$$f(x) = \omega^T \varphi(x) + b \tag{1}$$

where ω is the weight vector and b is the bias term.

In order to obtain ω and b, function should be minimized with structural risk function [2]:

$$R_{reg} = \frac{1}{2}\|\omega\|^2 + C \cdot R_{emp}[f] \tag{2}$$

Where $\|\omega\|^2$ is a function to describe the complexity of $f(\cdot)$; Constant C determines the trade-off between empirical risk and Vapnik Chervonenkis (VC) confidence.

For the sake of estimation, the regression can transfer to the optimization. The formula for optimization is shown below:

$$\min \frac{1}{2}\|\omega\|^2 + C \sum_{i=1}^{l} \xi_i \tag{3}$$

subject to constraints:

$$y_i(\omega \cdot \varphi(x_i) + b) \geq 1 - \xi_i,$$

$$\xi_i \geq 0, i = 1, \dots, C > 0.$$

$\varphi(x_i)$, in the above equation, maps x_i to a higher dimensional space, $\varphi(x_i)$ can be replaced by kernel formulas $K(x_i, x_j)$. The output of LS-SVM can then be described as follows:

$$y(x) = \sum_{i=1}^{l} \alpha_i k(x, x_i) + b \tag{4}$$

where α_i is the Lagrange multiplier. The kernels in common use are [2-5]:

(1) Polynomial Kernel:

$$K(x, x_i) = (\langle x, x_k \rangle + p)^d, d \in N, p > 0$$

(2) Radial Basis Function Kernel:

$$K(x, x_i) = \exp(-\|x - x_i\|^2 / 2\sigma^2)$$

(3) Hyperbolic Kernel:

$$K(x, x_i) = \tanh(\beta x^T x_i + k)$$

Chapter 3

PROPOSED METHOD APPLICATION IN SMART METER LOAD PROFILE ESTIMATION

3.1 Smart Meter Service Class and Stratum Category

Smart meters or Advanced Metering Infrastructure (AMIs) allow bi-directional, real-time communication between the utility and the consumer. Besides enabling the utility to measure consumer power consumption in near real-time, the AMIs act as a communication gateway for the utility to interact with the home or building power control system and appliances. Similar to other utility companies, Con Edison has implemented an AMI system and is exploring the possibility to utilize metered data in a more intelligent manner to improve the system efficiency, reliability, and security.

Load profile estimation has been identified as one of the most desirable applications after AMI implementation since the load consumption pattern and the daily peak load are vital factors in scheduling, operation, and control of the utility grid. The demand forecasting is useful to plan and purchase power supply, schedule equipment maintenance, and provide early warning to consumers of potential load curtailment or advance pricing information. For system operation, it is essential to know with as much accuracy as possible what the total and local system demand will be in the next minutes, hours, and days, so that generators with different costs and various constraints can be scheduled to optimize the total system efficiency.

Based upon the tariff, the Con Edison establishes the service classes definitions. Stratum Category represents a subgroup within a service class. It is the measure of the size of a customer as defined by a particular billing quantity. Table 3-1 relates the stratum variables to the service classes provided by Con Edison.

Table 3-1 Service Classes and Their Stratum Billing Variables

Service Class	Service Class Description	Stratum Billing Variable
SC 1	Residential (excluding Religious and Water Heating)	Annual kWh
SC 2	General Small Business	June-Sept kWh
SC 3	Backup Service	Annual kWh
SC 4	Commercial Redistribution	Average June-Sept kW
SC 5	Traction	Annual kWh
SC 6	Private Street Lighting	Annual kWh
SC 7	Residential Space and Water Heating	Annual kWh
SC 8	Multiple Dwelling Redistribution	Average June-Sept kW
SC 9	General Large Business	Average June-Sept kW
SC 12	Multiple Dwelling Space Heating	Nov-Feb kWh
SC 62	General Small	Nov-March kWh
SC 64	General Small	Annual kWh

There are six stratum groups (A thru F) within service class 1 (SC1) – residential (excluding religious and water heating). Table 3-2 shows the annual kWh ranges for the stratification variables. Group A includes unoccupied apartments.

Table 3-2 Service Class 1 – Residential and its Stratum Groups

# Class	Group	Stratification Criteria	Low (kWh)	High (kWh)
RESID	A	Total Annual kWh	0	1948
RESID	B	Total Annual kWh	1949	2897
RESID	C	Total Annual kWh	2898	3897
RESID	D	Total Annual kWh	3898	5239
RESID	E	Total Annual kWh	5240	7741
RESID	F	Total Annual kWh	7742	9999999999

There are five stratum groups (A thru E) within service class 2 (SC2) – general small business. Table 3-3 shows the total June-September kWh ranges for the stratification variables. Group A includes unoccupied small business customers.

Table 3-3 Service Class 2 – General Small Business and its Stratum Groups

# Class	Group	Stratification Criteria	Low (kWh)	High (kWh)
GENSM	A	Total June-Sept kWh	0	1500
GENSM	B	Total June-Sept kWh	1501	3000
GENSM	C	Total June-Sept kWh	3001	5000
GENSM	D	Total June-Sept kWh	5001	8000
GENSM	E	Total June-Sept kWh	8001	9999999999

There are four stratum groups (A thru D) within service class 4 (SC4) – commercial redistribution. Table 3-4 shows the total June-September kW ranges for the stratification variables. Group A includes unoccupied commercial redistributions.

Table 3-4 Service Class 4 – Commercial Redistribution and its Stratum Groups

# Class	Group	Stratification Criteria	Low (kW)	High (kW)
CMREDISTR	A	Average June-Sept kW	0	221
CMREDISTR	B	Average June-Sept kW	222	431
CMREDISTR	C	Average June-Sept kW	432	741
CMREDISTR	D	Average June-Sept kW	742	9999999999

There are four stratum groups (A thru D) within service class 7 (SC7) – residential space and water heating. Table 3-5 shows the annual kWh ranges for the stratification variables. Group A includes unoccupied residential spaces and water heating.

Table 3-5 Service Class 7 – Residential Space & Water Heating and its Stratum Groups

# Class	Group	Stratification Criteria	Low (kWh)	High (kWh)
RESPWH	A	Annual kWh	0	6698
RESPWH	B	Annual kWh	6699	10505
RESPWH	C	Annual kWh	10506	14247
RESPWH	D	Annual kWh	14248	9999999999

There are five stratum groups (A thru E) within service class 8 (SC8) – multiple dwelling redistribution. Table 3-6 shows the average June-September kW ranges for the stratification variables. Group A includes unoccupied multiple dwelling redistribution.

Table 3-6 Service Class 8 – Multiple Dwelling Redistribution and its Stratum Groups

# Class	Group	Stratification Criteria	Low (kW)	High (kW)
MDREDIS	A	Average June-Sept kW	0	120
MDREDIS	B	Average June-Sept kW	121	220
MDREDIS	C	Average June-Sept kW	221	340
MDREDIS	D	Average June-Sept kW	341	530
MDREDIS	E	Average June-Sept kW	531	9999999999

There are five stratum groups (A thru E) within service class 9 (SC9) – general large business. Table 3-7 shows the average June-September kW ranges for the stratification variables. Group A includes unoccupied general large business.

Table 3-7 Service Class 9 – General Large Business and its Stratum Groups

# Class	Group	Stratification Criteria	Low (kW)	High (kW)	# Class
GENLG	A	Average June-Sept kW	0	17	GENLG
GENLG	B	Average June-Sept kW	18	33	GENLG
GENLG	C	Average June-Sept kW	34	70	GENLG
GENLG	D	Average June-Sept kW	71	200	GENLG
GENLG	E	Average June-Sept kW	201	9999999999	GENLG

To characterize the customer load curves, Service Class and Stratum Group should be known as the priori for all the meters that are used for load profiles development based on Stratification information, but not when the load profiles are developed based on Clustering techniques, where the customers groups are inferred from the data themselves.

During the project period, we have received both the smart meter data and traditional meter data from Con Edison. For task one, smart meter data part, we have to choose the SC with relatively large customer number for the desired applications. Numbers of smart meter customers in the year 2012 and their service classes are shown in Table 3-8.

Table 3-8 Received Smart Meter Data in the Year of 2012

SC	**2012**										
	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
RESID/SC1	1084	1232	1235	1476	1586	1568	1562	1555	1552	1587	1583
SC 2	171	311	296	327	322	311	325	328	326	340	344
SC 7	10	23	23	25	25	25	25	25	25	25	25
SC 9	198	388	372	378	377	357	349	365	363	382	385
SC 91	6	17	18	14	15	15	14	13	13	14	14
WAHTD	1	6	6	6	6	6	6	6	6	3	3
RELIG	0	1	1	2	2	2	2	2	1	2	2
#N/A	72	115	105	108	103	94	90	90	90	88	90
TOTAL	1542	2093	2056	2336	2436	2378	2373	2384	2376	2441	2446

Based on the number of customers for each service class, SC1, SC2 and SC9 are selected to conduct the smart meter load profile forecasting. One can follow the same procedure to perform load profile forecasting for other SCs with sufficient customers.

3.2 Load Profile Estimation Algorithm Development Based on Stratification Group Information

The objective of this part is to develop load profiles estimation algorithm. First of all, the load profiles should be developed. Because smart meter is located at the endpoint, it can be aggregated in different ways to serve the purpose. Load profile development follows a straightforward, structural approach based on Service Class and Stratification information. Service Class represents a group of customer types with similar load characteristics, and Stratum Category represents a subgroup within a service class.

As required for load profile development, curves are defined at 15-minute interval, resulting in a 96-point daily curve. In general, only Channel 1 (total kWh) is needed for load profile estimation.

When more than one day is considered, the meter data should be available at all time, that's to say, the meters that have complete data set during the entire study period to be extracted. Then, the next step is to average the load consumption at each individual meter the whole time period. Once individual load consumption at the meter level is determined, the next step is to proceed with load profile for consumers that belong to a strata category. The load profile is characterized by the mean and standard deviation.

Due to the large amount of load profile cases that can be obtained from the current data, presenting exhaustive results in the main body of the report would be not convenient. The approaches for developing the load profile estimation algorithm by smart meter data are shown in the following parts.

3.2.1 Apply K-Means Clustering for Each Stratification Customers within Each SC

In order to develop load profile forecast algorithm, k-means clustering is utilized. Clustering techniques have a long and rich history in grouping samples with similar characteristics. Different clustering algorithms have been developed over time. However, none of the clustering is superior to the other, they all have pros and cons[39]. Since the structure of the data is not known a priori, it is up to the analyst to try competing and diverse approaches to determine a suitable algorithm for the clustering at hand. Based upon the results of former project, k-means is a suitable approach for clustering load profiles. Since we have obtained enough number of customers, k-means clustering approach is applied to SC1, SC2 and SC9 customers.

The stratification criteria for SC1, SC2 and SC9 are vary, also the stratification group for each SC are different. For each SC, we have received both 15-min interval and 1-min interval customers. We do not perform load profile estimation for SC2 and SC9 15-min interval customers, due to insufficient customer numbers. There are two groups of 1-min interval customers. Since only Channel 1 is needed, we have combined two 1-min interval groups for clustering and load profile forecasting.

In order to study the effectiveness of applying clustering in developing load profile forecast algorithm, the load profile forecast (day-ahead) for each stratification with and without applying k-means clustering are analyzed. Mean absolute percentage error (MAPE) is the criteria to evaluate the accuracy of load profile forecast. The results are shown in Table 3-9.

Table 3-9 MAPEs for 1-Min and 15-Min Interval Data without and with Clustering

	1-Min interval data		**15-Min interval data**	
	Without	**With**	**Without**	**With**
SC1 (k=6)	8.38%	7.03%	8.84%	9.38%
SC2 (k=5)	13.23%	13.09%	N/A	N/A
SC9 (k=5)	6.29%	6.24%	N/A	N/A

When there are enough customers in the stratification, the k-means performances on load profile forecast for each stratification customers show the advantage of developed forecast algorithm. The high MAPE values for some groups are due to the small customer numbers in those groups.

3.2.2 Perform Clustering by Combining 15-Min and 1-Min Data Together

The purpose for this part of study is to test whether increase the number of customers will improve k-means performance for algorithm development.

To have the same format, 1-min interval data is converted to 15-min interval data by summing up fifteen continuous 1-min interval data. The results are shown in Table 3-10.

Table 3-10 MAPE Comparison for Combined Data

	15-Min interval data	
	Without Clustering	**With Clustering**
SC1	9.31%	8.89%
SC2	11.07%	14.23%
SC9	6.18%	6.73%

The results in this part are not as expected, some MAPE values are higher than previous section. After investigation, it is found out that the differences of the consumption between 15-min meter data and the merged 15-min meter data are large. The combined two groups with large differences lead to negative effect on the load profile forecast.

3.3 Load Profile Estimation Algorithm Development Based on K-Means Clustering Group Information

Instead of using stratification subgroups to derive individual load profile, we re-group SC customers by means of k-means clustering and apply the forecast algorithm within the subgroups.

The curves for load profiles in this part are still defined at 15-minute interval, i.e. a 96-point daily curve. Also only Channel 1 is needed for load profiling.

Following the same case study as before, the load profiles obtained through clustering can be found in the APPENDIX A for SC9 (1-minute interval).

3.3.1 Perform K-Means Clustering within Each SC

There are 6 stratification groups (A thru F) for SC1, 5 stratification groups (A thru E) for SC2 and 5 stratification groups (A thru E) for SC9. In order to compare the forecast result for each SC, the k values of 6, 5, and 5 are used for SC1, SC2 and SC9 respectively.

3.3.1.1 Perform Clustering on 15-Min and 1-Min Data Separately

There are two groups of 1-min interval customers for each SC. The two groups are combined to perform clustering.

In order to study the clustering application in load profile forecast algorithm development, the forecast for each SC with and without k-means are both analyzed. The results are shown in Table 3-11.

Table 3-11 MAPEs for 1-Min and 15-Min Interval Data without and with Clustering

	1-Min interval data		**15-Min interval data**	
	Without	**With**	**Without**	**With**
SC1 (k=6)	10.47%	8.01%	9.494%	8.117%
SC2 (k=5)	8.65%	6.177%	N/A	N/A
SC9 (k=5)	4.05%	4.071%	N/A	N/A

Since we do not have sufficient 15-min interval customer data, load profile forecasting for SC2 and SC9 are not performed.

The k-means performances on load profile forecast for SC1 and SC2 customers show the improvements compared with the no k-means ones. For SC9, we have obtained satisfactory load profile forecasting results with and without k-means clustering, due to smaller variations among load consumption pattern of the General Large Business.

3.3.1.2 Perform K-Means Clustering on 15-Min and 1-Min Data Together

The purpose for this part of study is to test whether increase the number of customers will improve k-means performance for algorithm development.

First step for this part is to merge 1-min interval data to 15-min interval data by summing up fifteen continuous 1-min interval data. The accuracies are shown in Table 3-12.

Table 3-12 MAPE Comparison for Combined Data

	15-Min interval data	
	Without Clustering	**With Clustering**
SC1 (k=6)	6.84%	10.40%
SC2 (k=5)	7.82%	5.493%
SC9 (k=5)	3.38%	4.189%

The results in this part are not as expected, some MAPE values are higher than previous section. After investigation, it is found out that the differences of the consumption between 15-min meter data and the merged 15-min meter data are large. The combining two groups with large differences lead to negative effect on the load profile forecasting.

3.4 Load Growth Factor Application

With the economy development, electricity consumption increases every year. Elements those affect electricity consumption are various, the overall influence is more important than individual one.

Load growth factor is focused on overall effect on electricity consumption. The factor was injected to the estimation procedures, so that the original prediction results were updated. Load growth factor is not a certain value, after several experiments, the range for load growth factor variables is 1.04 ~ 1.08. The updated accuracies for smart meter customer consumptions are shown in following tables:

Table 3-13 MAPE Comparison for Stratification Group 1-Min Interval Data with or without Load Growth Factor

	1-Min interval data		
	Load Growth Factor	**Original Accuracy**	**Updated Accuracy**
SC1	1.05	7.04%	5.76%
SC2	1.08	13.09%	10.49%
SC9	1.05	6.24%	5.18%

Table 3-14 MAPE Comparison for Stratification Group 15-Min Interval Data with or without Load Growth Factor

	15-Min interval data		
	Load Growth Factor	**Original Accuracy**	**Updated Accuracy**
SC1	1.06	8.84%	5.33%
SC2		N/A	

SC9		N/A	

Table 3-15 MAPE Comparison for Stratification Group Combined 15-Min Interval Data with or without Load Growth Factor

	Combined 15-Min interval data		
	Load Growth Factor	**Original Accuracy**	**Updated Accuracy**
SC1	1.07	8.89%	8.85%
SC2	1.04	11.08%	10.53%
SC9	1.04	6.18%	6.078%

Table 3-16 MAPE Comparison for Cluster Group 1-Min Interval Data with or without Load Growth Factor

	1-Min interval data		
	Load Growth Factor	**Original Accuracy**	**Updated Accuracy**
SC1	1.05	8.01%	5.90%
SC2	1.06	6.18%	6.02%
SC9	1.02	4.05%	4.01%

Table 3-17 MAPE Comparison for Cluster Group 15-Min Interval Data with or without Load Growth Factor

	15-Min interval data		
	Load Growth Factor	**Original Accuracy**	**Updated Accuracy**
SC1	1.05	8.12%	6.79%
SC2		N/A	
SC9		N/A	

Table 3-18 MAPE Comparison for Cluster Group Combined 15-Min Interval Data with or without Load Growth Factor

	Combined 15-Min interval data		
	Load Growth Factor	**Original Accuracy**	**Updated Accuracy**
SC1	1.05	6.84%	6.26%
SC2	1.06	5.49%	4.44%
SC9	1.04	3.38%	3.37%

From the results in above tables, we can say that based on both customer behaviors and overall load growth factor, estimation for load profile can be improved.

Chapter 4

PROPOSED METHOD APPLICATION IN TRADITIONAL METER LOAD PROFILE ESTIMATION

In current Con Edison system, many customers do not have a smart meter installed. However, it is desired to extend load profile estimation to customers without smart meters. Therefore, the current load profile estimation algorithm derived from the customers with smart meters can be applied to predict the load profile of the customers without smart meters.

Following the procedure that has been established based upon the received data of SC1 smart meter customers in 2012, the load profile forecasting are conducted to the customers in Residential Service Class 1 with traditional meters. Similar procedure is applied to SC2 and SC9. The load profile estimation is performed for these types of customers in both 2012 and 2013.

4.1 Data Preparation

The traditional meter data files provided by Con Edison are from the year 2012 through the year 2014. The information in the files consists of division, account number, service class, network, site ID, channel number, unit of measure, device ID, device channel number, device type, date, and the consumption for every 15 minutes.

The SCs in the files including SC1, SC2, SC5, SC8, SC9, SC11, SC12, SC15, SC39, SC45, SC50, SC51, SC52, SC58, SC67, SC87, SC116, SC240, SC267, SC414, SC467, SC901. For side-by-side comparison, we plan to perform load profile analysis on SC1, SC2, and SC9 customers, they are extracted from traditional meter data files.

Figure 4-1 shows the raw information from the traditional meter file. There are 96 data point in each row to represent the 15-min energy consumption for each day. Only I1 thru I9 are shown for illustration.

Based upon SC, UOM, and ChanNum, the filtered data are shown in Figure 4-2. The updated data information will be used for traditional meter load profile analysis.

Division	Account	erviceClas	Network	SITEID	ChanNum	UnitOfMeasure	DeviceId	viceChanN	DeviceType	Day	I1	I2	I3	I4	I5	I6	I7	I8	I9
QN	2.95E+13	8	RP	9919	20	kVARh	7929803	2	KV2C	4/1/2012	2.24	2.36	2.4	2.4	2.2	2.36	2.52	2.48	2.24
QN	2.95E+13	8	RP	9919	20	kVARh	7929803	2	KV2C	4/2/2012	2.56	2.52	2.6	2.48	2.48	2.4	2.32	2.28	2.24
QN	2.95E+13	8	RP	9919	20	kVARh	7929803	2	KV2C	4/3/2012	2.24	2.08	2.12	2.2	2.32	2.32	2.24	2.28	2.2
QN	2.95E+13	8	RP	9919	20	kVARh	7929803	2	KV2C	4/4/2012	2.48	2.4	2.36	2.4	2.52	2.36	2.32	2.32	2.32
QN	2.95E+13	8	RP	9919	20	kVARh	7929803	2	KV2C	4/5/2012	2.28	2.28	2.28	2.28	2.36	2.28	2.2	2.2	2.2
QN	2.95E+13	8	RP	9919	20	kVARh	7929803	2	KV2C	4/6/2012	2.32	2.32	2.28	2.24	2.32	2.28	2.32	2.28	2.2
QN	2.95E+13	8	RP	9919	20	kVARh	7929803	2	KV2C	4/7/2012	2.56	2.48	2.48	2.44	2.36	2.36	2.36	2.32	2.28
QN	2.95E+13	8	RP	9919	20	kVARh	7929803	2	KV2C	4/8/2012	2.48	2.48	2.44	2.36	2.36	2.44	2.4	2.36	2.28
QN	2.95E+13	8	RP	9919	20	kVARh	7929803	2	KV2C	4/9/2012	2.48	2.4	2.4	2.36	2.28	2.32	2.24	2.24	2.16
QN	2.95E+13	8	RP	9919	20	kVARh	7929803	2	KV2C	4/10/2012	2.32	2.32	2.32	2.36	2.36	2.32	2.28	2.32	2.4
QN	2.95E+13	8	RP	9919	20	kVARh	7929803	2	KV2C	4/11/2012	2.32	2.32	2.28	2.32	2.24	2.12	2.24	2.24	2.2
QN	2.95E+13	8	RP	9919	20	kVARh	7929803	2	KV2C	4/12/2012	2.56	2.44	2.52	2.4	2.52	2.52	2.44	2.48	2.44
QN	2.95E+13	8	RP	9919	20	kVARh	7929803	2	KV2C	4/13/2012	2.6	2.68	2.64	2.6	2.6	2.56	2.52	2.48	2.48
QN	2.95E+13	8	RP	9919	20	kVARh	7929803	2	KV2C	4/14/2012	2.6	2.56	2.56	2.56	2.4	2.48	2.56	2.64	2.6
QN	2.95E+13	8	RP	9919	20	kVARh	7929803	2	KV2C	4/15/2012	2.64	2.68	2.76	2.6	2.6	2.76	2.64	2.6	2.52
QN	2.95E+13	8	RP	9919	20	kVARh	7929803	2	KV2C	4/16/2012	2.92	2.92	2.92	2.8	2.88	2.8	2.8	2.84	2.84
QN	2.95E+13	8	RP	9919	20	kVARh	7929803	2	KV2C	4/17/2012	2.64	2.64	2.72	2.8	2.8	2.68	2.6	2.56	2.56
QN	2.95E+13	8	RP	9919	20	kVARh	7929803	2	KV2C	4/18/2012	2.64	2.76	2.88	2.84	2.76	2.8	2.64	2.6	2.76
QN	2.95E+13	8	RP	9919	20	kVARh	7929803	2	KV2C	4/19/2012	2.24	2.16	2.2	2.28	2.2	2.2	2.2	2.08	2.16
QN	2.95E+13	8	RP	9919	20	kVARh	7929803	2	KV2C	4/20/2012	2.2	2.16	2.2	2.2	2.12	2.12	2.08	2.2	2.12
QN	2.95E+13	8	RP	9919	20	kVARh	7929803	2	KV2C	4/21/2012	2.04	2.16	2.12	2	2.08	2.08	2.08	2.04	2
QN	2.95E+13	8	RP	9919	20	kVARh	7929803	2	KV2C	4/22/2012	2.44	2.32	2.16	2.04	2.16	2.12	2.16	2.16	2.2
QN	2.95E+13	8	RP	9919	20	kVARh	7929803	2	KV2C	4/23/2012	1.88	1.88	1.88	1.8	1.84	1.84	1.88	1.88	1.88
QN	2.95E+13	8	RP	9919	20	kVARh	7929803	2	KV2C	4/24/2012	2.2	2.2	2.16	2.2	2.08	2.2	2.2	2.04	2.08
QN	2.95E+13	8	RP	9919	20	kVARh	7929803	2	KV2C	4/25/2012	2.4	2.48	2.48	2.48	2.48	2.44	2.44	2.4	2.44
QN	2.95E+13	8	RP	9919	20	kVARh	7929803	2	KV2C	4/26/2012	2.52	2.56	2.48	2.36	2.4	2.28	2.28	2.32	2.36
QN	2.95E+13	8	RP	9919	20	kVARh	7929803	2	KV2C	4/27/2012	2.76	2.76	2.8	2.68	2.68	2.56	2.64	2.64	2.64
QN	2.95E+13	8	RP	9919	20	kVARh	7929803	2	KV2C	4/28/2012	2.12	2.08	2.12	2.16	2.08	2.2	2.12	2.08	1.96
QN	2.95E+13	8	RP	9919	20	kVARh	7929803	2	KV2C	4/29/2012	2.36	2.44	2.44	2.36	2.36	2.4	2.36	2.28	2.32
QN	2.95E+13	8	RP	9919	20	kVARh	7929803	2	KV2C	4/30/2012	2.32	2.32	2.24	2.36	2.24	2.28	2.28	2.32	2.28
QN	2.95E+13	8	RP	9919	21	kWh	7929803	1	KV2C	4/1/2012	7.6	7.68	7.12	7.4	7.08	7.24	7.28	7.12	6.84
QN	2.95E+13	8	RP	9919	21	kWh	7929803	1	KV2C	4/2/2012	8.32	8.16	8.16	8.08	7.4	7.24	7.04	6.92	7
QN	2.95E+13	8	RP	9919	21	kWh	7929803	1	KV2C	4/3/2012	7.48	7.2	6.92	6.72	6.76	6.72	6.72	6.76	6.44
QN	2.95E+13	8	RP	9919	21	kWh	7929803	1	KV2C	4/4/2012	7.28	7.4	7.12	7.4	7.12	7.12	6.76	6.56	6.6
QN	2.95E+13	8	RP	9919	21	kWh	7929803	1	KV2C	4/5/2012	8.32	8.08	7.84	7.84	7.72	7.48	7.36	7.2	7.12
QN	2.95E+13	8	RP	9919	21	kWh	7929803	1	KV2C	4/6/2012	7.96	7.84	7.52	7.48	7.32	7.16	7.36	7.28	7.16
QN	2.95E+13	8	RP	9919	21	kWh	7929803	1	KV2C	4/7/2012	8.08	7.68	7.28	7.16	7.08	6.92	6.96	7.24	7.16
QN	2.95E+13	8	RP	9919	21	kWh	7929803	1	KV2C	4/8/2012	8.48	8.32	8.2	7.76	7.72	7.76	8	7.88	7.92
QN	2.95E+13	8	RP	9919	21	kWh	7929803	1	KV2C	4/9/2012	7.96	7.8	7.84	7.76	7.56	7.56	7.44	7.44	7
QN	2.95E+13	8	RP	9919	21	kWh	7929803	1	KV2C	4/10/2012	7.52	7.12	7.12	6.92	6.68	6.12	6.44	6.24	6.52
QN	2.95E+13	8	RP	9919	21	kWh	7929803	1	KV2C	4/11/2012	7.72	7.48	7.44	7.52	7.2	6.84	6.72	6.52	6.56
QN	2.95E+13	8	RP	9919	21	kWh	7929803	1	KV2C	4/12/2012	8	8.44	8.32	8.04	7.96	7.8	7.56	7.52	7.2
QN	2.95E+13	8	RP	9919	21	kWh	7929803	1	KV2C	4/13/2012	8.6	8.2	8.04	7.88	7.84	7.44	7.48	7.36	7.28

Figure 4-3 Original Information from Traditional Meter File

Divisio	Accoun	erviceC	Netwo	SITEID	ChanNu	itOfMea	DeviceId	eviceChar	DeviceTy	Day	I1	I2	I3	I4	I5	I6	I7	I8	I9
BX	3.011E+13	1	NE	13058	1	kWh	8132570	1	KV2C	4/1/2012	0.003	0.003	0.006	0.003	0.003	0.003	0.003	0.006	0.003
BX	3.011E+13	1	NE	13058	1	kWh	8132570	1	KV2C	4/2/2012	0.003	0.006	0.003	0.003	0.003	0.003	0.006	0.003	0.003
BX	3.011E+13	1	NE	13058	1	kWh	8132570	1	KV2C	4/3/2012	0.003	0.006	0.003	0.003	0.003	0.003	0.006	0.003	0.003
BX	3.011E+13	1	NE	13058	1	kWh	8132570	1	KV2C	4/4/2012	0.003	0.003	0.003	0.006	0.003	0.003	0.003	0.006	0.003
BX	3.011E+13	1	NE	13058	1	kWh	8132570	1	KV2C	4/5/2012	0.003	0.003	0.006	0.003	0.003	0.003	0.003	0.003	0.006
BX	3.011E+13	1	NE	13058	1	kWh	8132570	1	KV2C	4/6/2012	0.003	0.003	0.006	0.003	0.003	0.003	0.006	0.003	0.003
BX	3.011E+13	1	NE	13058	1	kWh	8132570	1	KV2C	4/7/2012	0.003	0.003	0.003	0.006	0.003	0.003	0.003	0.003	0.006
BX	3.011E+13	1	NE	13058	1	kWh	8132570	1	KV2C	4/8/2012	0.087	0.051	0.051	0.051	0.102	0.102	0.06	0.042	0.048
BX	3.011E+13	1	NE	13058	1	kWh	8132570	1	KV2C	4/9/2012	0.015	0.006	0.006	0.03	0.018	0.006	0.003	0.027	0.021
BX	3.011E+13	1	NE	13058	1	kWh	8132570	1	KV2C	4/10/2012	0.024	0.006	0.018	0.027	0.006	0.006	0.036	0.006	0.006
BX	3.011E+13	1	NE	13058	1	kWh	8132570	1	KV2C	4/11/2012	0.012	0.036	0.072	0.027	0.036	0.054	0.045	0.042	0.06
BX	3.011E+13	1	NE	13058	1	kWh	8132570	1	KV2C	4/12/2012	0.003	0.003	0.024	0.009	0.003	0.003	0.027	0.006	0.003
BX	3.011E+13	1	NE	13058	1	kWh	8132570	1	KV2C	4/13/2012	0.033	0.033	0.072	0.126	0.072	0.156	0.111	0.018	0.021
BX	3.011E+13	1	NE	13058	1	kWh	8132570	1	KV2C	4/14/2012	0.021	0.054	0.057	0.012	0.021	0.027	0.006	0.021	0.021
BX	3.011E+13	1	NE	13058	1	kWh	8132570	1	KV2C	4/15/2012	0.006	0.003	0.021	0.015	0.003	0.039	0.033	0.015	0.021
BX	3.011E+13	1	NE	13058	1	kWh	8132570	1	KV2C	4/16/2012	0.024	0.033	0.012	0.012	0.036	0.018	0.003	0.027	0.03
BX	3.011E+13	1	NE	13058	1	kWh	8132570	1	KV2C	4/17/2012	0.036	0.021	0.021	0.03	0.027	0.036	0.027	0.036	0.06
BX	3.011E+13	1	NE	13058	1	kWh	8132570	1	KV2C	4/18/2012	0.012	0.003	0.03	0.015	0.03	0.024	0.021	0.036	0.024
BX	3.011E+13	1	NE	13058	1	kWh	8132570	1	KV2C	4/19/2012	0.036	0.057	0.03	0.027	0.054	0.039	0.03	0.042	0.054
BX	3.011E+13	1	NE	13058	1	kWh	8132570	1	KV2C	4/20/2012	0.033	0.078	0.099	0.066	0.066	0.051	0.015	0.003	0.027
BX	3.011E+13	1	NE	13058	1	kWh	8132570	1	KV2C	4/21/2012	0.102	0.111	0.084	0.069	0.075	0.069	0.096	0.099	0.066
BX	3.011E+13	1	NE	13058	1	kWh	8132570	1	KV2C	4/22/2012	0.072	0.066	0.111	0.066	0.081	0.063	0.045	0.06	0.054
BX	3.011E+13	1	NE	13058	1	kWh	8132570	1	KV2C	4/23/2012	0.021	0.009	0.009	0.036	0.018	0.009	0.006	0.039	0.015
BX	3.011E+13	1	NE	13058	1	kWh	8132570	1	KV2C	4/24/2012	0.021	0.033	0.012	0.027	0.063	0.039	0.018	0.012	0.048
BX	3.011E+13	1	NE	13058	1	kWh	8132570	1	KV2C	4/25/2012	0.057	0.048	0.084	0.081	0.081	0.066	0.045	0.102	0.039
BX	3.011E+13	1	NE	13058	1	kWh	8132570	1	KV2C	4/26/2012	0.015	0.015	0.042	0.021	0.015	0.015	0.018	0.045	0.027
BX	3.011E+13	1	NE	13058	1	kWh	8132570	1	KV2C	4/27/2012	0.036	0.021	0.015	0.03	0.024	0.015	0.018	0.039	0.015
BX	3.011E+13	1	NE	13058	1	kWh	8132570	1	KV2C	4/28/2012	0.015	0.021	0.015	0.042	0.015	0.015	0.015	0.039	0.018
BX	3.011E+13	1	NE	13058	1	kWh	8132570	1	KV2C	4/29/2012	0.033	0.033	0.015	0.015	0.015	0.033	0.03	0.015	0.015
BX	3.011E+13	1	NE	13058	1	kWh	8132570	1	KV2C	4/30/2012	0.015	0.024	0.063	0.054	0.036	0.006	0.006	0.018	0.057
BX	3.253E+13	1	RI	11563	1	kWh	8087394	1	KV2C	4/1/2012	0.108	0.117	0.141	0.096	0.126	0.129	0.117	0.081	0.096
BX	3.253E+13	1	RI	11563	1	kWh	8087394	1	KV2C	4/2/2012	0.045	0.054	0.054	0.018	0.018	0.057	0.039	0.018	0.033
BX	3.253E+13	1	RI	11563	1	kWh	8087394	1	KV2C	4/3/2012	0.063	0.033	0.03	0.054	0.024	0.018	0.042	0.048	0.018
BX	3.253E+13	1	RI	11563	1	kWh	8087394	1	KV2C	4/4/2012	0.018	0.024	0.057	0.03	0.018	0.039	0.072	0.099	0.018
BX	3.253E+13	1	RI	11563	1	kWh	8087394	1	KV2C	4/5/2012	0.033	0.054	0.027	0.015	0.048	0.048	0.018	0.021	0.057
BX	3.253E+13	1	RI	11563	1	kWh	8087394	1	KV2C	4/6/2012	0.084	0.069	0.096	0.096	0.018	0.027	0.054	0.036	0.018
BX	3.253E+13	1	RI	11563	1	kWh	8087394	1	KV2C	4/7/2012	0.051	0.048	0.015	0.021	0.057	0.036	0.018	0.03	0.057
BX	3.253E+13	1	RI	11563	1	kWh	8087394	1	KV2C	4/8/2012	0.024	0.018	0.045	0.057	0.018	0.018	0.069	0.063	0.054
BX	3.253E+13	1	RI	11563	1	kWh	8087394	1	KV2C	4/9/2012	0.114	0.03	0.033	0.057	0.039	0.039	0.042	0.051	0.018
BX	3.253E+13	1	RI	11563	1	kWh	8087394	1	KV2C	4/10/2012	0.018	0.054	0.048	0.018	0.021	0.057	0.042	0.018	0.027
BX	3.253E+13	1	RI	11563	1	kWh	8087394	1	KV2C	4/11/2012	0.027	0.057	0.036	0.018	0.039	0.054	0.021	0.018	0.054
BX	3.253E+13	1	RI	11563	1	kWh	8087394	1	KV2C	4/12/2012	0.042	0.081	0.042	0.018	0.039	0.054	0.018	0.018	0.057
BX	3.253E+13	1	RI	11563	1	kWh	8087394	1	KV2C	4/13/2012	0.105	0.099	0.132	0.123	0.099	0.108	0.057	0.018	0.018
BX	3.253E+13	1	RI	11563	1	kWh	8087394	1	KV2C	4/14/2012	0.144	0.135	0.171	0.156	0.141	0.168	0.162	0.12	0.105

Figure 4-4 Updated Data Information after Filtered by Criteria

4.1.1 Aggregate Traditional Meter Customer into Monthly Consumption

According to the criteria for each stratification in SC1, the annual consumption of each meter should be known; similarly, the total consumption from June to September for every SC2 customer and the average consumption from June to September for every SC9 customer should also be known.

Since some other traditional meters only provide monthly reading, we have aggregated the traditional meter data into monthly consumption for customers in SC1, SC2, and SC9 to mimic this situation.

4.1.2 Based upon the Definition of Con Edison, Assign Stratification to Each Customer in Its Service Class

According to the definition of the stratifications in each service class shown as Table 3-2, Table 3-3 and Table 3-7, the customers come from SC1, SC2 and SC9 are assigned to the corresponding stratification.

4.2 Load Profile Estimation Algorithm Development with Traditional Meter Data

From the traditional meter file information, the electricity for each customer is 15-minute interval. Similar to task one, k-means clustering are applied to each stratification within each SC and directly applied to each SC.

Besides customer number, alien customers e.g. customer with very large consumption, will bring negative effect on forecast algorithm performance. The alien customers should be separated from the whole group of customers and be analyzed separately.

4.2.1 Results Comparison

4.2.1.1 Results for Applying K-Means Clustering in Each Stratification

Due to the small number for SC1 and SC2, the k-means clustering does not perform satisfied for forecast algorithm. Using customer behaviors to estimate their monthly billing would lead to high error.

Table 4-5 MAPE Comparisons for SCs with and without Clustering and Monthly Billing

	Without	**With**	**For Monthly Billing Cycle**
SC1 (k=6)	27.86%	28.83%	36.26%
SC2 (k=5)	32.62%	32.37%	41.42%
SC9 (k=5)	6.25%	5.89%	15.72%

4.2.1.2 Results for Applying K-Means Clustering in each SC

The forecast algorithm with k-means owns higher accuracy than no k-means. The accuracies are listed in Table 4-2.

Table 4-6 MAPE Comparison for SCs with and without Clustering

	Without Clustering	With Clustering
SC1 (k=6)	18.99%	18.30%
SC2 (k=5)	24.32%	17.08%
SC9 (k=6)	5.99%	5.22%

4.3 Load Profile Estimation Algorithm Development with Combined Data from Smart Meter and Traditional Meter Data

In order to see whether increasing the numbers of the customer can help the estimation algorithm development, smart meter and traditional meter data are combined to conduct the load profile forecast.

4.3.1 Apply K-Means Clustering for Each Stratification within Each SC

To evaluate the potential estimation accuracy improvement by applying k-means clustering application in each stratification, the customers from same stratification for smart meter and traditional meter are combined. The detailed procedures for stratification A are shown in Figure 4-1 and Figure 4-2.

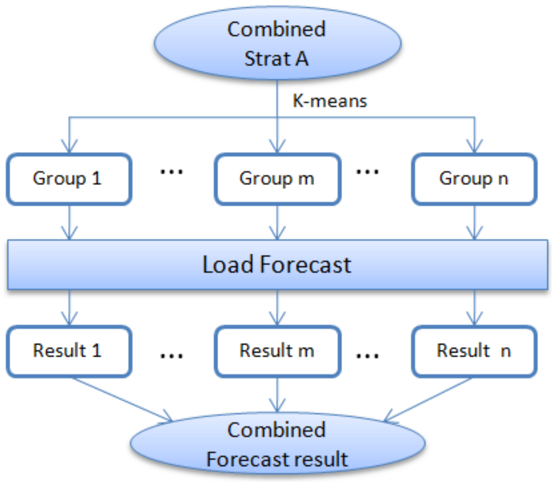


Figure 4-3 Perform load profile forecast estimation algorithm with k-means clustering on each stratification

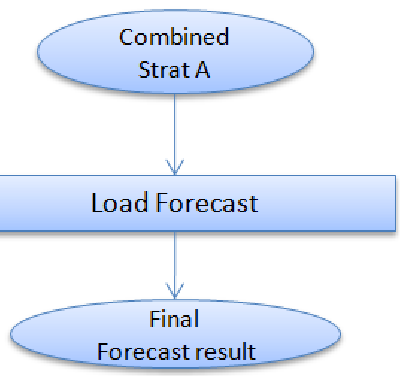


Figure 4-4 Perform load profile forecast estimation algorithm for each stratification without clustering

The forecast results with k-means clustering within each stratification for SC1 and SC2 are not satisfied due to the small number of customers in each stratification. The results for SC9 are shown in the table below:

Table 4-5 MAPE Comparison for SC9

	With Clustering	Without Clustering
SC9	3.62%	5.83%

4.3.2 Apply K-Means Clustering within Each SC

To evaluate the potential estimation accuracy improvement by applying k-means clustering application in each SC, the customers from same SC for smart meter and traditional meter are combined. The detailed procedures for SC1 are shown in Figure 4-6 and Figure 4-7. The same procedure is applied to other SCs.

Load profile estimation algorithm development by means of k-means clustering shows better accuracies than no k-means application one. The results are listed in Table 4-5.

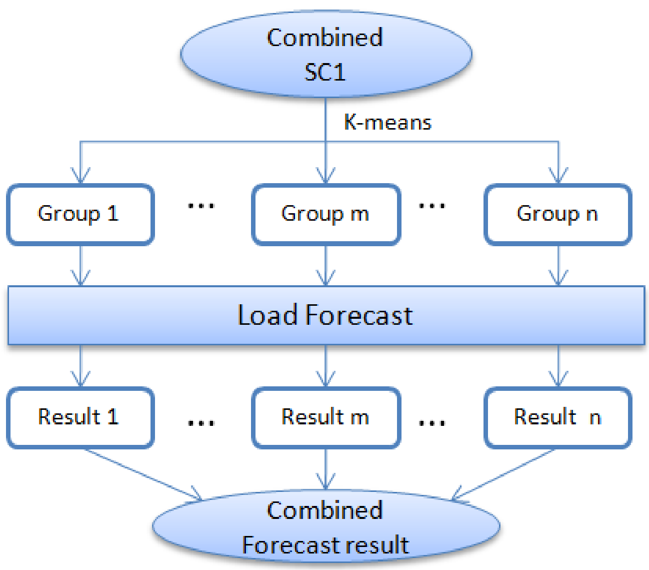


Figure 4-8 Perform load profile forecast algorithm with k-means clustering on each SC

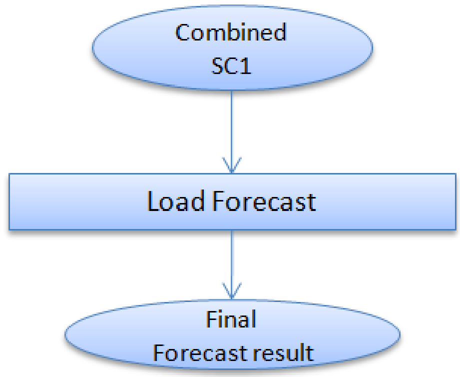


Figure 4-9 Perform load profile forecast algorithm for each SC without clustering

With k-means clustering, the forecast accuracy for SC2 is improved and is shown in Table 4-4.

Table 4-10 MAPE Comparison for SC2

	With Clustering	**Without Clustering**
SC2	12.47%	24.31%

However, no improvement has been observed for SC1 and SC9 with k-means clustering due to the presence of several alien customers in SC1 and SC9. Alien customers are the customers have so different behaviors or have so large amount of consumptions compared other customers. The load profiles for SC1 customers and SC9 customers are shown in Figure 4-11 and Figure 4-12 respectively. One can see that the blue, green and light green lines are alien customer for SC1 and the green and pink lines are the alien customer for SC9.

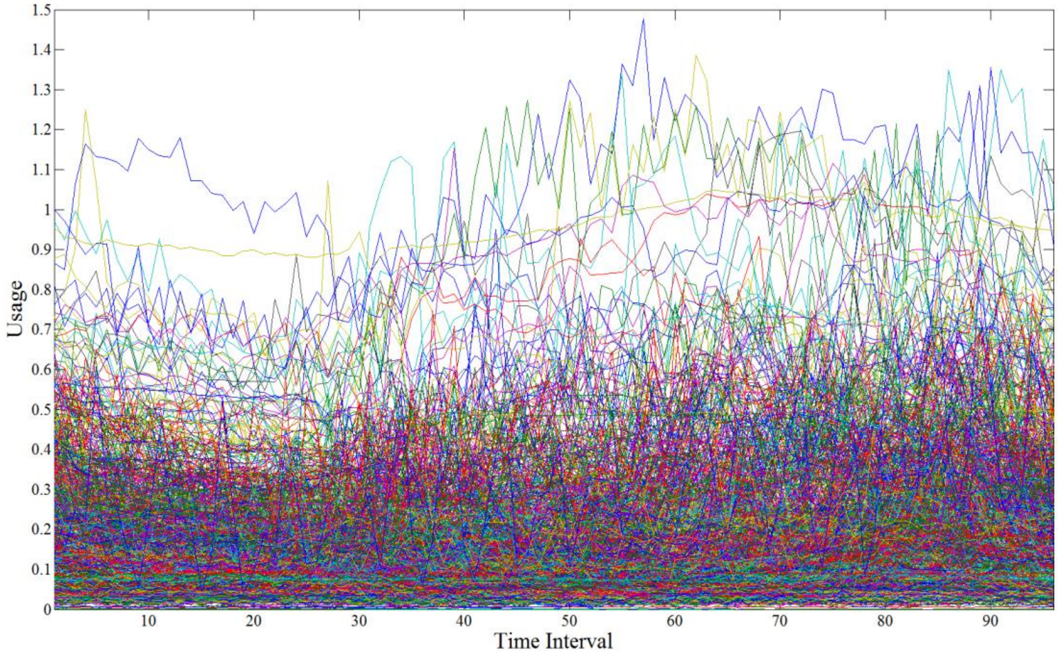


Figure 4-13 SC1 Customers Load Profiles

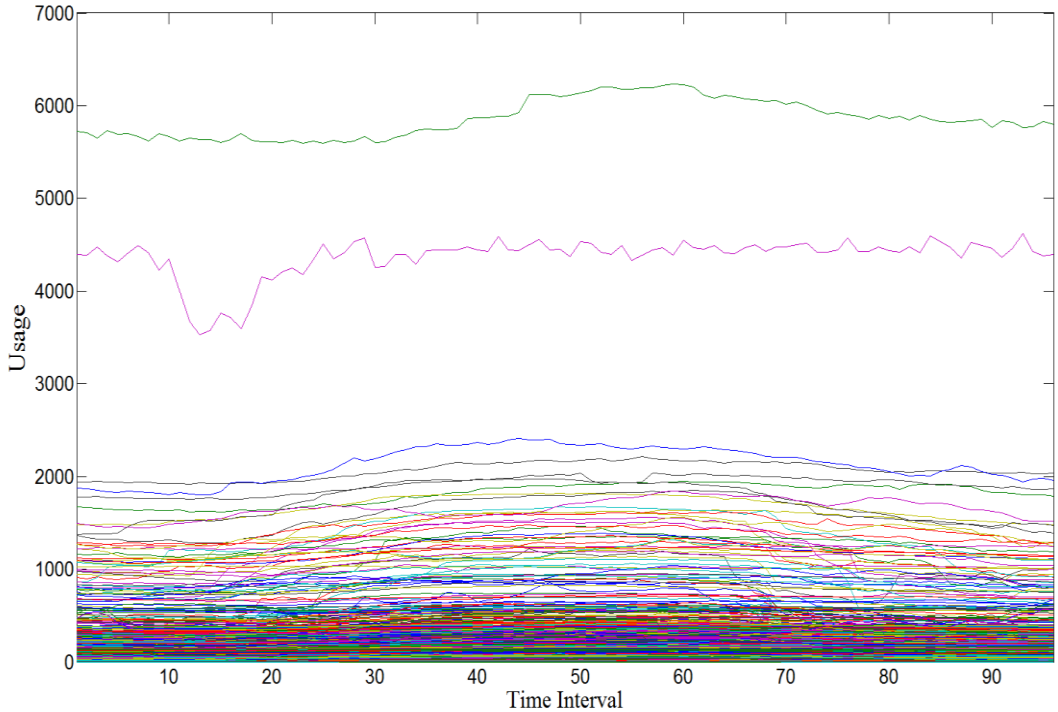


Figure 4-14 SC9 Customers Load Profiles

To overcome this issue, the alien customers are separated from the rest of the customers and the k-means clustering is applied to the rest of the customers. We then perform the load forecast on the alien customer individually and on each group after applying k-means. The individual results are combined together to obtain the final prediction outcome. The detailed procedure is described in Figure 4-15.

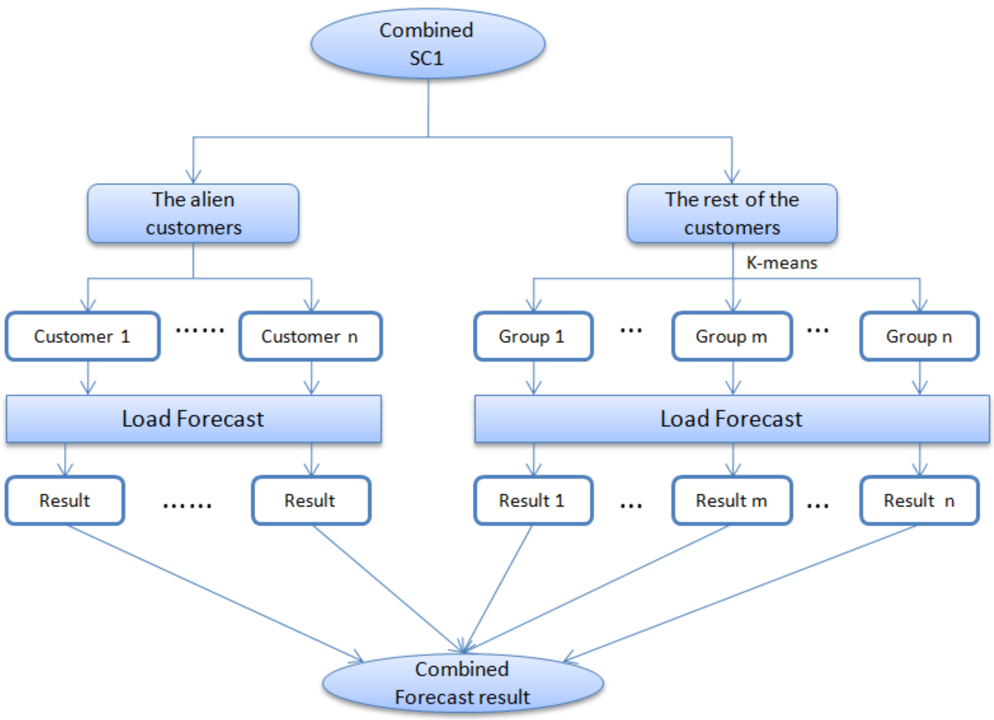


Figure 4-16 Detail procedures to perform load profile forecast algorithm with alien customers

Updated with this process, the algorithm shows improvement compared with no k-means. The results for SC1 and SC9 are shown as below:

Table 4-17 MAPE Comparison for SC1 and SC9

	With Clustering	**Without Clustering**
SC1	10.99%	14.93%
SC9	5.22%	5.99%

4.4 Load Profile Estimation Algorithm Development with Aggregate Traditional Meter Data

In current Con Edison system, many customers do not have a smart meter installed. It is desired to extend load profile estimation algorithm to customers without smart meters. It is assumed that traditional meter customers' consumption can only be obtained from the monthly billing data. This part of the research explores the possibility to forecast the aggregate traditional meter data by using the developed forecast algorithm.

4.4.1 Apply Weighting Factor to Establish Combined Load Profiling for Load Forecasting on Aggregate Traditional Meter Data at Stratification Level

The weighting factor here means the percentage of that group of customer versus the certain stratification customer number. For example, there are 200 customers in stratification A within SC1, and group 1 in stratification A has 50 customers, then the weighting factor for group 1 in stratification A is 50/200=0.25. The method for forecasting on the aggregate traditional meter data is that if the traditional meter customer belongs to stratification A, then multiply the weighting factor with the corresponding group load profile within stratification A and sum up all the groups within stratification A to obtain the combined load profile for stratification A. Repeat the same procedure for stratification B through stratification F for SC1 traditional meter data. Apply the same procedure for stratification A through stratification E for SC2 and SC9 traditional meter data. The procedure is shown in the Figure 4-18:

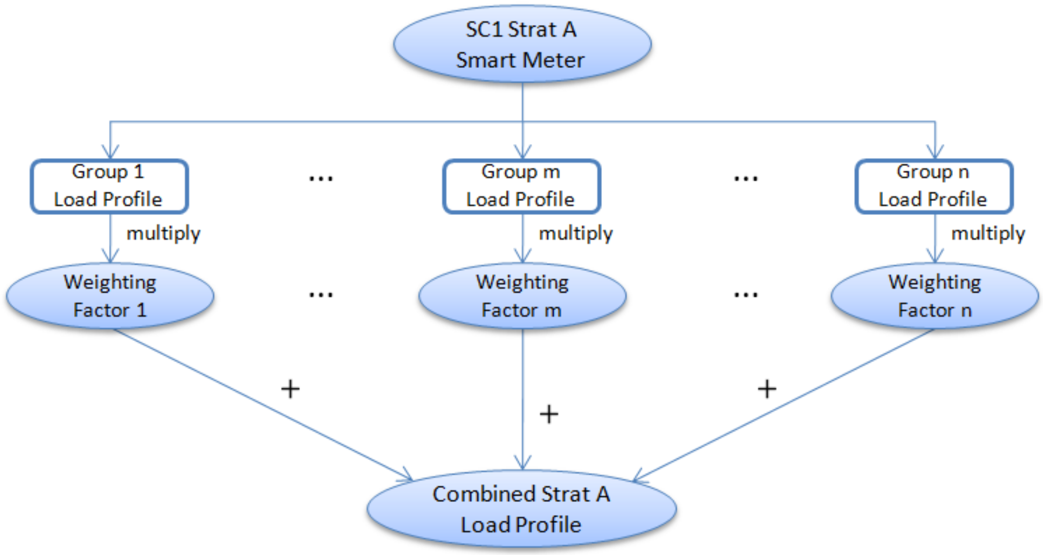


Figure 4-19 Apply Weighting Factor to Establish Combined Load Profiling at Stratification Level

4.4.2 Apply Weighting Factor to Establish Combined Load Profiling for Load Forecasting on Aggregate Traditional Meter Data at SC Level

The weighting factor here is similar as last part, the difference is that it's the percentage of that group of customer versus the certain SC customer number. The method for forecast on the aggregate traditional meter data is that if the traditional meter customer belongs to certain SC, then multiply the weighting factor with the corresponding group load profile within that SC and sum up all the groups within that SC to obtain the combined load profile. The procedure is shown in the Figure 4-20.

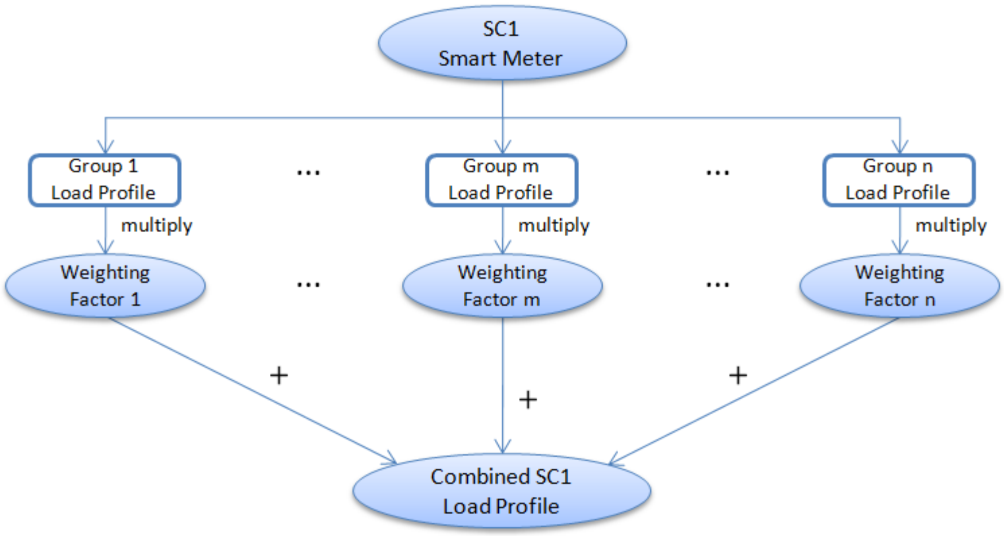


Figure 4-21 Apply Weighting Factor to Establish Combined Load Profiling at SC Level

4.4.3 Compare the Accuracy with the Actual Traditional Meter Load Profile

The accuracies of each stratification for SC1 are shown in the Table 4-6.

Table 4-22 MAPE of Traditional Meter Load Profile Forecasting for SC1

A	**B**	**C**	**D**	**E**	**F**
46.09%	27.85%	29.4%	4.79%	12.41%	25.93%

For high MAPE value groups in above table, the trends of smart meter data and traditional meter data are similar. However, the differences between the consumption are large.

The accuracies of each stratification for SC2 are shown in the Table 4-7.

Table 4-23 MAPE of Traditional Meter Load Profile Forecasting for SC2

A	**B**	**C**	**D**	**E**	**F**
68.59%	34.99%	37.61%	20.87%	99.53%	68.59%

For stratification A, the consumption of smart meter customers is nearly twice of the traditional one; For stratification B to D, the consumptions of traditional meter customers are nearly twice of the smart meter customers; For stratification E, the consumption of traditional meter customers are much higher than the smart meter customers.

In part 4.4.1, the MAPE values of each stratification are too large for all SC9 traditional meter. That is because the consumption of traditional meter customers are much larger than the corresponding smart meter ones, which makes the results for this part unsatisfactory.

In part 4.4.2, the accuracy for SC1 is 13.053%, traditional meter customers' electricity consumptions are less than the ones from smart meter customers.

In part 4.4.2, the result for SC2 is not good. The reason for the unsatisfied results is that, SC2 traditional meter customers' electricity consumptions are too large compared with its smart meter customers.

In part 4.4.2, similar to SC2, we could not obtain satisfactory result for SC9. Because SC9 traditional meter customers' electricity consumptions are too large. This difference between traditional meter data and smart meter data for SC9 is even larger than the difference between traditional meter data and smart meter data in the SC2.

Chapter 5

REFINE SERVICE DEMAND ESTIMATORS SLOPE AND INTERCEPT – PEAK DEMAND ESTIMATION

In Con Edison, the Service Demand Estimator uses a slope and intercept to estimate peak demand from annual usage. The existing slope and intercept values were calculated from a few points available from the Rate Department report. The Smart Meter data provides the opportunity to refine these slopes and intercept values. To refine service demand estimators slope and intercept, it is important to understand the original design philosophies and approaches of Service Demand Estimator.

The spreadsheet Regression & Spline data for Service Class Estimation.xls contains the service class, stratification group, and total annual kilowatt hours for different service classes plus June to September kilowatt-hours for SC2. Besides, the peak day is July 17 for the year of 2012 and July 18 for the year of 2013. With all these information, it's enough to determine a slope and intercept for each service class.

In this dissertation, both stratification grouping and K-means grouping are used to generate necessary information for the service demand estimators slope and intercept. When using stratification grouping approach, the breakpoints at the boundary of the groups will have to match. Linear regression method is applied to minimize the mismatch at the breakpoints. Though annual total consumption is the base of current stratification grouping, the consumption patterns are not determined by annual total consumption. Therefore, the K-means clustering is adopted in this project to perform the customer segregation/regrouping within each stratification group based upon their consumption patterns and then re-integrate them based upon each group's aggregated total load to refine Service Demand Estimators.

The other approach is conducting k-means clustering to perform the customer segregation within each service class based upon their consumption patterns and then re-integrate them based upon each group's aggregated load to refine service demand estimators.

5.1 Refine the Parameters with Smart Meter Data

Only smart meter data is used to refine the parameters in this section.

5.1.1 Refine the Parameters for Each SC

According to the design philosophies, the peak demand forecast is different for SC1, SC2, and SC9.

In order to refine the parameters for each SC, the updated new data should be obtained for calculation. The calculation is shown in the regression & spline data for SC1&2 estimation.xls (SC9 is not shown in the original file). T_A_kWHR represents for Total_Annual_kWHR and T_U represents for Total_Usage for Table 5-1 and Table 5-2.

Table 5-1 Original Parameters for SC1 from Con Edison

Stratification	low	high	T_A_kWHR	m	b	y
A	0	1948	1312	0.000280438	0.0196	0.38753423
B	1949	2897	2594	0.000299298	-0.0293	0.7470784
C	2898	3897	3622	0.000197609	0.339	1.05474144
D	3898	5239	4849	0.000337584	-0.3397	1.29724272
E	5240	7741	6772	0.000204492	0.5616	1.94641794
F	7742	9999999	12051	0.000204492	0.5616	3.02592974

Table 5-2 Original Parameters for SC2 from Con Edison

Stratification	low	high	T_U Jun-Sep	m	b	y
A	0	1500	2198	0.00029379	-0.1451	0.50066486
B	1501	3000	6748	0.00029324	-0.1414	1.83741123
C	3001	5000	11618	0.00035766	-0.8898	3.26550356
D	5000	8000	18338	0.00014507	3.0087	5.66904318
E	8000	9999999	32734	0.00014507	3.0087	7.75750978

The formula for parameters m and b are listed below:

$$m_A = \frac{\text{Coin kW}_B - \text{Coin kW}_A}{\text{Ave annual}_B - \text{Ave annual}_A}$$

$$b_A = \frac{\text{Ave annual}_B * \text{Coin kW}_A - \text{Ave annual}_A * \text{Coin kW}_B}{\text{Ave annual}_B - \text{Ave annual}_A}$$

In SC1, the m for stratification B to stratification E have the similar formula as stratification A, m for stratification F is equal to m for stratification E; the b for stratification B to stratification E have the similar formula as stratification A, b for stratification F is equal to b for stratification E.

As for SC2 and SC9, the m for stratification B to stratification D have the similar formula as stratification A, m for stratification E is equal to m for stratification D; the b for stratification B to stratification D have the similar formula as stratification A, b for stratification E is equal to b for stratification D.

For SC1, the average annual kWHR usage per customer for each stratification should be calculated, the coincident peak demand for each stratification in the year 2012 (July 17th) is obtained from the smart meter data.

For SC2, the total June-September kWHR usage per customer for each stratification should be calculated, the coincident peak demand for each stratification in the year 2012 (July 17th) is obtained from the smart meter data.

For SC9, the average June-September kW usage per customer for each stratification should be calculated, the coincident peak demand for each stratification in the year 2012 (July 17th) is obtained from the smart meter data.

The following table shows the calculated parameters from recent data for this part.

Table 5-3 Refined Parameters for each SC with Con Edison Recent Data

SC1		SC2		SC9	
m	**b**	**m**	**b**	**m**	**b**
0.000339556	0.0964	0.000641862	0.1469	0.000044813	0.6029
0.000152914	0.5690	0.001303211	-1.2265	0.000040789	1.5889
0.000348915	-0.1156	-9.7285E-05	3.5845	0.000054886	-0.0281
0.000312038	0.0523	0.000527879	0.5390	0.000033607	0.2538
0.000252030	0.4191	0.000527879	0.5390	0.000033607	0.2538
0.000252030	0.4191				

5.1.2 Data Fitting for Refined Parameters

From Con Edison's original file for peak demand forecast, columns min and max are the minimum and maximum peak kW obtained from applying the slope and intercept to the stratification groups low and high bounds. There is a problem in group C and D. Group C's high Total Annual kWHR is 3897. Customers here will have a peak demand of 1.10909. Adding one additional kWHR moves to group D at 3898. Customers here will have a peak of 0.97620. It seems unlikely that increasing the annual consumption lowers the peak. Table 5-4 is for the original Con Edison file.

Table 5-4 The Limitations Calculated with Original Parameters for SC1 from Con Edison

Stratification	Low	High	b	m	Min	Max
A	0	1948	0.0196	0.00028044	0.02	0.57
B	1949	2897	-0.0293	0.0002993	0.55	0.84
C	2898	3897	0.3390	0.00019761	0.91	1.11
D	3898	5239	-0.3397	0.00033758	0.98	1.43
E	5240	7741	0.5616	0.00020449	1.63	2.14
F	7742	9999999	0.5616	0.00020449	2.14	2044917

When calculate the m and b values, the breakpoints at the boundary of the stratification or clustering groups will have to match. The linear regression method is utilized to minimize the mismatch at the breakpoints.

For SC1, with recent data, m and b values are refined and listed in the Table 5-5:

Table 5-5 The Limitations Calculated with Refined Parameters for SC1

Stratification	Low	High	b	m	Min	Max
A	0	1948	0.096	0.00034	0.10	0.76
B	1949	2897	0.569	0.000153	0.87	1.01
C	2898	3897	-0.1156	0.000349	0.90	1.24
D	3898	5239	0.0523	0.000312	1.29	1.69
E	5240	7741	0.4191	0.000252	1.74	2.37
F	7742	9999999	0.4191	0.000252	2.37	2520000

Similarly, refined results with recent received data from Con Edison also have the problem described as above. The linear regression method is used to resolve this problem. The correct results are shown in Table 5-6.

Table 5-7 The Limitations Calculated with Updated Refined Parameters for SC1

Stratification	Low	High	b	m	Min	Max
A	0	1948	0.096	0.00034	0.09	0.76
B	1949	2897	0.465	0.000153	0.76	0.91
C	2898	3897	-0.1	0.000349	0.91	1.26
D	3898	5239	0.0523	0.000312	1.27	1.69
E	5240	7741	0.4191	0.000252	1.73	2.37
F	7742	9999999	0.4191	0.000252	2.37	2520000

With the revised parameters for m and b, there is no min or max values exceed the boundaries.

The updated parameters for SC1 and SC2 after data fitting are shown in the Table 5-7. For SC9, consumptions on some lower stratification are higher than some higher stratification in the training data, which makes it hard to process this part for SC9.

Table 5-8 m&b for SC1 and SC2

SC1		SC2	
m	**b**	**m**	**b**
0.000339556	0.0964	0.00034186	0.1469
0.000152914	0.465	0.00130321	-1.2265
0.000348915	-0.1000	9.7285E-05	3.5845
0.000312038	0.0523	0.00097978	-0.7784
0.000252030	0.4191	0.00097978	-0.7784
0.000252030	0.4191		

5.1.3 Coincident Peak Demand Forecast with Original, Refined and Updated Refined Parameters

The criterion to compare the results' accuracy is the mean absolute percentage error (MAPE). For peak demand forecast, the predicted value and the actual value are in 15-minute. Table 5-8 uses SC1 (there are six groups) as an example to explain how MAPE is calculated.

Table 5-9 Procedures for MAPE Calculation

Number	Predicted	Actual	Predicted	Actual
#1	P1	A1	P1*#1	A1*#1
#2	P2	A2	P2*#2	A2*#2
#3	P3	A3	P3*#3	A3*#3
#4	P4	A4	P4*#4	A4*#4
#5	P5	A5	P5*#5	A5*#5
#6	P6	A6	P6*#6	A6*#6

$$\sum_{i=1}^{6}(P_i * \#i) = P_{Sum}$$

$$\sum_{i=1}^{6}(A_i * \#i) = A_{Sum}$$

$$MAPE = \left|\frac{A_{sum} - P_{sum}}{A_{sum}}\right| * 100$$

#1, #2... and #6 represent for the customer No. for corresponding group. A is short for actual value; P is short for predicted value.

With the MAPE calculation, the tables below show the coincident peak demand accuracies from updated refined parameters, refined parameters and the original parameters.

Table 5-10 MAPE for SC1

SC1	MAPE
Forecast with Refined Parameters with Data Fitting	17.43%
Forecast with Refined Parameters	15.98%
Forecast with Original Parameters	27.94%

Table 5-11 MAPE for SC2

SC2	MAPE
Forecast with Refined Parameters with Data Fitting	11.31%
Forecast with Refined Parameters	0.50%
Forecast with Original Parameters	56.12%

Table 5-12 MAPE for SC9

SC9	MAPE
Forecast with Refined Parameters with Data Fitting	N/A
Forecast with Refined Parameters	4.15%
Forecast with Original Parameters	N/A

From the above tables we can say, in order to increase the precision, the parameters should be refined by recent data.

5.1.4 K-means Application in Stratification to Refine the Parameters

K-means clustering is used to perform the customer segregation/regrouping within each stratification group based upon their consumption patterns and then re-integrate them based upon each group's aggregated total load to refine the parameters for peak demand forecast.

For SC1, SC2, and SC9, there are six (stratification A thru stratification F), five (stratification A thru stratification E) and five (stratification A thru stratification E) groups respectively. K-means is performed individually within each stratification. The flowcharts below describe the k-means application in refining the parameters.

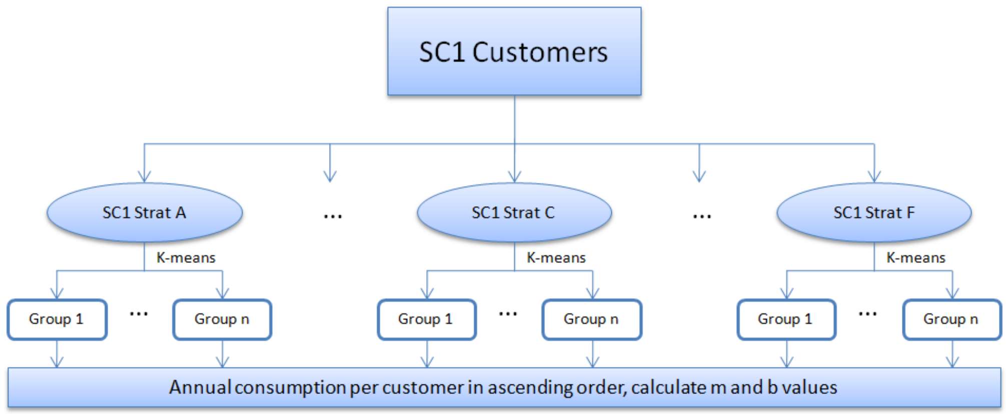


Figure 5-13 K-means Application in SC1 Stratifications

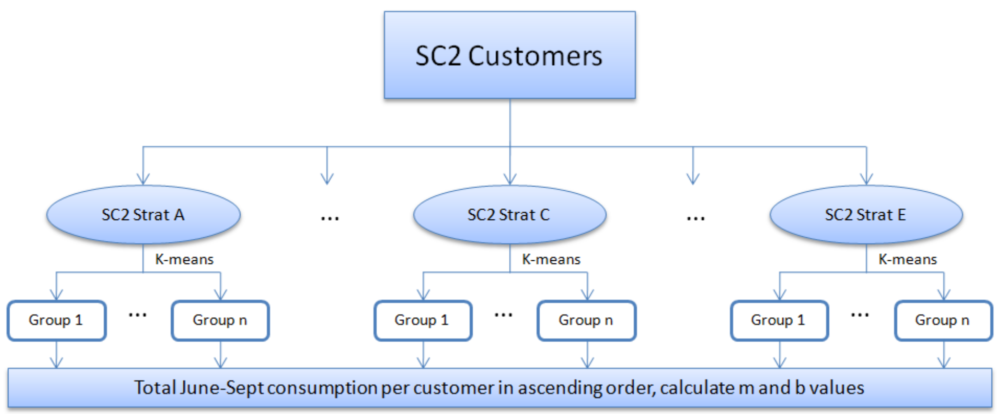


Figure 5-14 K-means Application in SC2 Stratifications

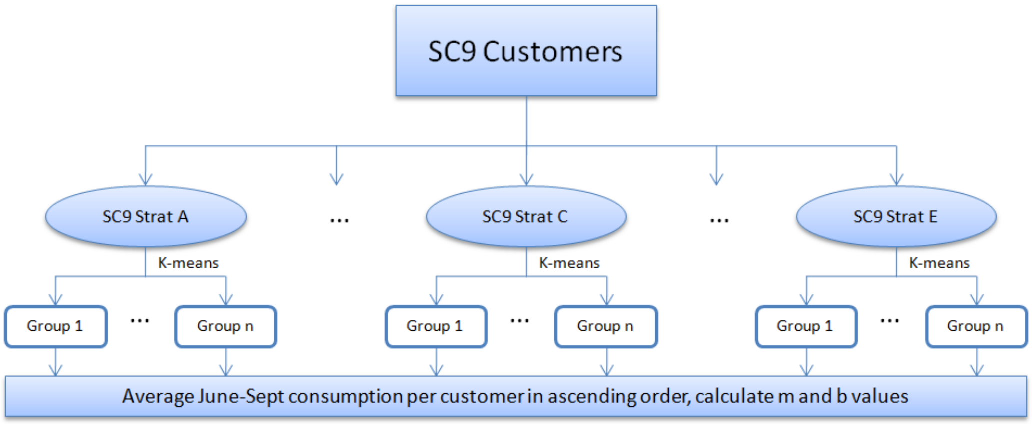


Figure 5-15 K-means Application in SC9 Stratifications

Due to the limited customer number in each stratification group, k-means performance for peak demand estimation is unsatisfied. After k-means applied within stratification, the boundaries between stratifications are uncertain. For stratification group, there is a clear consumption amount limitation between each group. However, k-means clustering is based on the customers' behavior patterns to divide the customers into different groups. So, k-means application in stratification makes it more complex for parameter calculation.

5.1.5 K-Means Application in Service Class (SC) to Refine the Parameters

Similar to part 5.1.3, apply k-means clustering in each SC to analyze the clustering performance in peak load forecast. In order to compare with no k-means stratification one, set k equals to 6, 5 and 5 for SC1, SC2 and SC9. K-means application in each SC is shown in the flowcharts below.

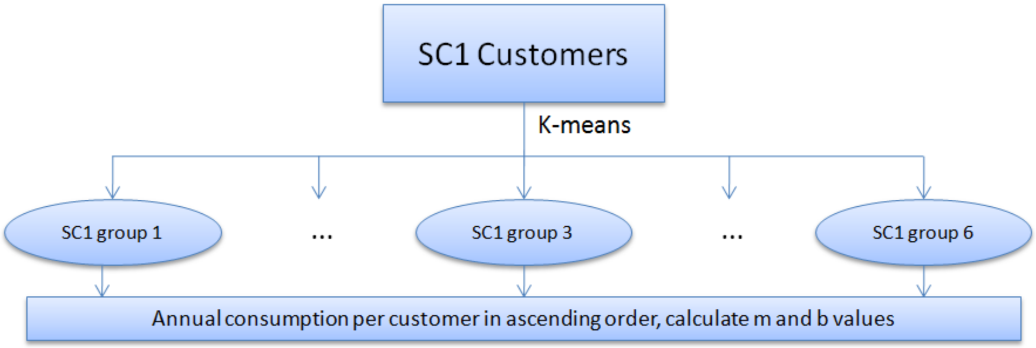


Figure 5-16 K-means Application in SC1

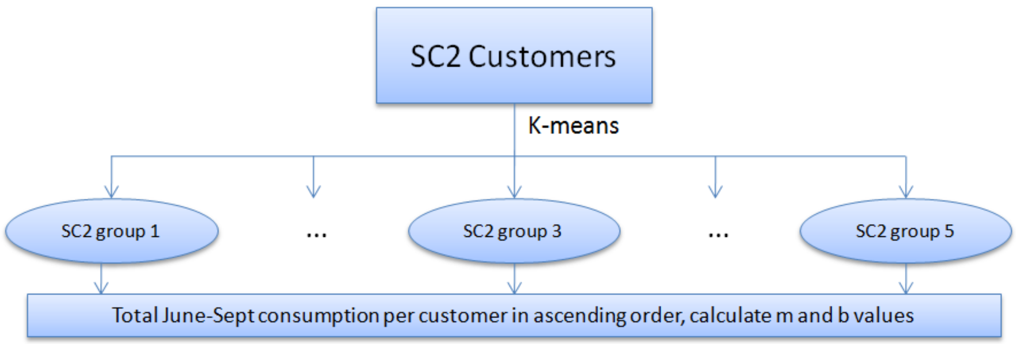


Figure 5-17 K-means Application in SC2

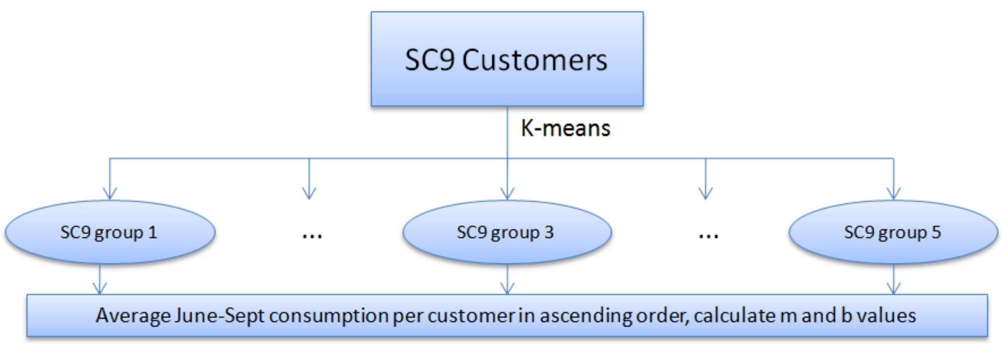


Figure 5-18 K-means Application in SC9

5.1.5.1 Results

The following table shows the parameters calculated by certain groups of customers after k-means application.

Table 5-19 Parameters for Each SC

SC1		SC2		SC9	
m	b	m	b	m	b
0.00032903	0.2185	0.000177272	-0.0296	0.297163231	-0.0222
0.000344722	0.167	0.000836028	-2.0068	0.231088316	0.0205
0.000270415	0.532	0.000113489	0.6046	0.262752217	-0.0055
0.000269871	0.5357	-0.001130958	10.0347	0.261058146	0.0048
-0.000254825	5.798	-0.001130958	10.0347	0.261058146	0.0048
-0.000254825	5.798				

The tables below present the coincident peak demand estimation accuracies for each SC.

Table 5-20 MAPE for SC1

SC1	MAPE
Forecast with Refined Parameters	2.82%

Table 5-21 MAPE for SC2

SC2	MAPE
Forecast with Refined Parameters	65.05%

Table 5-22 MAPE for SC9

SC9	MAPE
Forecast with Refined Parameters	1.15%

The original parameters from Con Edison are not applicable to this part, due to their grouping criteria's difference. For this part, grouping depends on the customer load behaviors; for the original parameters, grouping depends on the customer load consumptions. For peak demand estimation, the original philosophy with stratification to refine parameters is better.

5.1.6 Add Extra Data to Improve Coincident Peak Demand Forecast Accuracies

For both 5.1.4 and 5.1.5, the numbers for customers are limited. In this part, by applying the random number generator in Excel, extra data is used to compensate the limited number of current meter data.

SC2 is the example for illustration. The theory for this example is: setting k equals to 5 (there are 5 stratification groups within SC2), after k-means application, each group has a number for its original customers. Increase each group number to 1000 with extra data. Here are the steps:

Step one:

Generate random numbers range from 0.9 to 1.1 in Excel with 96 rows (15-min, 96 data per day) and 1000 columns (up to 1000 new data).

	A	B	C	D	E	F	G
1	0.9572	0.9375	1.0154	0.9402	1.0907	1.0588	0.9216
2	1.0239	0.9664	0.9378	1.0982	0.9789	1.0388	1.0962
3	1.0249	1.0006	0.9844	0.9775	0.9351	1.094	1.017
4	0.9571	1.0073	1.0127	1.0738	1.0182	0.961	0.9311
5	0.9696	1.0114	0.9614	1.0493	0.9614	0.9069	0.979
6	0.9218	0.9798	0.9335	1.0649	1.002	1.0408	1.0375
7	1.0398	0.9849	0.9106	0.9938	1.0429	0.9684	1.0273
8	0.9129	1.0849	0.9992	0.9892	1.0433	0.9277	0.9718
9	0.9241	1.0065	1.0251	1.0153	0.9942	0.9082	0.9865
10	1.0637	1.0005	0.9511	1.0965	0.9956	1.0998	0.9449
11	1.0875	0.9585	1.042	0.9795	1.0167	1.0208	0.9932
12	0.9403	1.0798	0.9901	0.9589	0.9907	1.0068	1.0911
13	0.9842	1.0256	1.0607	0.9027	1.0016	1.0602	1.0627
14	0.9331	0.9778	1.0941	1.0013	1.0412	0.9184	0.9744
15	1.0378	1.0374	1.0527	1.0758	1.0825	0.9215	1.0535
16	1.0876	1.0087	0.9398	0.9441	1.0608	1.0089	0.9048
17	1.0666	1.0682	0.9203	0.9149	0.9789	0.9506	0.9939
18	1.0809	0.9287	0.9056	0.9439	1.0981	1.0872	0.9084
19	1.0179	0.9596	0.9388	0.9527	1.0318	0.9684	0.9864
20	0.937	0.966	1.0649	1.027	1.046	1.0662	1.0642
21	0.9776	1.0863	1.0992	0.9092	1.0146	1.0101	1.0168
22	•••	•••	•••	•••	•••	•••	•••

Figure 5-23 Generate Random Numbers with Excel

Step two:

Multiply the original 15-min consumptions with the random numbers, this step will generates up to 1000 customer consumptions.

Step three:

According to the algorithm, the total usages from June to September for all customers within one group should be summed up. The summation should be divided by the customer number of that group. Also, the coincident peak demand is renewed. Calculate the parameters with renewed values.

Step four:

Repeat step two (the same random number set for training data) for original testing data.

Step five:

Do the peak demand forecast in testing part with the calculated parameters.

Step six:

With customer numbers for each group, obtain the total peak usages for both forecast and actual consumptions in whole SC2 and then calculate the MAPE.

5.1.6.1 Results

With the extra data principle, the results for SC1, SC2 and SC9 are showing in the table below. In the table, for SC9 is the comparison for k-means within SC, for SC1 and SC2 are the comparison for stratification.

Table 5-24 Accuracy Comparisons for Coincident Peak Demand Forecast

	Updated MAPE	**MAPE**
SC1	2.17%	2.82%
SC2	24.42%	65.05%
SC9	2.93%	4.15%

The accuracies are improved with the extra data, which proves the customer numbers do have influence on k-means performance.

5.2 Refine the Parameters with Combined Data from Smart Meter and Traditional Meter

In order to analyze the k-means clustering application in peak demand forecast, k-means clustering also performed on the combined data of both smart meter data and traditional meter data.

5.2.1 Coincident Peak Demand Prediction for Each SC

First step to do this part is to combine the smart meter data and traditional data for each stratification within each SC. According to m&b calculation theory, refine the parameters with combined data.

The flowcharts below present the procedure for this part.

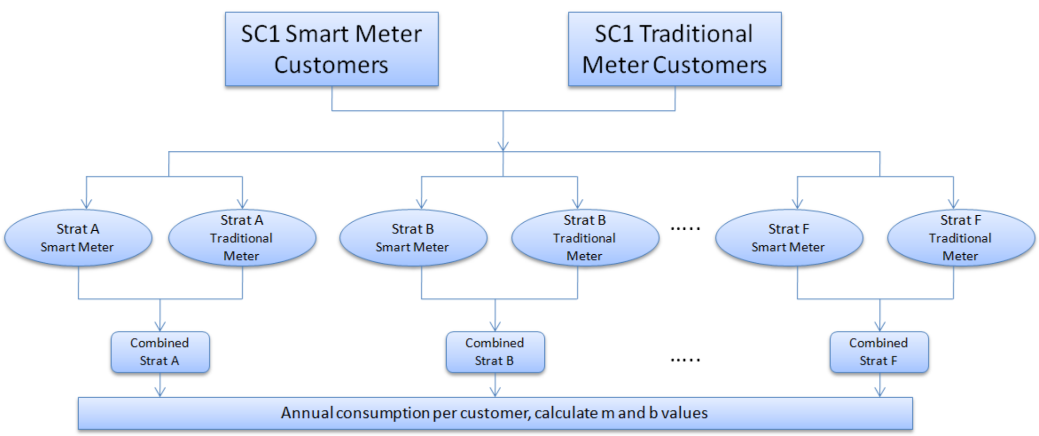


Figure 5-25 Procedures for Combined SC1 Data

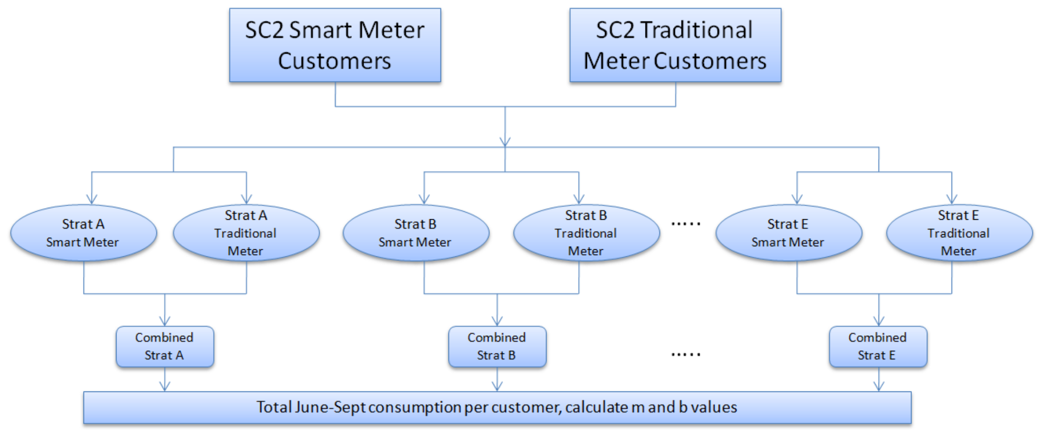


Figure 5-26 Procedures for Combined SC2 Data

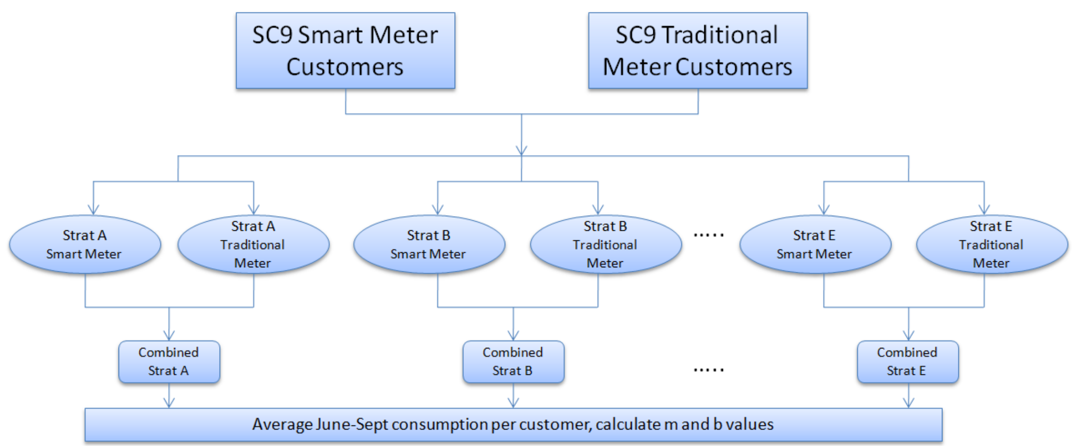


Figure 5-27 Procedures for Combined SC9 Data

5.2.2 Results

The original m&b values in the file "Regression & Spline data for Service Class Estimation" are not applicable in this part, as for this part is the combining of smart meter data and traditional meter data, the one in that file is smart meter only.

The coincident peak demand forecast accuracies for SC1, SC2 and SC9 are presented in the following table:

Table 5-28 The MAPE Comparisons for Combined Data and Smart Meter Data

	MAPE for SM&TM	**MAPE for SM Only**
SC1	0.46%	15.98%
SC2	14.72%	0.50%
SC9	9.71%	4.19%

From above table, for SC1 the precision is improved, but for SC2 and SC9 are not. The reason is that the traditional customer consumptions are much larger than the smart meter customer consumptions in SC2 and SC9.

SC9 is selected to present the smart meter customer and traditional meter customer load profiles for each stratification.

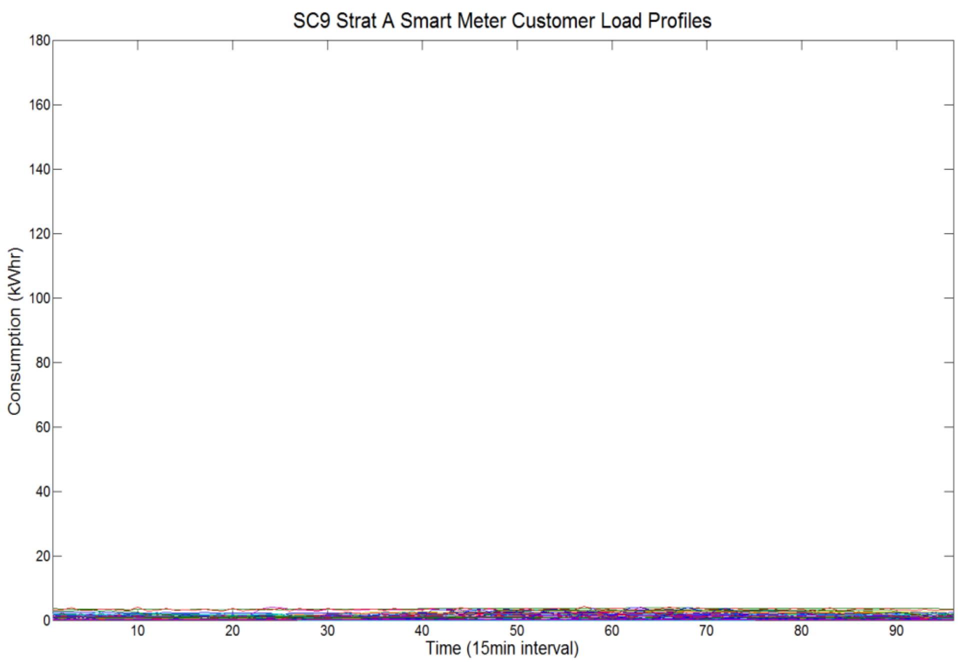


Figure 5-29 SC9 Strat A Smart Meter Customer Load Profiles

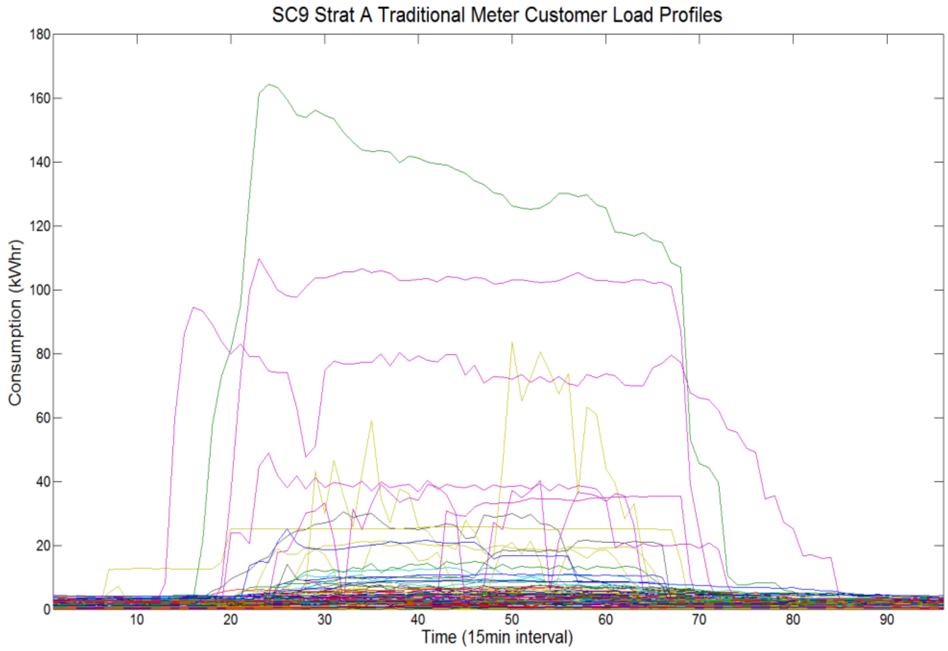


Figure 5-30 SC9 Strat A Traditional Meter Customer Load Profiles

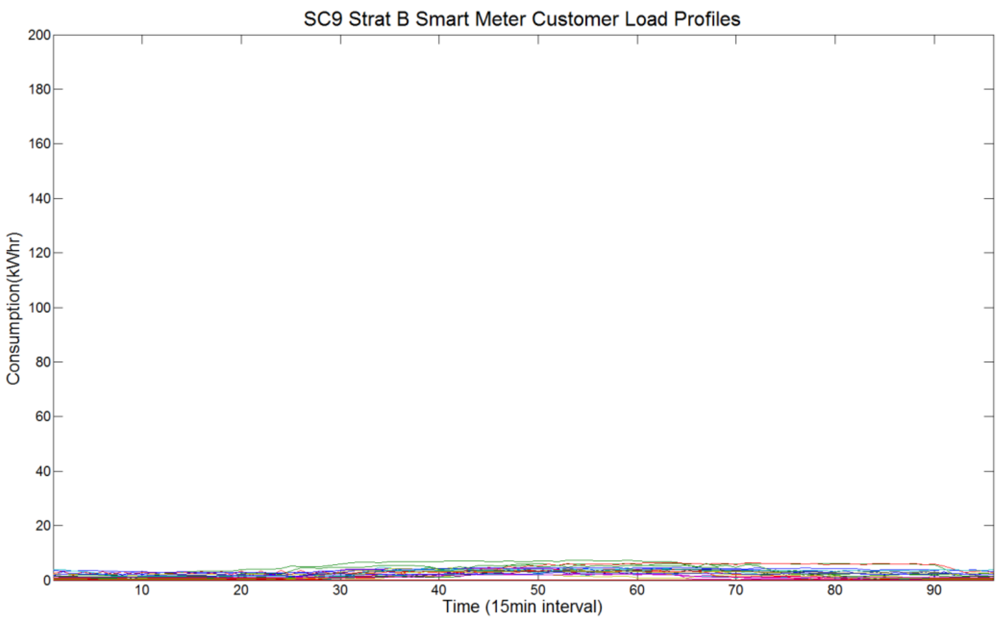


Figure 5-31 SC9 Strat B Smart Meter Customer Load Profiles

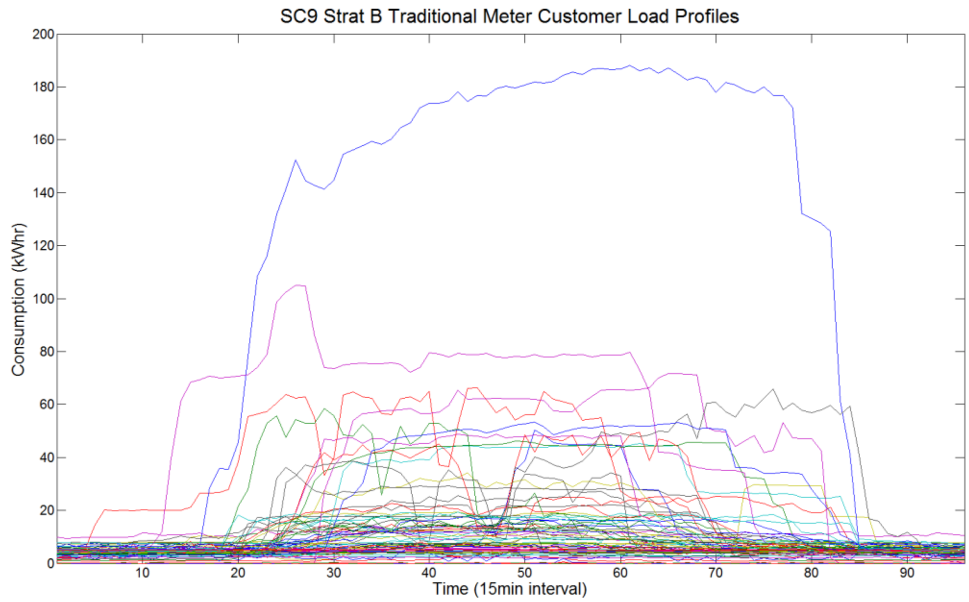


Figure 5-32 SC9 Strat B Traditional Meter Customer Load Profiles

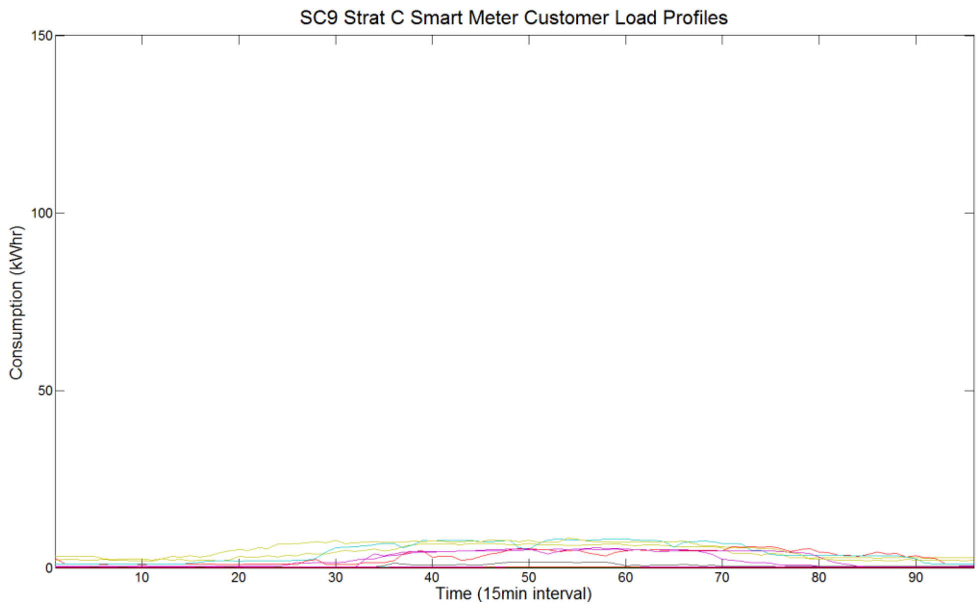


Figure 5-33 SC9 Strat C Smart Meter Customer Load Profiles

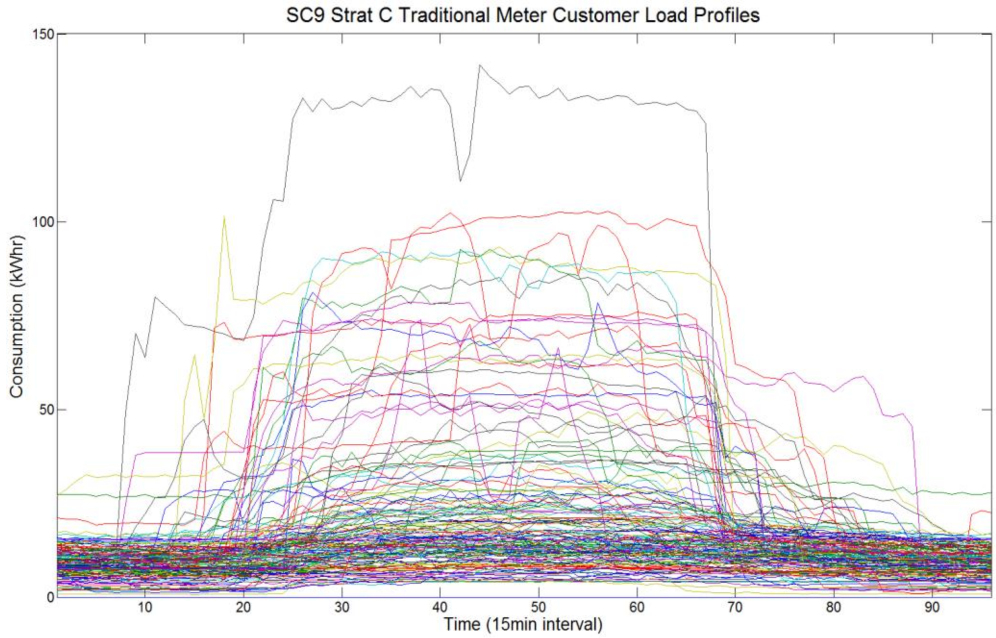


Figure 5-34 SC9 Strat C Traditional Meter Customer Load Profiles

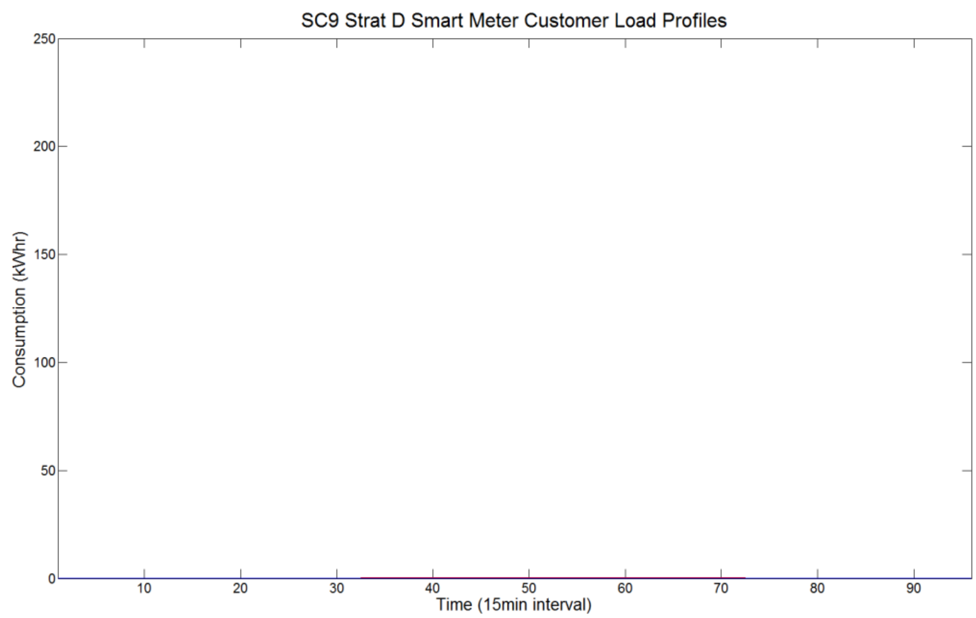


Figure 5-35 SC9 Strat D Smart Meter Customer Load Profiles

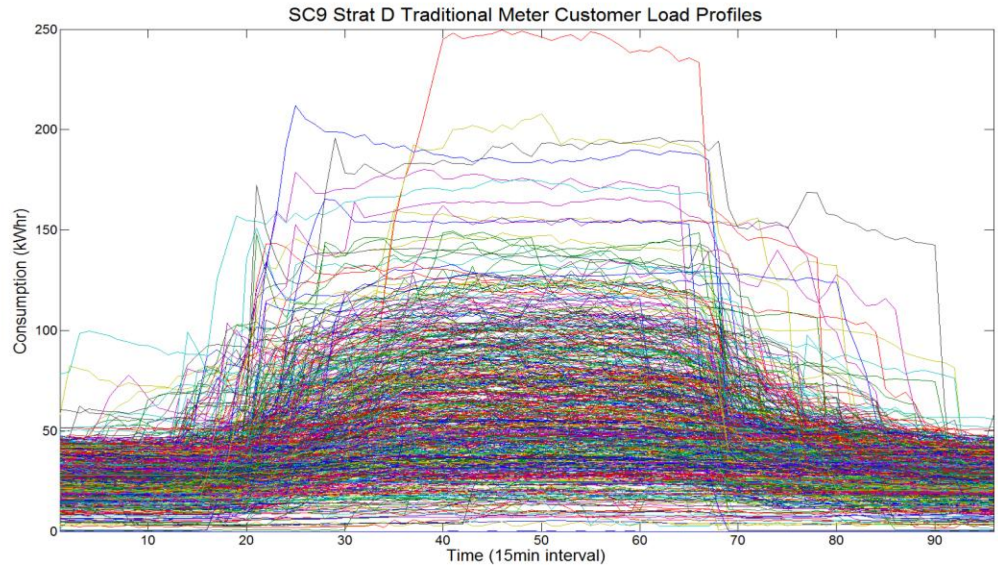


Figure 5-36 SC9 Strat D Traditional Meter Customer Load Profiles

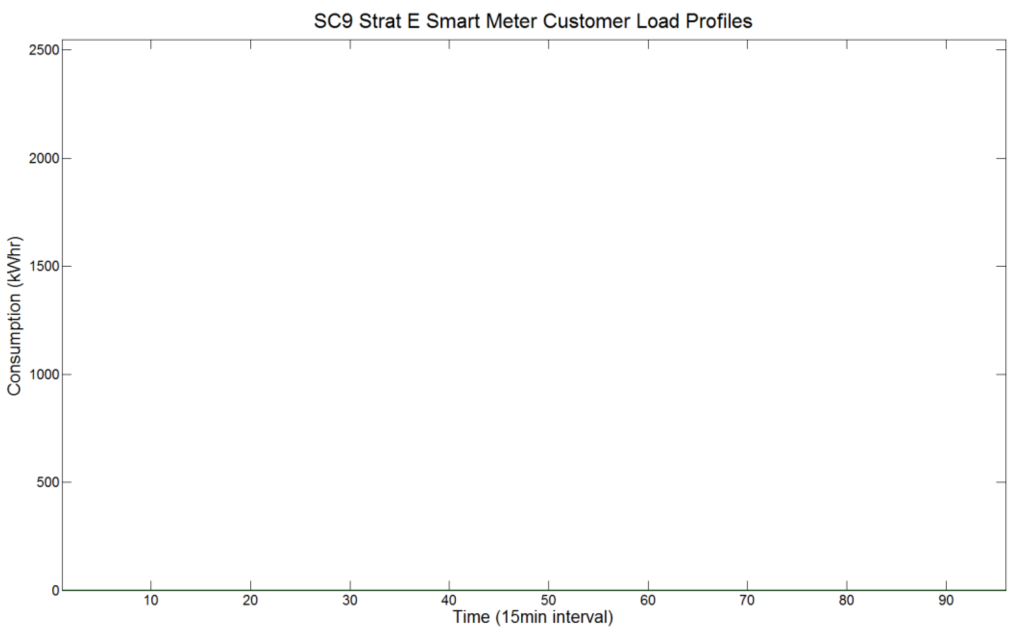


Figure 5-37 SC9 Strat E Smart Meter Customer Load Profiles

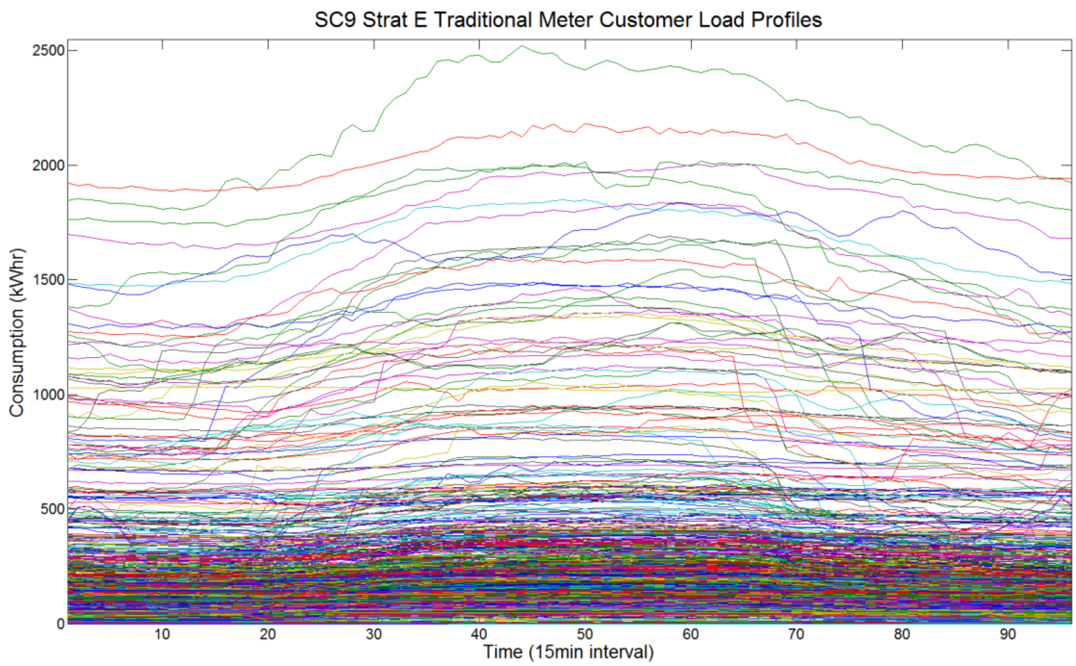


Figure 5-38 SC9 Strat E Traditional Meter Customer Load Profiles

5.3 Apply K-Means within Each Stratification for Combined Data

Apply k-means clustering in each combined stratification for each SC, and rank each group by consumption amount in ascending order, then calculate the m and b values to refine these parameters.

The detailed procedures are shown in the following flowchart:

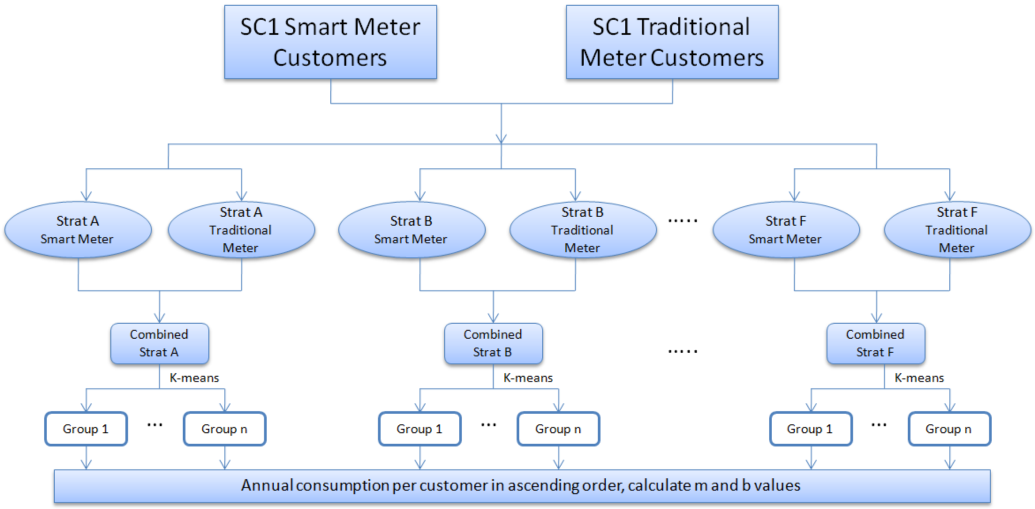


Figure 5-39 K-Means Application within Each Stratification for Combined SC1 Data

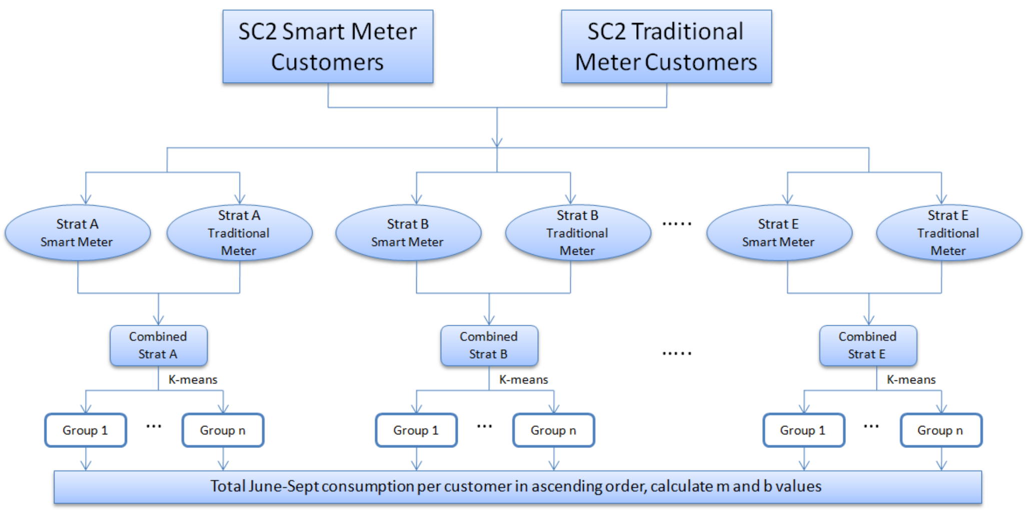


Figure 5-40 K-Means Application within Each Stratification for Combined SC2 Data

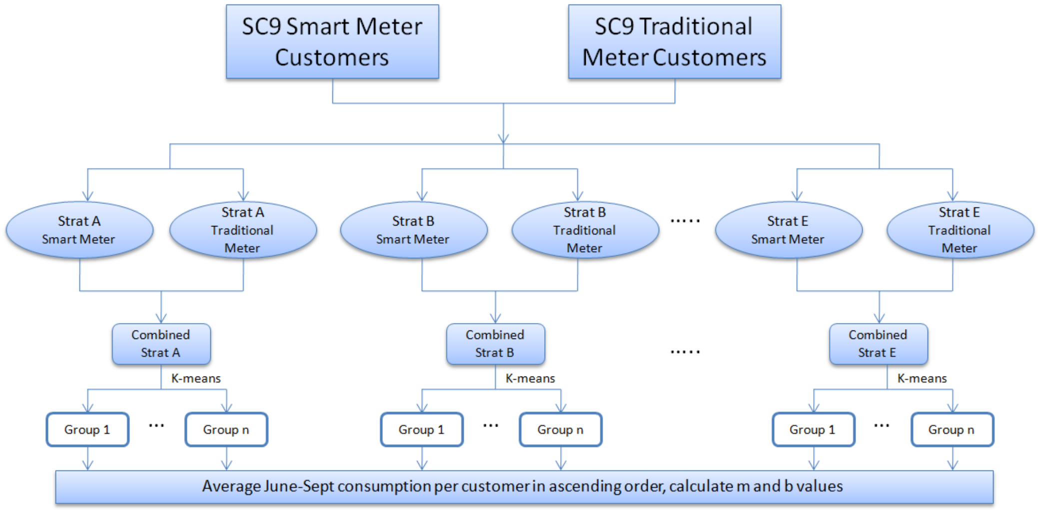


Figure 5-41 K-Means Application within Each Stratification for Combined SC9 Data

5.3.1 Results

The results for this part are not satisfactory due to the small customer number and uncertain boundaries between each group in each stratification after k-means application. The coincident peak demand forecast for each SC shown in the table below:

Table 5-42 MAPEs for Each SC

	MAPE with K-Means	MAPE without K-Means
SC1	162.96%	0.46%
SC2	79.75%	14.72%
SC9	83.66%	9.71%

5.4 Apply K-Means within Each SC for Combined Data

K-means clustering is performed on combined SCs to calculate the parameters. The detailed procedures for this part are displayed in the following flowcharts.

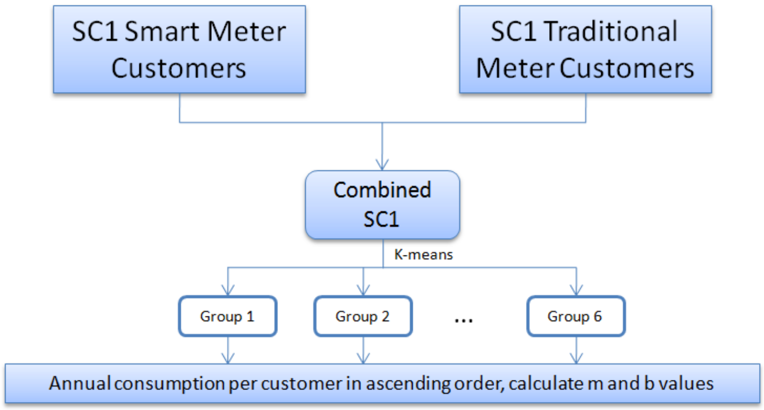


Figure 5-43 K-Means Application within Each SC for Combined SC1 Data

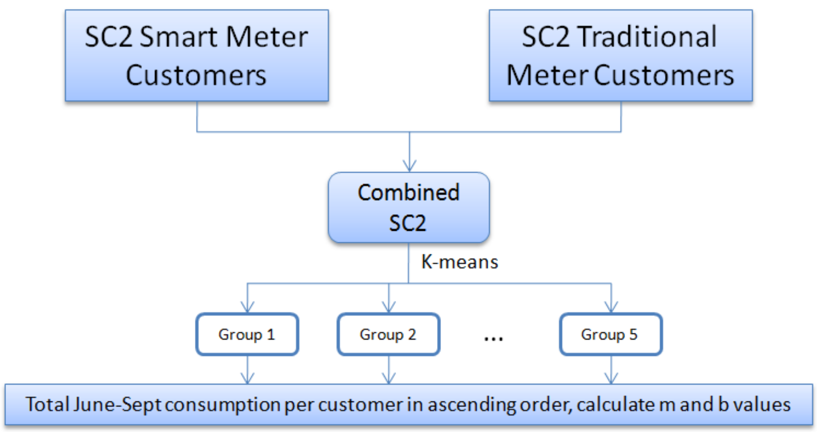


Figure 5-44 K-Means Application within Each SC for Combined SC2 Data

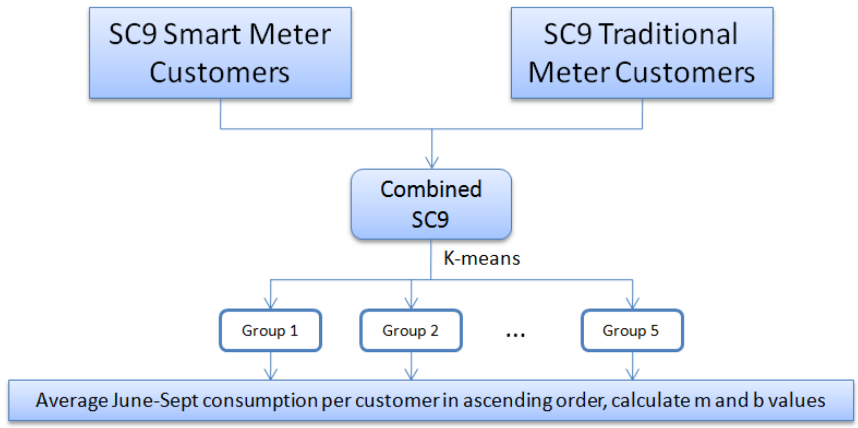


Figure 5-45 K-Means Application within Each SC for Combined SC9 Data

5.4.1 Results

This part presents the forecast results with original data and also the results with extra data.

5.4.1.1 Forecast with Original Data

Table 5-22 shows the MAPE values for each SC. In second column, they are the accuracies for coincident peak demand forecast with parameters calculated from clustering groups. In third column, the parameters are calculated from stratification groups.

Table 5-46 MAPEs for Forecast with Clustering Group and Stratification Group

	MAPE for Clustering Group	**MAPE for Stratification Group**
SC1	0.40%	0.46%
SC2	13.88%	14.72%
SC9	2.62%	2.02%

Above table shows that if the parameters are calculated from clustering groups, the accuracies for SC1 and SC2 are improved, but for SC9 the accuracy reduces. For SC9 customers, according to their consumptions in the test year, some customers are not belong to the same clustering as in the training year. This affects the forecast accuracy.

5.4.1.2 Forecast with Extra Data

Update the results in 5.2.4.1 with extra data to validate the improvement in coincident peak demand forecast accuracy. Still, after the k-means application in each SC, add extra data to each group to make the customer number to 1000. And check the changes in accuracies.

Table 5-47 Comparisons for Forecast with Extra Data and Original Data

	MAPE with Extra Data	**MAPE with Original Data**
SC1	0.62%	0.40%
SC2	2.23%	13.88%
SC9	8.03%	10.66%

From table 5-23, there is an accuracy improvement for SC2 and SC9; as for SC1, the k-means application with more data does not necessary produce better results., This is because that the prediction for each group is either smaller or larger than the actual amount, the errors may be added or subtracted after combining all the data together. Although the accuracy for SC1 doesn't improve, both accuracies are satisfactory.

5.5 With Received 2014 Traditional Data, Test the Refined Parameters for SC2 and SC9 Traditional Data

Similar procedures as part 5.1 smart meter data only, the refined m&b parameter values obtained from 2012 traditional meter customers. Information from the year 2013 is the input for testing part to validate the m&b values.

Both the refined m&b and the original m&b from the "Regression&Spline data for Service Class Estimation.xls" are taken into the testing part for comparison.

5.5.1 Apply K-Means within each Stratification for SC2 and SC9

Apply k-means within each stratification then conduct the algorithm to obtain the parameters. In training part, calculate m&b values with data from the year 2012. In testing part, the information from the year 2013 is treated as input. The coincident peak demand in the year 2014 for traditional meter customers is used to validate the refined m&b values.

5.5.1.1 Results

In order to validate the refined parameters, the coincident peak forecast results from both refined and original parameters are compared. The comparison for SC2 and SC9 are shown in the tables below:

Table 5-48 MAPE Comparison for SC2

SC2	**MAPE**
Refined m&b	24.09%
Original m&b	69.65%

Table 5-49 MAPE Comparison for SC9

SC9	**MAPE**
Refined m&b	14.55%
Original m&b	N/A

From the two tables above, the coincident peak forecast accuracy is improved with refined parameters compared with former ones. As for SC9, there is no parameter in the "Regression&Spline data for Service Class Estimation.xls" file.

5.5.2 Apply K-Means within SC2 and SC9

Apply k-means within each SC then conduct the algorithm to obtain the parameters. In training part, calculate m&b values with data from the year 2012. In testing part, the information from the year 2013 is treated as input. The coincident peak demand in the year 2014 for traditional meter customers is used to validate the refined m&b values.

5.5.2.1 Results

For this part, the parameters from the "Regression&Spline data for Service Class Estimation.xls" file are not applicable, as the clustering criteria is different, for this part is k-means, the one in the Excel file is customer usages. So the table below shows the coincident peak demand forecast for refined parameters only.

Table 5-50 MAPE for SC2

SC2	MAPE
Refined m&b	22.91%

Table 5-51 MAPE for SC9

SC9	MAPE
Refined m&b	34.71%

For peak demand estimation parameter refinery, the original philosophy for stratification groups with up to date meter information provides better results.

5.6 Temperature Variable Invited to Peak Demand Forecast

Temperature has mainly influence on peak demand forecast. In order to adjust the impact of the temperature on the peak demand prediction, the temperature variables are taken into consideration in this part. The table below shows the temperature and the corresponding variable for the peak day in the year 2013 and 2012:

Table 5-52 Temperature Variables

Year	T for Peak Day	T Variable
2013	86.3	1
2012	85	0.985

With temperature variables the forecast result should be updated as below:

$$\text{Updated Forecast} = \frac{\text{Forecast}*1}{0.985} = \text{Forecast}/0.985$$

Then the updated equation for the accuracy is:

$$\text{MAPE} = \frac{|\text{Forecast}/0.985 - \text{Acutal}|}{\text{Actual}} * 100\%$$

Updated the accuracy for SC1, SC2 and SC9 smart meter data from stratification groups with temperature variable to validate its efficiency. Following table shows the MAPE with and without temperature variable (TV).

Table 5-53 MAPE for SC1, SC2 and SC9 with and without Temperature Variable

	With TV	**Without TV**
SC1	16.17%	17.43%
SC2	9.96%	11.31%
SC9	2.69%	4.15%

With the results from above table, we can see the accuracies got improved with temperature variable.

5.7 Drive the Parameters and Average Curves for Other Seasons

Part 5.1 through Part 5.6 are dealing with the parameters in summer time; in this section, the parameters and the average curves for spring, fall, and winter are also calculated.

5.7.1 Drive the Parameters for Each SC

For SC1, there are three groups of customers: Min15Ch2, Min1Ch2 and Min1Ch20, and there are enough customers in these groups. For SC2 and SC9, only Min1Ch20 group has enough customers.

5.7.1.1 Parameters for SC1

The parameters for each group in each season are shown in the tables below:

Table 5-54 Parameters for Min15Ch2 in Each Season

Min15Ch2					
Spring		**Fall**		**Winter**	
m	**b**	**m**	**b**	**m**	**b**
0.000407519	-0.2325	0.000163654	0.0582	0.000202188	-0.0031
0.000207935	0.2726	0.000161086	0.0647	0.000184337	0.0421
0.000365640	-0.2776	0.000190539	-0.0381	0.000103935	0.3226
0.000244434	0.2757	0.000155041	0.1240	0.000235322	-0.2772
0.000256664	0.2014	0.000204527	-0.1766	0.000218460	-0.1747

Table 5-55 Parameters for Min1Ch2 in Each Season

Min1Ch2					
Spring		**Fall**		**Winter**	
m	**b**	**m**	**b**	**m**	**b**
0.000378321	0.1000	0.000138854	0.3727	0.000294817	0.2366
0.000145778	0.6374	0.000136230	0.3788	0.000048387	0.8062
0.000581486	-0.8411	0.000518092	-0.9170	0.000225968	0.2036
0.000602021	-0.9218	0.000245268	0.1549	0.000333427	-0.2186
0.000676136	-1.2926	0.000424781	-0.7431	0.000140092	0.7486

Table 5-56 Parameters for Min1Ch20 in Each Season

Min1Ch20					
Spring		**Fall**		**Winter**	
m	**b**	**m**	**b**	**m**	**b**
0.000956318	-1.6712	0.000095653	0.3217	0.000346332	1.4814
0.000116196	0.4916	0.000288794	-0.1755	0.000228485	0.0016
0.000867276	-2.3510	0.000121214	0.4587	0.000320604	-0.3470
0.000132515	0.9436	0.000173383	0.2248	0.000209653	0.1505

0.000607671	-2.2362	0.000443756	-1.5847	0.000008788	1.4947

Table 5-57 Parameters for Combined 1min in Each Season

Combined 1 min					
Spring		**Fall**		**Winter**	
m	**b**	**m**	**b**	**m**	**b**
0.000893097	-1.3247	0.000187945	0.1610	0.000067406	0.5682
0.000145943	0.5125	0.000226829	0.0654	0.000165846	0.3261
0.000747523	-1.6424	0.000308458	-0.2270	0.000282419	-0.0915
0.000309092	0.1932	0.000202683	0.2159	0.000257491	0.0129
0.000714037	-2.1393	0.000460728	-1.2705	0.000105144	0.8904

In addition to the accuracies in Table 5-26, Table 5-31 also shows the accuracies for SC1 in Spring, Fall, and Winter with and without temperature variable.

Table 5-58 MAPE Comparisons for SC1 Min15Ch2 Stratification Customers Peak Demand Estimation in 2013 With and Without Temperature Variables

	Spring	**Summer**	**Fall**	**Winter**
With TV	11.1%	16.17%	12.1%	9.7%
Without TV	12.4%	17.43%	13.4%	11.2%

5.7.1.2 Parameters for SC2

Table 5-59 Parameters for Min1Ch20 in Each Season

Min1Ch20					
Spring		**Fall**		**Winter**	
m	**b**	**m**	**b**	**m**	**b**
0.002336128	-0.2128	0.001882588	-0.1716	0.000870949	-0.2008
0.001333312	0.5863	0.001370473	0.2649	0.000600969	0.3595
0.001514812	0.3423	0.000243772	1.9575	0.000346958	1.3390
0.000772758	1.9719	0.000944442	0.4266	0.000185001	2.1385
0.000772758	1.9719	0.000944442	0.4266	0.000185001	2.1385

Besides the accuracies in Table 5-26, Table 5-33 also shows the accuracies for SC2 in Spring, Fall, and Winter with and without temperature variable.

Table 5-60 MAPE Comparisons for SC2 Min1Ch20 Stratification Customers Peak Demand Estimation in 2013 With and Without Temperature Variables

	Spring	**Summer**	**Fall**	**Winter**
With T	10.19%	9.96%	9.75%	11.53%
Without T	11.53%	11.31%	11.10%	12.86%

5.7.1.3 Parameters for SC9

Table 5-61 Parameters for Min1Ch20 in Each Season

Min1Ch20					
Spring		**Fall**		**Winter**	
m	**b**	**m**	**b**	**m**	**b**
1.227623264	-0.4757	1.234476907	-0.8037	0.959952558	0.5136
1.200202127	-0.2531	1.198903478	-0.5134	0.990127401	0.2811
1.156637506	-0.0419	1.112316072	-0.0072	1.055138889	-0.0156
0.996542783	0.0459	1.240110590	-0.0674	1.003985383	0.0060
0.996542783	0.0459	1.240110590	-0.0674	1.003985383	0.0060

Besides the accuracies in Table 5-26, Table 5-35 also shows the accuracies for SC9 in Spring, Fall, and Winter with and without temperature variable.

Table 5-62 MAPE Comparisons for SC9 Min1Ch20 Stratification Customers Peak Demand Estimation in 2013 With and Without Temperature Variables

	Spring	**Summer**	**Fall**	**Winter**
With T	3.58%	2.69%	4.25%	3.36%
Without T	5.02%	4.15%	5.69%	4.81%

5.7.2 Drive the Average Curves for Each SC

5.7.2.1 The Average Curves for SC1

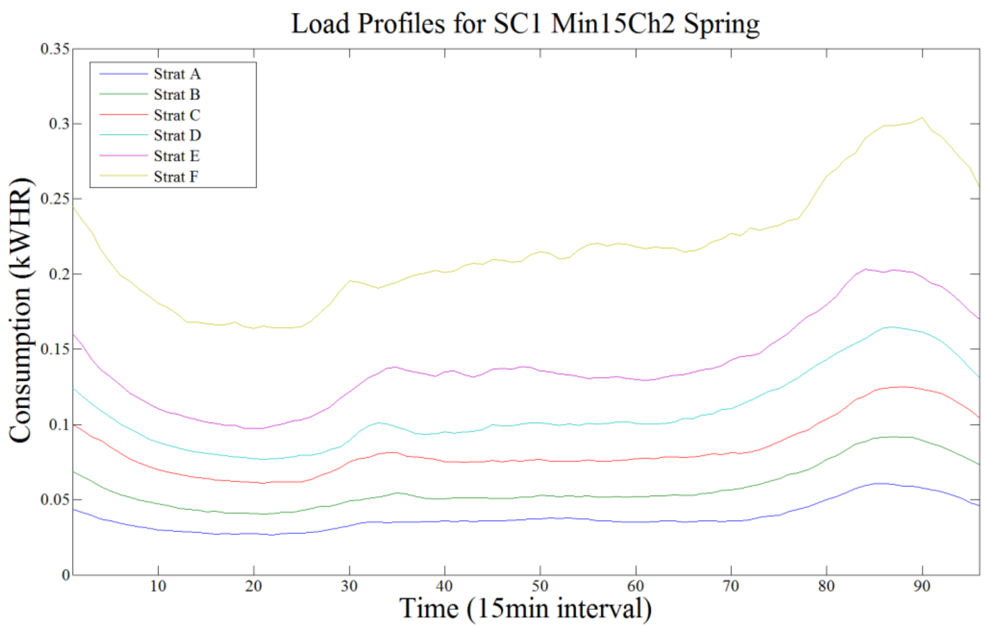


Figure 5-63 Load Profiles for SC1 Min15Ch2 Spring

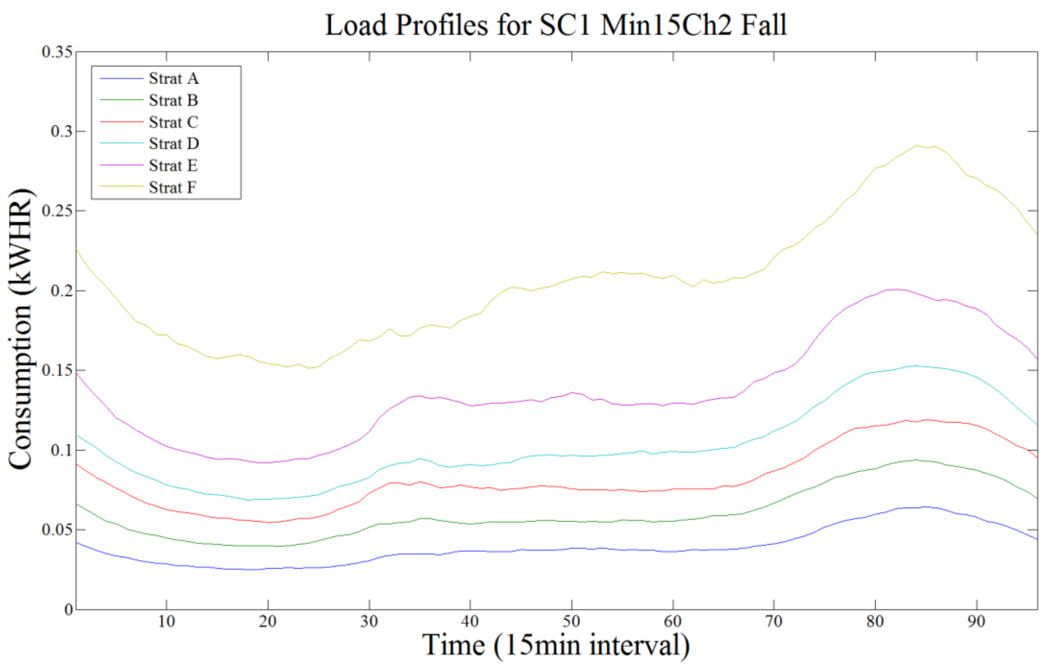


Figure 5-64 Load Profiles for SC1 Min15Ch2 Fall

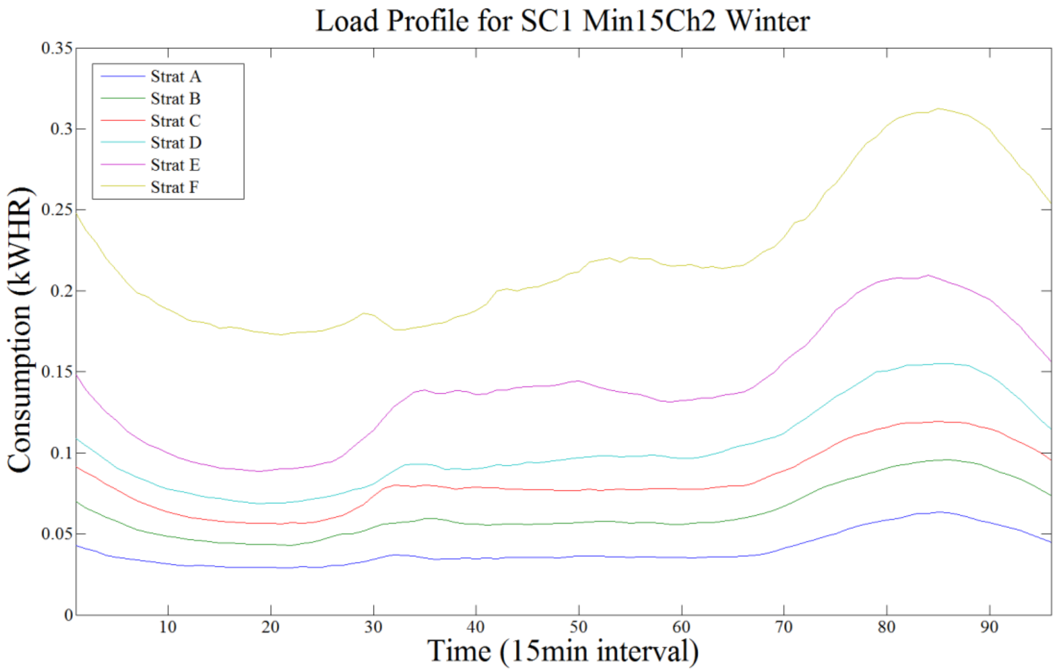


Figure 5-65 Load Profiles for SC1 Min15Ch2 Winter

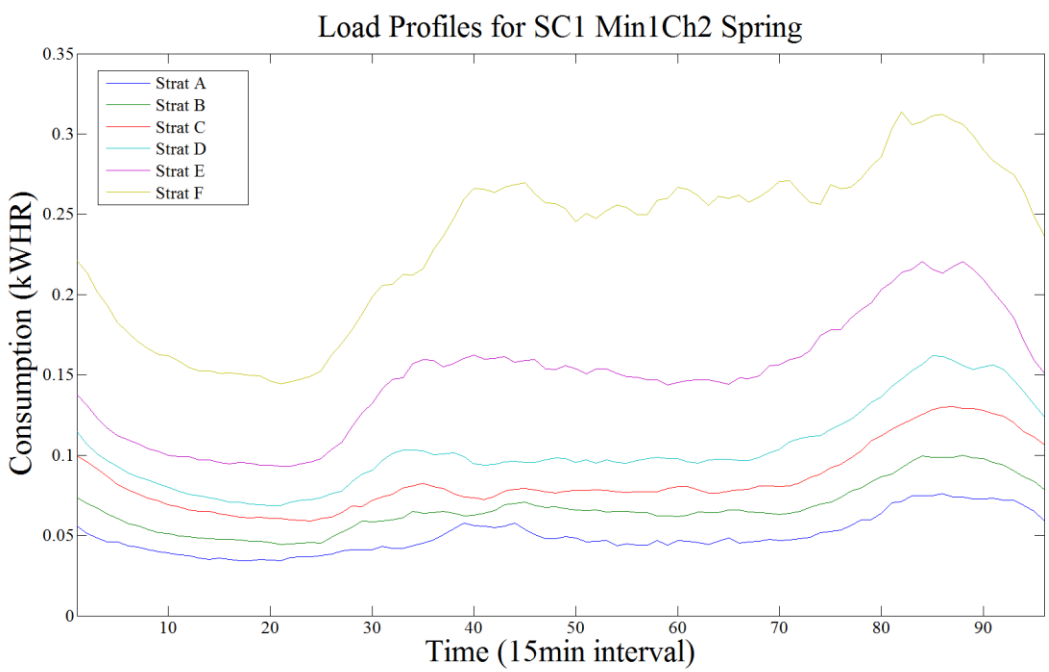


Figure 5-66 Load Profiles for SC1 Min1Ch2 Spring

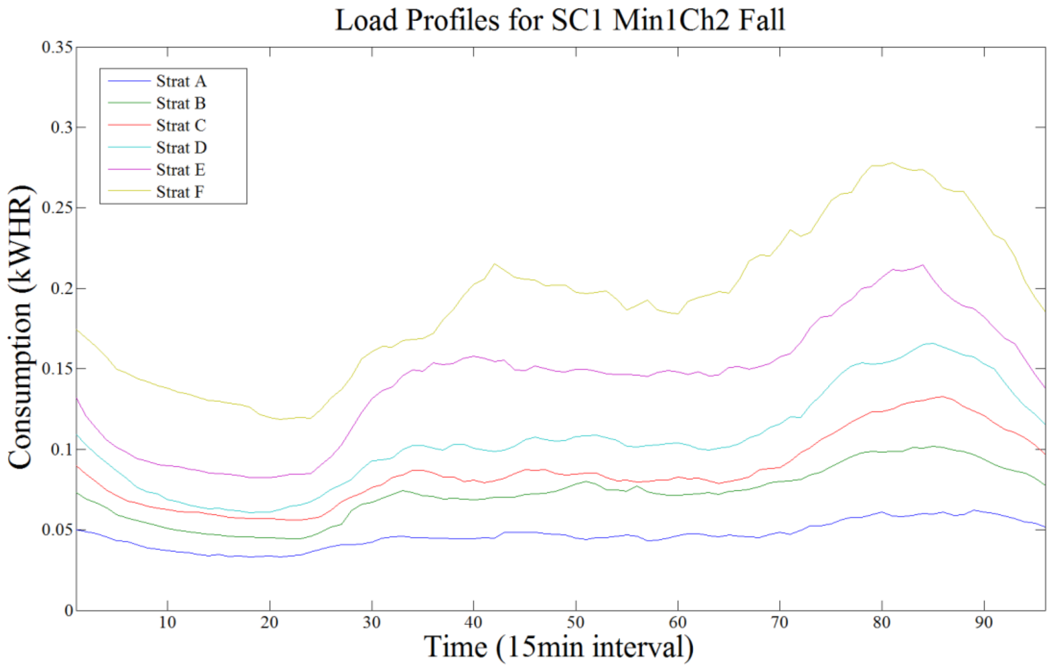


Figure 5-67 Load Profiles for SC1 Min1Ch2 Fall

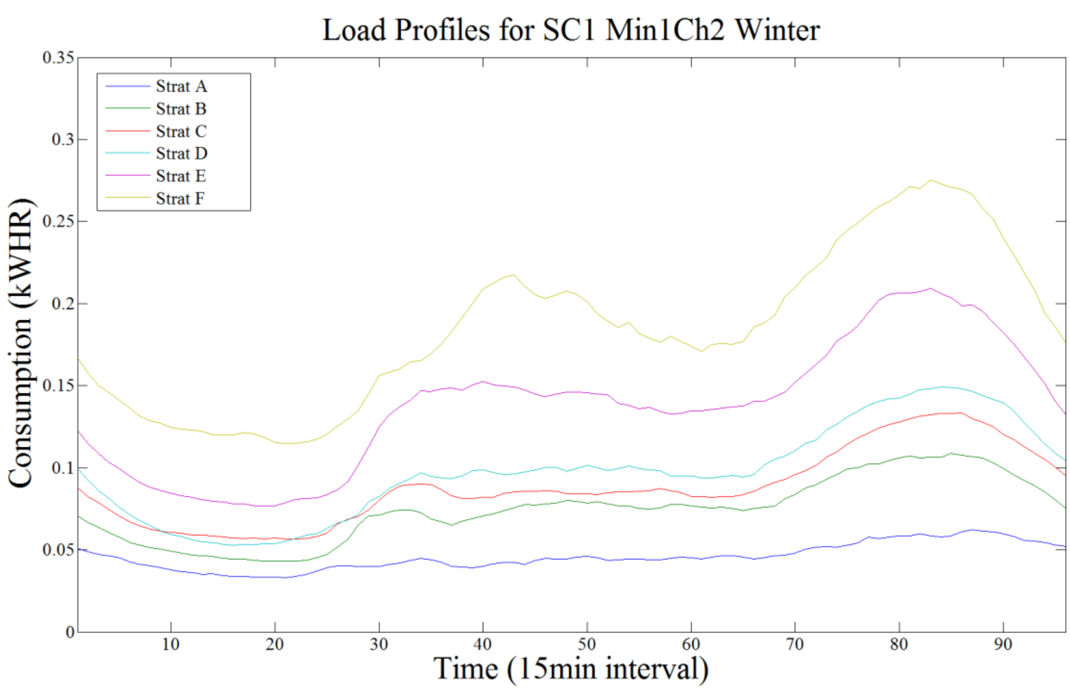


Figure 5-68 Load Profiles for SC1 Min1Ch2 Winter

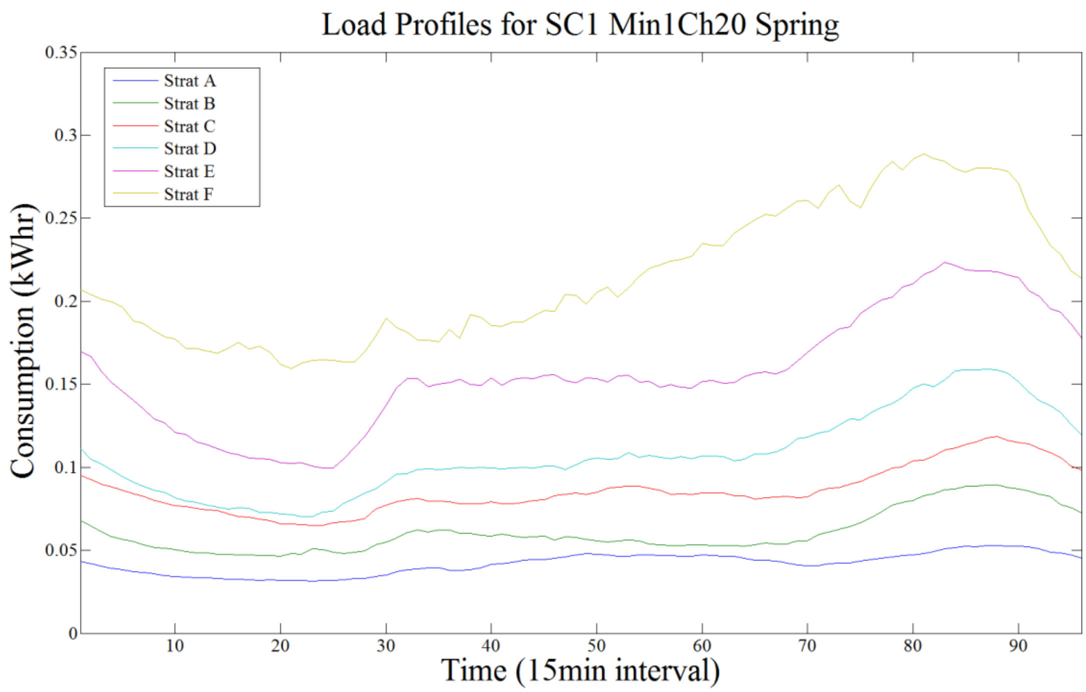


Figure 5-69 Load Profiles for SC1 Min1Ch20 Spring

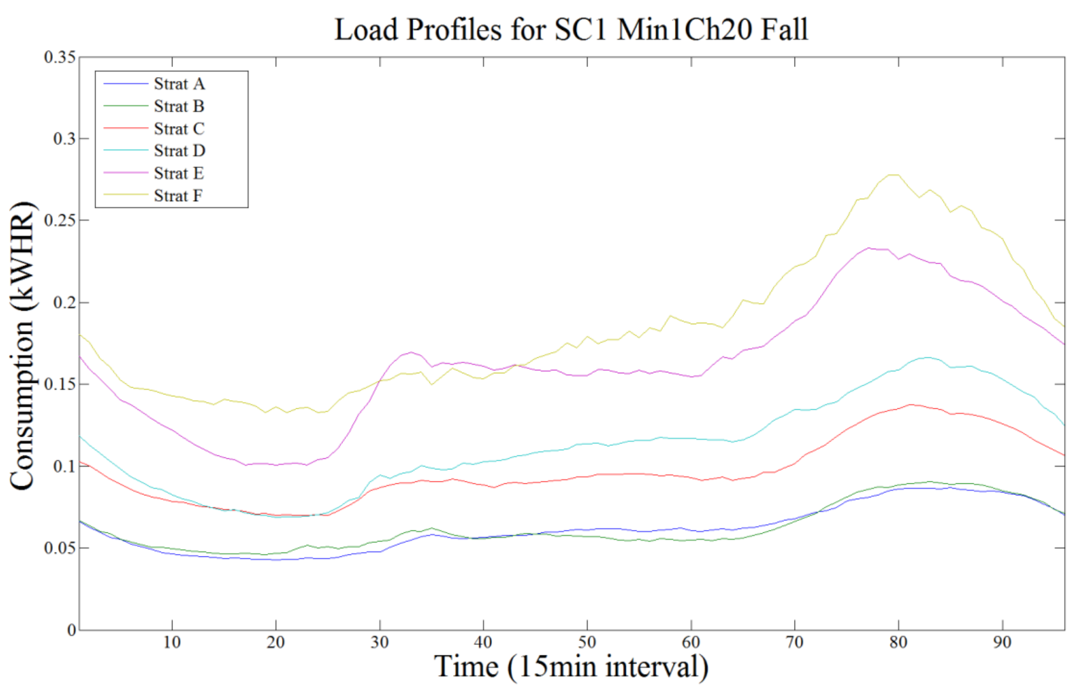


Figure 5-70 Load Profiles for SC1 Min1Ch20 Fall

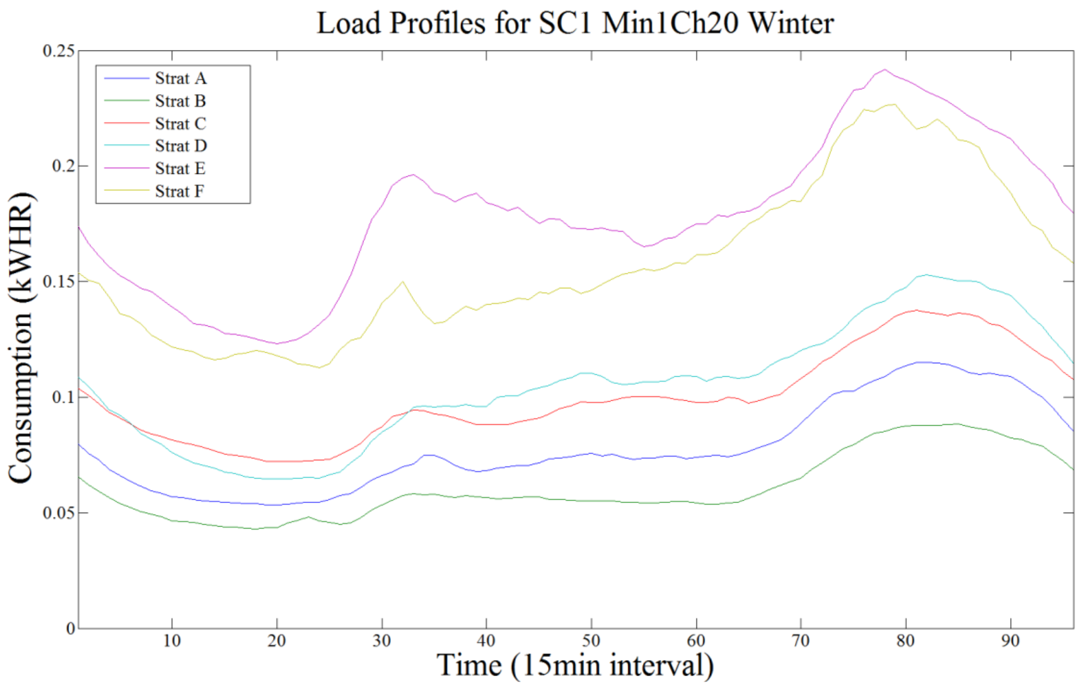


Figure 5-71 Load Profiles for SC1 Min1Ch20 Winter

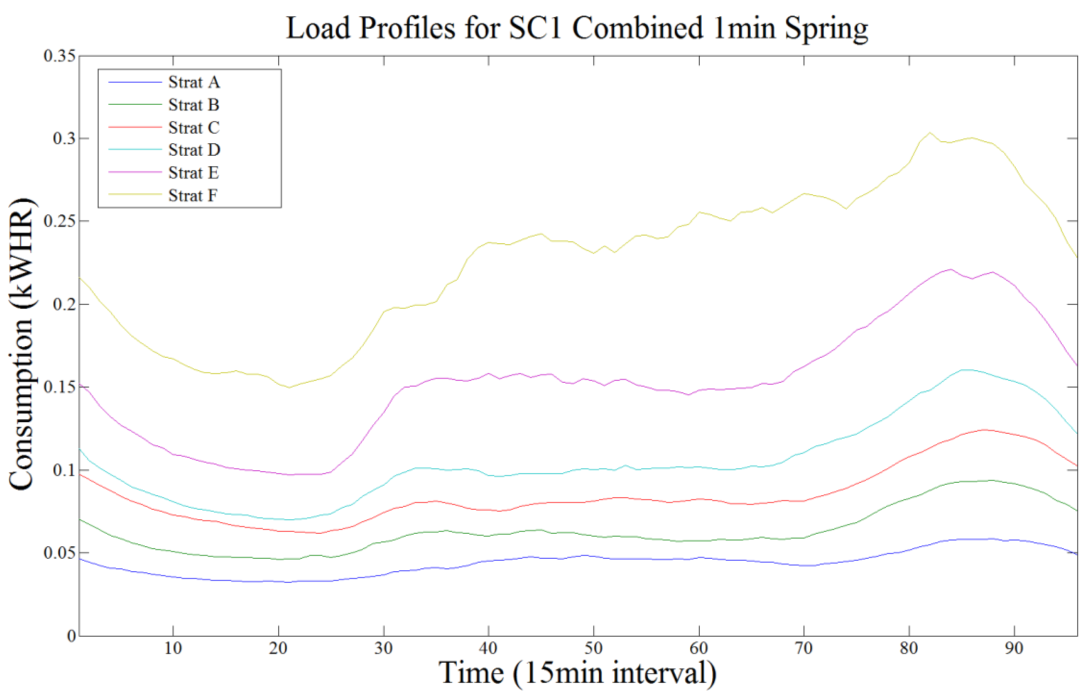


Figure 5-72 Load Profiles for SC1 Combined 1Min Spring

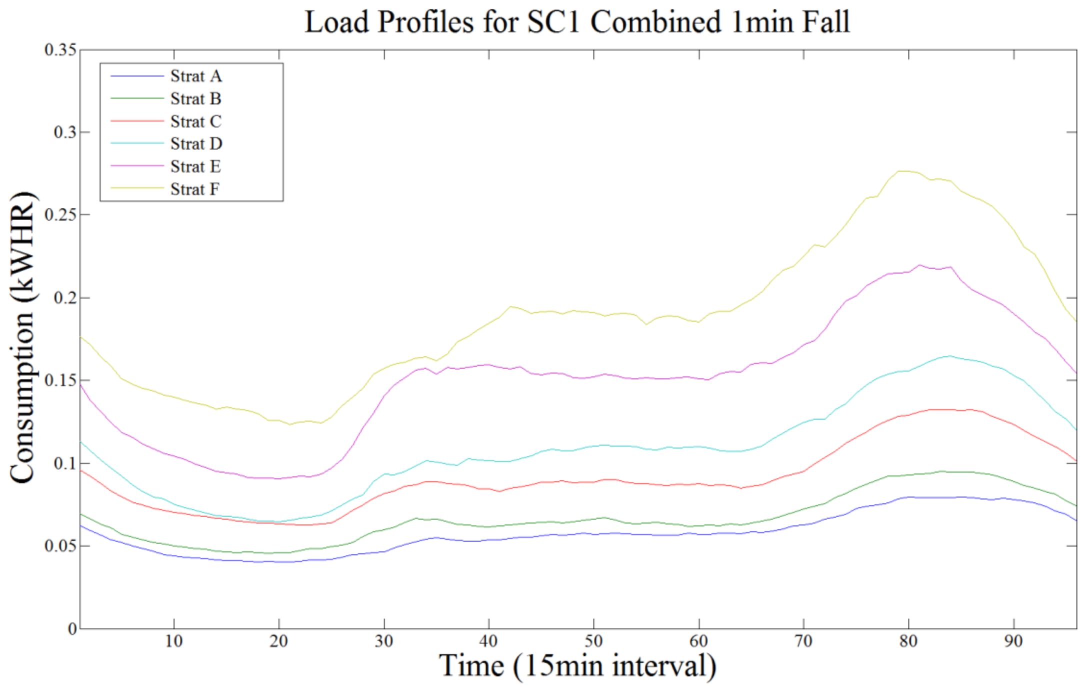


Figure 5-73 Load Profiles for SC1 Combined 1Min Fall

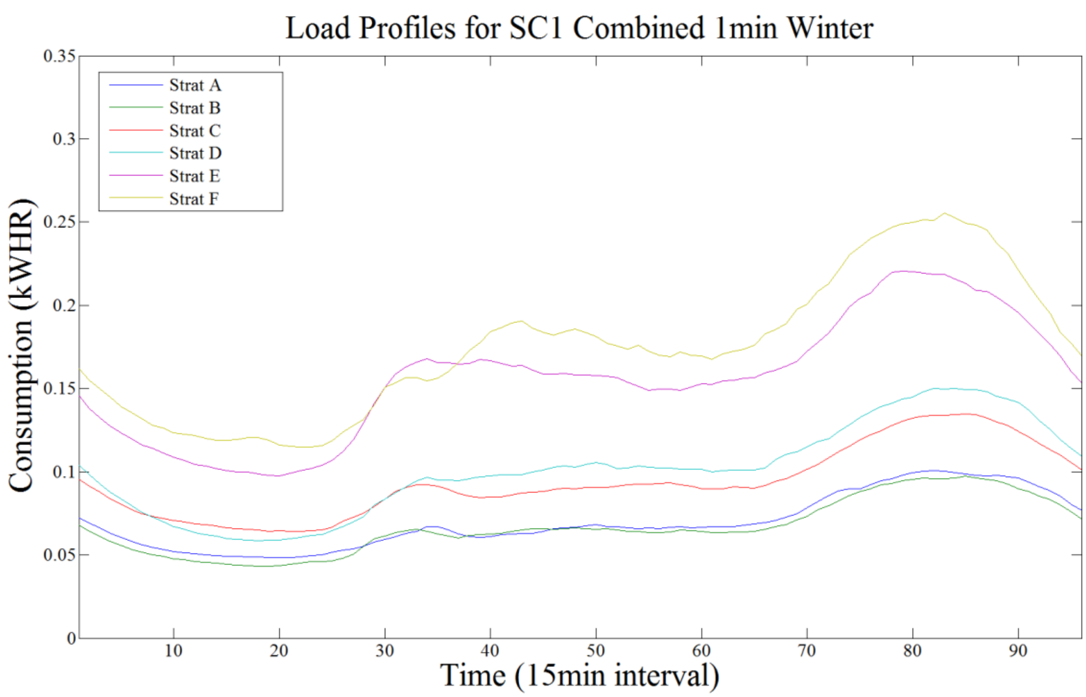


Figure 5-74 Load Profiles for SC1 Combined 1Min Winter

5.7.2.2 The Average Curves for SC2

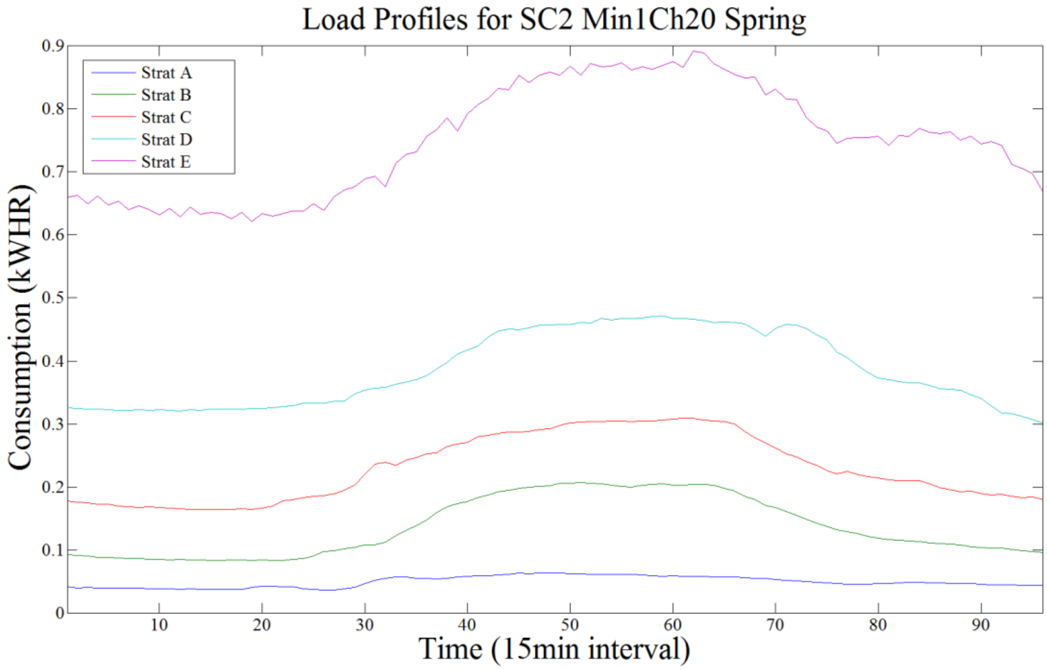


Figure 5-75 Load Profiles for SC2 Min1Ch20 Spring

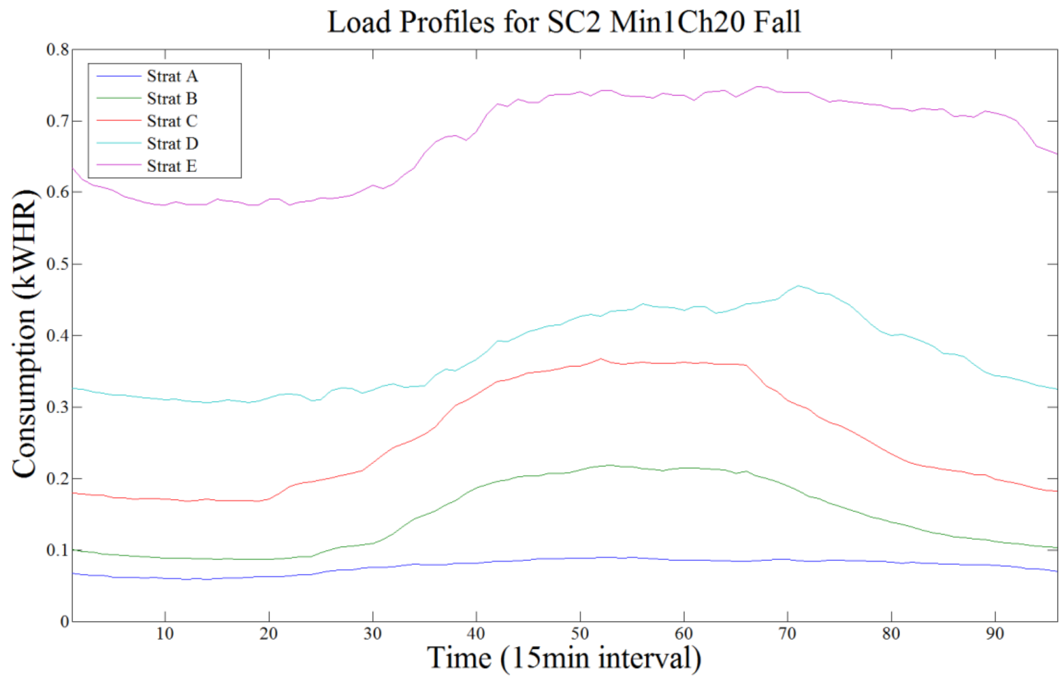


Figure 5-76 Load Profiles for SC2 Min1Ch20 Fall

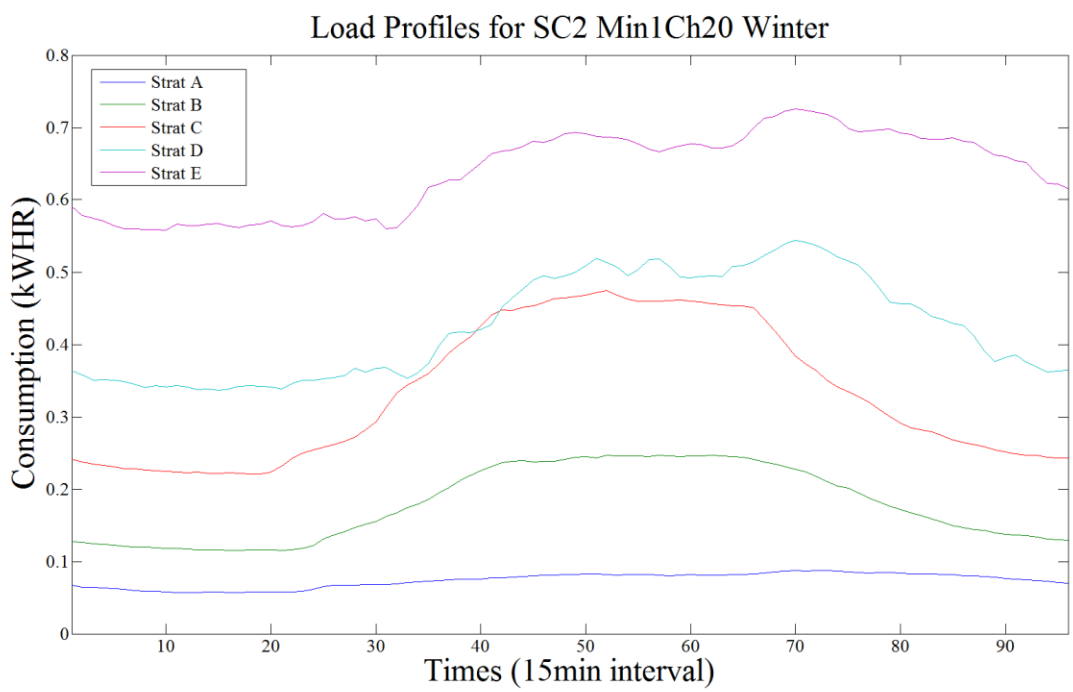


Figure 5-77 Load Profiles for SC2 Min1Ch20 Winter

5.7.2.3 The Average Curves for SC9

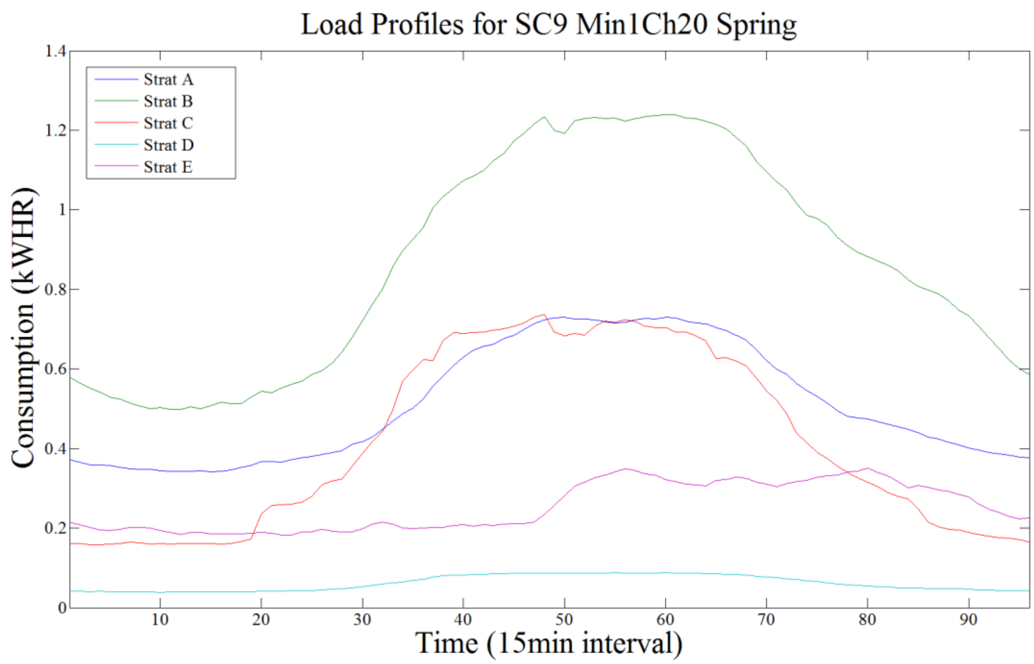


Figure 5-78 Load Profiles for SC9 Min1Ch20 Spring

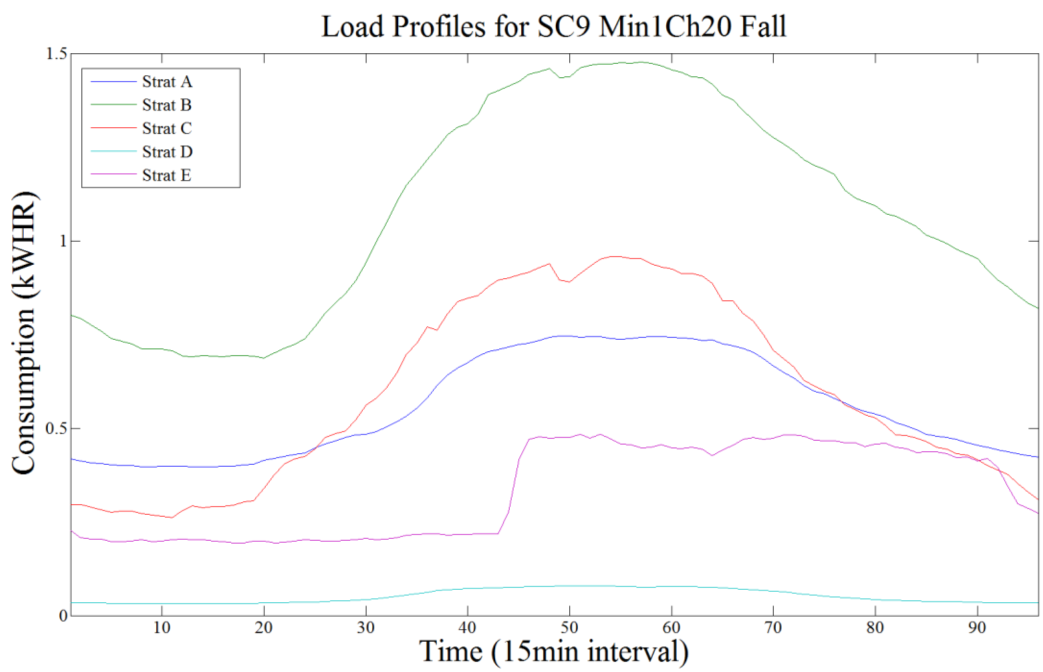


Figure 5-79 Load Profiles for SC9 Min1Ch20 Fall

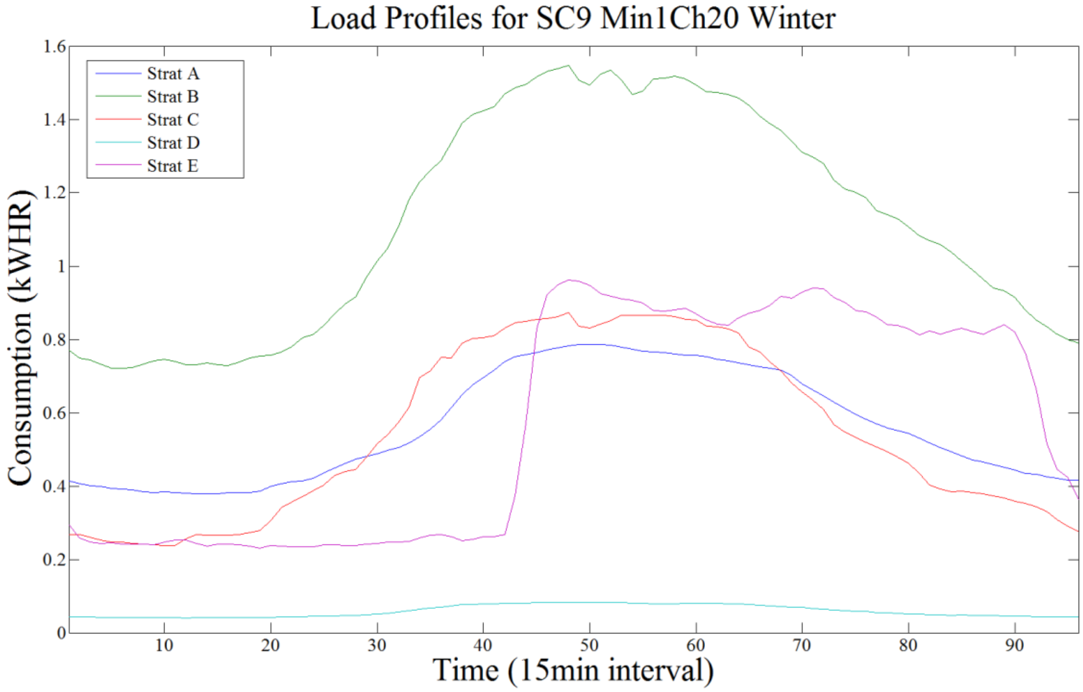


Figure 5-80 Load Profiles for SC9 Min1Ch20 Winter

Chapter 6

CONCLUSIONS AND FUTURE RESEARCH

6.1 Conclusions

For the whole dissertation, data preparation and data preprocessing before the application of analytical functions were two crucial procedures that transformed raw data into actionable information.

This dissertation is focused on residential, general small business, and general large business sectors. From these three sectors, the algorithm can be effectively developed as an application to other sectors.

The capacity of alien customers (the customers whose behaviors show marked consumption contrasts with other customers) influences the k-means clustering performance on load profile forecast algorithms. If there are any alien customers, they should be handled separately from the rest of the customers.

In order to refine the parameters of peak demand forecast, the original philosophy for stratification group with up to date data should be adopted. Applying temperature variable in coincident peak demand forecast can improve the accuracy. K-means clustering is not recommended to be applied in refining the parameters for peak demand forecast, because k-means clustering is based upon customer behavior patterns not the consumption amount.

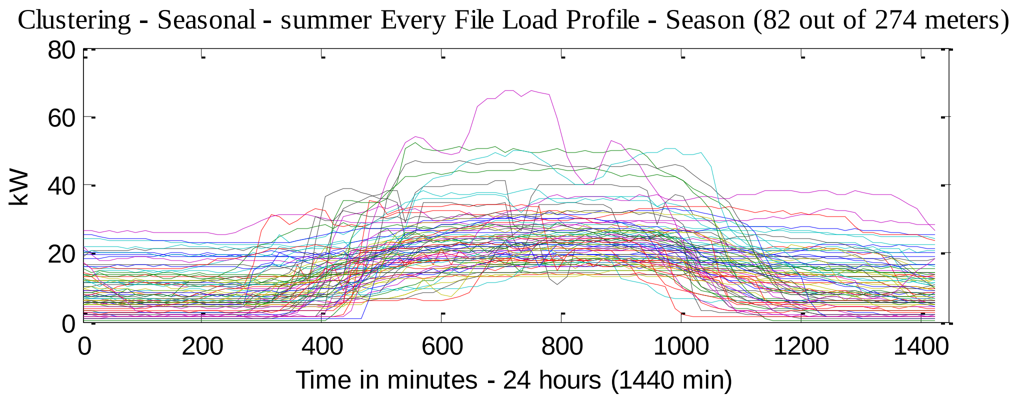
6.2 Potential Future Work

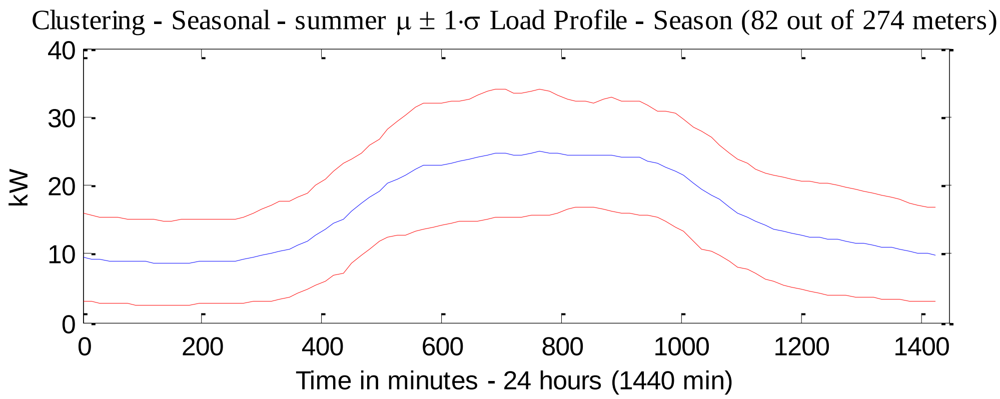
Electricity consumption is increasing dramatically, the day ahead estimation algorithm for load profile proved to increase the forecast accuracy, if the algorithm can be extended to weekly ahead, monthly ahead or used for medium-term load forecast, it may benefit utilities, customers, and save energy.

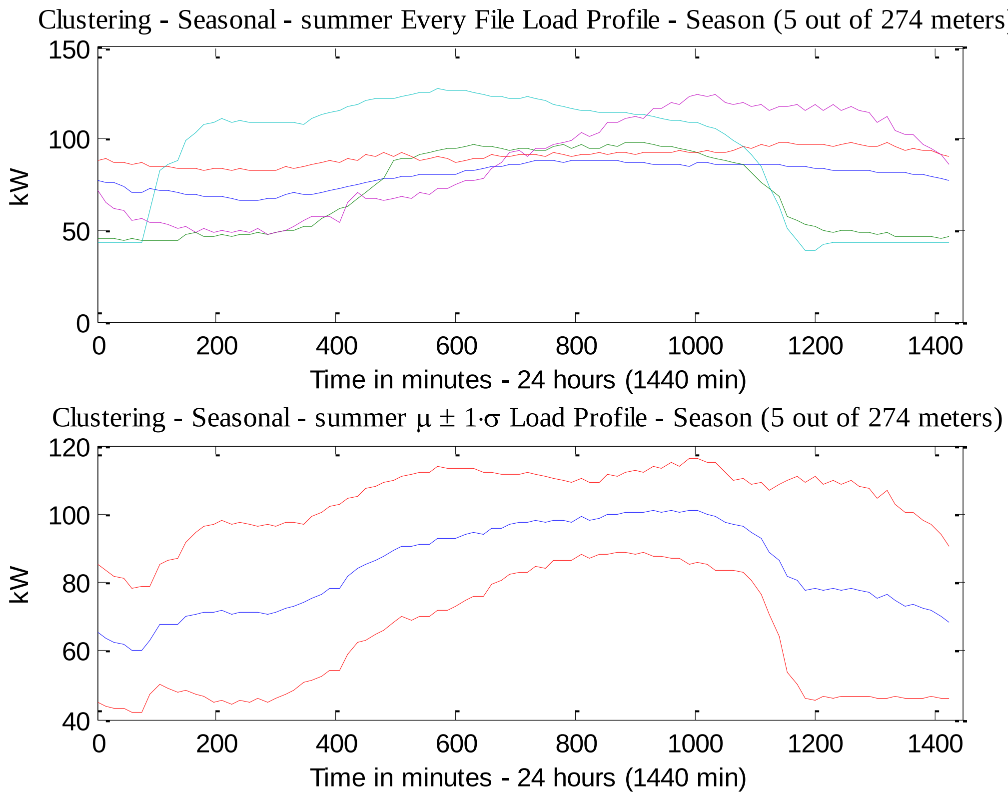
Appendix A

Load Profile Development Based on Clustering Techniques Case Example for SC9

Summer Season: July, August, and September 2012

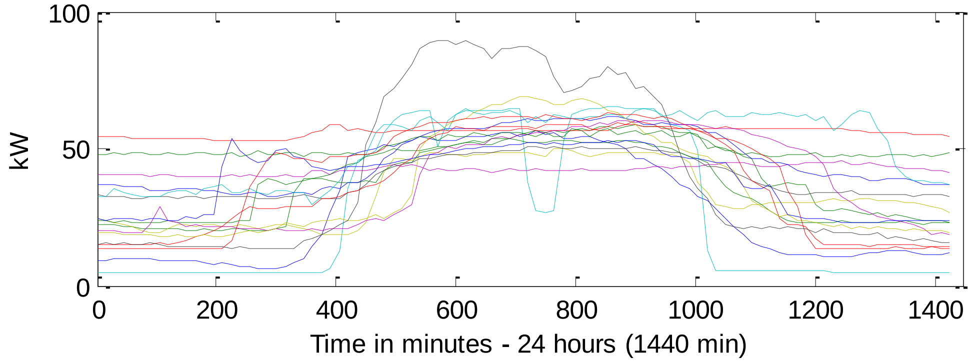

Clustering - Seasonal - summer Every File Load Profile - Season (82 out of 274 meters)
kW
0
20
40
60
80
0
200
400
600
800
1000
1200
1400
Time in minutes - 24 hours (1440 min)


Clustering - Seasonal - summer μ ± 1·σ Load Profile - Season (82 out of 274 meters)
kW
0
10
20
30
40
0
200
400
600
800
1000
1200
1400
Time in minutes - 24 hours (1440 min)

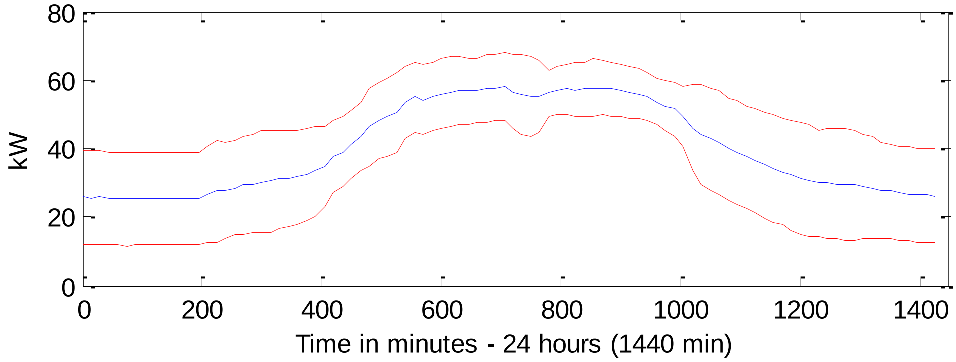


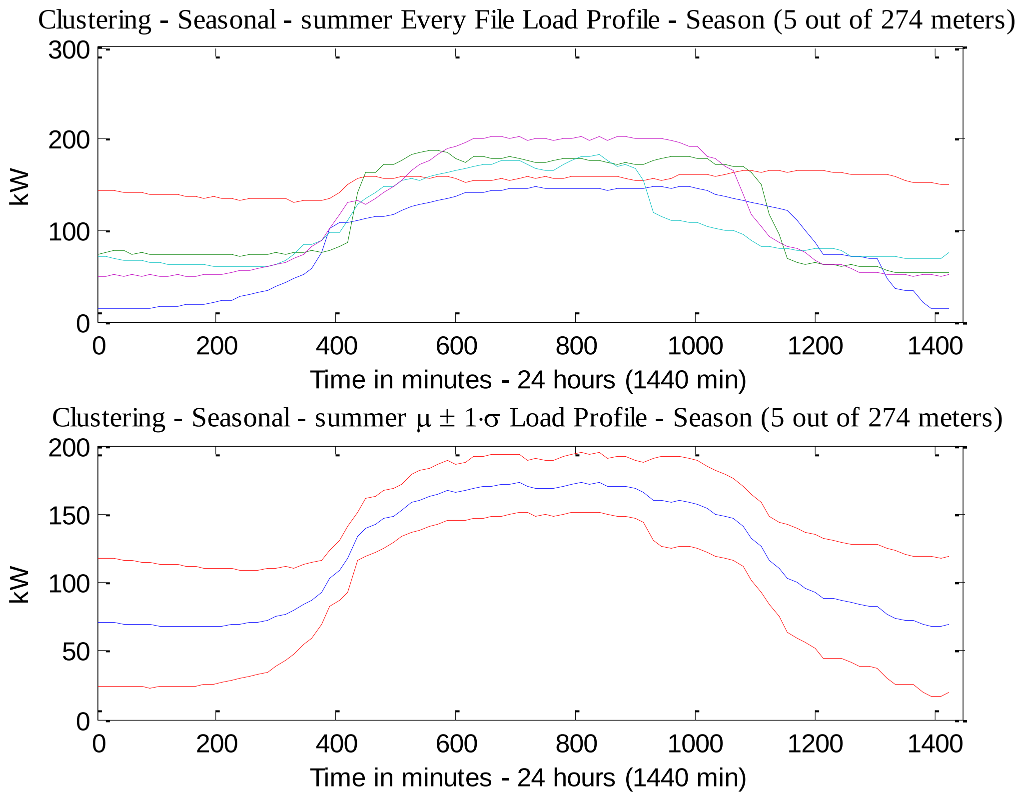
Clustering - Seasonal - summer Every File Load Profile - Season (5 out of 274 meters)
150
100
50
0
kW
0
200
400
600
800
1000
1200
1400
Time in minutes - 24 hours (1440 min)
Clustering - Seasonal - summer μ ± 1·σ Load Profile - Season (5 out of 274 meters)
120
100
80
60
40
kW
0
200
400
600
800
1000
1200
1400
Time in minutes - 24 hours (1440 min)

Clustering - Seasonal - summer Every File Load Profile - Season (17 out of 274 meters)

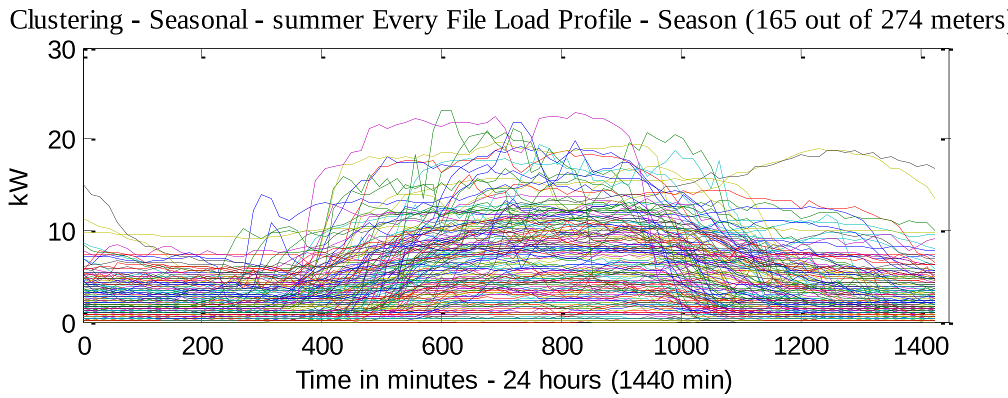


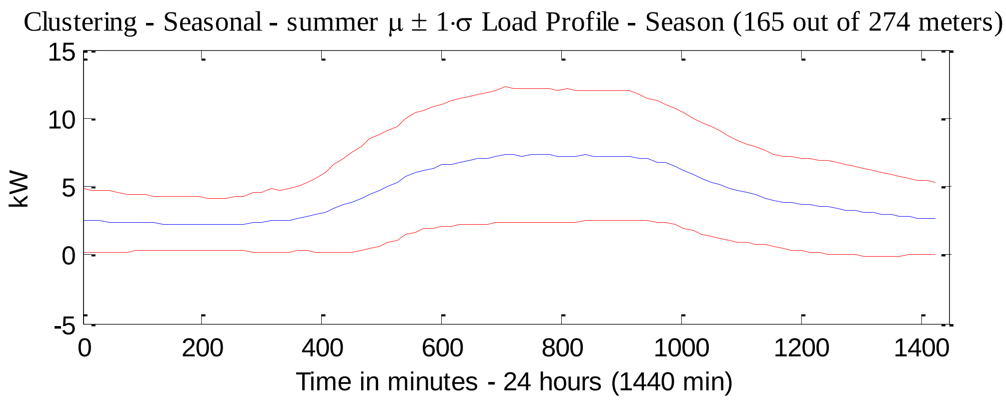
Clustering - Seasonal - summer μ ± 1·σ Load Profile - Season (17 out of 274 meters)





Clustering - Seasonal - summer Every File Load Profile - Season (5 out of 274 meters)
kW
300
200
100
0
0
200
400
600
800
1000
1200
1400
Time in minutes - 24 hours (1440 min)
Clustering - Seasonal - summer μ ± 1·σ Load Profile - Season (5 out of 274 meters)
kW
200
150
100
50
0
Time in minutes - 24 hours (1440 min)


Clustering - Seasonal - summer Every File Load Profile - Season (165 out of 274 meters)
kW
30
20
10
0
0
200
400
600
800
1000
1200
1400
Time in minutes - 24 hours (1440 min)


Clustering - Seasonal - summer μ ± 1·σ Load Profile - Season (165 out of 274 meters)
kW
15
10
5
0
-5
0
200
400
600
800
1000
1200
1400
Time in minutes - 24 hours (1440 min)

References

[1] Chicco, Gianfranco, Roberto Napoli, and Federico Piglione, "Load pattern clustering for short-term load forecasting of anomalous days," in *Proc. 2001 IEEE Power Tech*, Vol. 2, pp. 2-6.

[2] Yongqian Liu, Jie Shi, Yongping Yang, and Wei-Jen Lee, "Short-Term Wind-Power Prediction Based on Wavelet Transform–Support Vector Machine and Statistic-Characteristics Analysis," *Industry Applications, IEEE Transactions,* 48.4 (2012): 1136-1141.

[3] S. Fan, L. Chen, and W.-J. Lee, "Machine learning based switching model for electricity load forecasting,"2008*Energy Conversion and Management,* vol. 49, pp. 1331-1344.

[4] S. R. Gunn, "Support vector machines for classification and regression," Image Speech Intell. Syst. Res. Group, Univ. Southampton, Southampton, U.K., Tech. Rep., 1997.

[5] N. Vapnik, *The Nature of Statistical Learning Theory*. Hoboken, NJ: Wiley, 1999, pp. 156–160.

[6] I. Steinwart and A. Christmann, *Support Vector Machines*. New York: Springer-Verlag, 2008, pp. 333–354.

[7] Meng, M., Lu, J. C., and Sun, W, "Short-term load forecasting based on ant colony clustering and improved BP neural networks," in *2006 IEEE Machine Learning and Cybernetics International Conf.,* pp. 3012-3015.

[8] Li, Xiaoyan, and Ming Ding, "Load Model Based on Integrative K-means Clustering For Reliability Evaluation in Operational Planning,"*2010 IEEE Asia-Pacific Power and Energy Engineering Conference.*

[9] Mori, Hiroyuki, and Tadahiro Itagaki, "A precondition technique with reconstruction of data similarity based classification for short-term load forecasting,"*2004 IEEE Power Engineering Society General Meeting*.

[10] Qingming Wang, Jian Ren, Buxiang Zhou, and Zuo Li, "Forecasting of short-term load based on fuzzy clustering and improved BP algorithm,"*2011 IEEE Electrical and Control Engineering International Conference*.

[11] Tapas Kanungo, David M. Mount, Nathan S. Netanyahu, Christine D. Piatko, Ruth Silverman, and Angela Y. Wu, "An efficient k-means clustering algorithm: Analysis and implementation," *IEEE Trans. Pattern Analysis and Machine Intelligence, 24(7)*, pp. 881-892, 2002.

[12] Escobar, A. M., and L. P. Perez, "Application of support vector machines and ANFIS to the short-term load forecasting," *2008 IEEE/PES Transmission and Distribution Conference and Exposition*.

[13] Yu-Jun He, You-Chan Zhu, Dong-Xing Duan, and Wei Sun, "Application of neural network model based on combination of fuzzy classification and input selection in short term load forecasting,"*2006 IEEE Machine Learning and Cybernetics International Conference*.

[14] Huang, Yuan-Sheng, and Jia-Jia Deng, "Short-term load forecasting based on ant colony fuzzy clustering and SVM algorithm,"*2008 IEEE Natural Computation, Fourth International Conf.,* Vol. 2.

[15] Jain, Amit, and B. Satish, "Clustering based short term load forecasting using artificial neural network,"*2009 IEEE/PES Power Systems Conference and Exposition*.

[16] Jain, Amit, and B. Satish, "Short term load forecasting by clustering technique based on daily average and peak loads,"*2009 IEEE Power & Energy Society General Meeting*.

[17] Jinhui, Zhang, and Deng Jiajia, "Application of SVM based on rough sets to short-term load forecasting,"*2009 IEEE Intelligent Information Technology Application, Third International Symposium on* .Vol. 3.

[18] Junfang, Wu, and Niu Dongxiao, "Short-Term Power Load Forecasting Using Least Squares Support Vector Machines (LS-SVM),"*2009 IEEE Computer Science and Engineering. Second International Workshop on*. Vol. 1.

[19] Ye, Ning, Yong Liu, and Yong Wang, "Short-term power load forecasting based on SVM,"*2012 IEEE World Automation Congress*.

[20] Wang, Haifeng, and DejinHu," Comparison of SVM and LS-SVM for regression,"*2005 IEEE Neural Networks and Brain. International Conf.*,vol. 1.

[21] Jain, Amit, and B. Satish, "Clustering based short term load forecasting using support vector machines," *2009 IEEE Bucharest Power Tech Conference*.

[22] Guo, Ying-Chun, and Dong-Xiao Niu, "Short-Term Load Forecasting Using Support Vector Regression Based on Pattern-Base," *2009 IEEE Intelligent Information and Database Systems. First Asian Conference*.

[23] Xinhui D, Liang W, Jiancheng S, et al,"Application of neural network and support vector machines to power system short-term load forecasting*," 2010 IEEE Computational Aspects of Social Networks (CASoN) International Conf: 729-732.*

[24] Esteves, R. M., Pais, R., &Rong, C. (2011, March), "K-means clustering in the cloud--a Mahout test,"*2011 IEEE Workshops of International Conference on Advanced Information Networking and Applications (WAINA): pp. 514-519.*

[25] Tangkitjaroenmongkol R, Kaittisin S, Ongwattanakul S, "Inbound logistics cassava starch planning: With application of GIS and K-means clustering methods in Thailand," *2011 IEEE Computer Science and Software Engineering (JCSSE) Eighth International Joint Conf: 204-209.*

[26] Holanda Filho R, Maia J E B, "Network traffic prediction using PCA and K-means,"*2010 IEEE Network Operations and Management Symposium (NOMS): 938-941.*

[27] Gao X, Wang P, Qi Y, et al,"Comparison studies of LS_SVM and SVM on modeling for fermentation processes," *2009 IEEE Natural Computation Fifth International Conf: 478-484.*

[28] Wang B, Shi Q, Mei Q,"Improvement of LS-SVM for time series prediction,"*2014 IEEE Service Systems and Service Management (ICSSSM) 11th International Conf: 1-5.*

[29] Li Y, He C, Li B, et al, "An efficient computational model for LS-SVM and its applications in time series prediction,"*2010 IEEE Computer Science and Education (ICCSE) 5th International Conf: 467-470.*

[30] Yang X, "Comparison of the LS-SVM based load forecasting models,"*2011 IEEE Electronic and Mechanical Engineering and Information Technology (EMEIT) International Conf: 2942-2945.*

[31] Jain A, Satish B,"Integrated approach for short term load forecasting using SVM and ANN,"*2008 IEEE TENCON: 1-6.*

[32] Yan X, Chowdhury N A,"A comparison between SVM and LSSVM in mid-term electricity market clearing price forecasting,"*201326th Annual IEEE Canadian Conference Electrical and Computer Engineering (CCECE): 1-4.*

[33] Mustafa M W, Sulaiman M H, Khalid S A, et al, "Real power transfer allocation via Continuous Genetic Algorithm-Least Squares Support Vector Machine technique," *2010 IEEE Power and Energy (PECon), Conf: 12-17.*

[34] Yu G, Soh L K, Bond A,"K-means clustering with multiresolution peak detection," *2005 IEEE Electro Information Technology Conf: 6 pp.-6.*

[35] Schölkopf B, Platt J, Hofmann T, "Stability of k-means clustering,"*2007 IEEE Advances in Neural Information Processing Systems: 1121-1128.*
[36] Yu Z, Wong H S, "Genetic-based K-means algorithm for selection of feature variables,"*2006 IEEE Pattern Recognition: 744-747.*
[37] Li G, Cheng C, Lin J, et al, "Short-term load forecasting using support vector machine with SCE-UA algorithm,"*2007 IEEE Third International Conference on Natural Computation, 1: 290-294.*
[38] *Kardi Teknomo, "Support Vector Machines Tutorial" ebook, http://people.revoledu.com/kardi/tutorial/SVM/purchase.html*
[39] Xin Wang*, et al.*, " Factors that Impact the Accuracy of Clustering Based Load Forecasting," 2016 IEEE Transactions on Industry Application.
[40]R. G. Pratt*, et al.*, "Metered end-use consumption and load shapes from the ELCAP residential sample of existing homes in the Pacific Northwest," *Energy and Buildings,* vol. 19, pp. 179-193, 1993.
[41]*Consolidated Edison Company of New York, 2012 Scheduled Holiday Day-Types*. Available: https://apps.coned.com/radata/HolidayDayType.doc
[42]NOAA. *Why Do We have Seasons?* Available: http://www.crh.noaa.gov/lmk/?n=seasons
[43]R. Xu and D. C. Wunsch, *Clustering - IEEE Press Series on Computational Intelligence*. Piscataway, N.J. Hoboken, N.J.: IEEE Press, John Wiley & Sons, Inc., 2009.
[44] B. Everitt, *Cluster analysis*, 5th ed. Chichester, West Sussex, U.K.: Wiley, 2011.
[45] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognition Letters,* vol. 31, pp. 651-666, 2010.
[46] G. Gan*, et al.*, *Data clustering : theory, algorithms, and applications*: SIAM, American Statistical Association, 2007.
[47] A. K. Jain*, et al.*, "Data clustering: a review," *ACM Comput. Surv.,* vol. 31, pp. 264-323, 1999.
[48] S. Das*, et al.*, *Metaheuristic clustering*. Berlin: Springer, 2009.
[49] B. J. Frey and D. Dueck, "Clustering by Passing Messages Between Data Points," *Science,* vol. 315, pp. 972-976, February 16, 2007 2007.
[50] J. Handl*, et al.*, "Computational cluster validation in post-genomic data analysis," *Bioinformatics,* vol. 21, pp. 3201-3212, 2005.
[51] A. K. Jain and R. C. Dubes, *Algorithms for clustering data*. Englewood Cliffs, NJ: Prentice Hall, 1988.
[52] Q. Zhao, "Cluster Validity in Clustering Methods," Ph.D. Dissertation, Faculty of Science and Forestry, School of Computing, University of Eastern Finland, 2012.
[53] Jinhui, Zhang, and Deng Jiajia, "Application of SVM based on rough sets to short-term load forecasting," *2009 IEEE Intelligent Information Technology Application, Third International Symposium on.* Vol. 3.
[54] Junfang, Wu, and Niu Dongxiao, "Short-Term Power Load Forecasting Using Least Squares Support Vector Machines (LS-SVM)," *2009 IEEE Computer Science and Engineering. Second International Workshop on*. Vol. 1.

Biographical Information

Xin Wang received the Bachelor's degree in Electrical Engineering from North China Electric Power University, China, in 2012.

In August 2012, she started PhD program in the area of power systems at the University of Texas at Arlington (UTA). She is a member of the research group at the Energy Systems Research Center (ESRC) at UTA. Her areas of interest are power system analysis, planning, operation, load forecasting, renewable energy and electricity market.