

2019 年度

PhD Thesis

博士学位論文

Application of Clustering Method for Energy Pattern Analysis and  
Energy Management in University Buildings

大学建物におけるエネルギーパターン分析とエネルギー管理の  
ためのクラスタリング手法の応用

中部大学大学院

工学研究科 建設工学専攻

Bishnu Nepal

## *Acknowledgements*

Undertaking this PhD has been a truly a life changing experience for me and it would not have been possible to do without the support and guidance that I receive from many people. Firstly, I would like to express my sincere gratitude to my Prof. Motoi Yamaha for the continuous support of my PhD study and related research, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my PhD thesis. He also helped me financially by providing the job of research assistant and helped to improve my Japanese.

Besides my advisor, I would like to thank associate Prof. Aya Yokoe for insightful comments and encouragement.

I am very grateful to Dr. Kyoko Shuku who gave me shelter in her mansion and always encouraged to be positive and move forward. She was like a guardian to me and in her presence I always had a feeling of support and protection. Without her support perusing PhD would be difficult.

My sincere thanks goes to Dr. Mitsugu Kawamura for continuous support and suggestion regarding my research. He also helped by providing the necessary data and comments for my research.

I am also thankful to friends Mr. Sahashi and Mr. Yamaji from my laboratory who introduced me different aspects of Japanese culture and supported with my Japanese and also my research activities. Similar thanks goes to my friend Mr. Lyujunlin, Mr. Wanghui, Mr. Yang for making my stay in Japan so wonderful.

I am greatly thankful to Mr. Simamura from the Facility Administration Department of Chubu University for providing data and advices for my research.

At last but not the least, I would like to thank my family members for the continuous support and encouragement for making this journey of perusing my PhD successful.

## Abstract

The energy demand in the building sector accounts for more than 30% of the total energy use and more than 55% of the global electricity demand. The energy demand of the building sector is increasing rapidly, driven by the improved access to energy in developing countries, greater ownership and use of energy-consuming devices, and rapid growth in building floor area. Efforts to develop sustainable buildings are progressing but are still not keeping pace with the growing building sector and the rising demand for energy.

Analyzing the energy use pattern of buildings and planning for energy conservation is essential for the conservation of energy in buildings. In this research, we used K-means clustering method for the analysis of energy consumption patterns in buildings. Initial centroids in K-means clustering are chosen randomly so the clustering number changes every time in the default process. To remove this instability, we selected the initial centroids using a percentile method based on the empirical cumulative distribution. As a result, the initial centroids were fixed irrespective of the number of trials. The result produced by this method had better accuracy, and the internal cohesion and separation between the cluster were better than the random initialization method. In comparing with the university schedule, six clusters were found to be appropriate so, the whole one-year electricity consumption data were classified into six clusters.

Knowing the energy consumption pattern of buildings is essential for the conservation of energy in buildings. Thus in this research, we used the clustering method to classify the energy consumption of university buildings into the base energy consumption, human activity factor energy consumption and air-conditioning factor energy consumption. This classification helps in deciding which factor of the building should be given priority for energy conservation.

On analyzing the buildings of Engineering department, Administrative buildings and Business Administration department it was found that Engineering department has large base energy consumption and the difference between the base energy of clusters were found high whereas in the case of Administrative buildings and buildings of Business Administration department base energy consumption was found small and the difference between the base energy of cluster 1 and 6 was found very small. On analyzing the peak of peak days and the base of peak days, it was found that the peak electricity consumption of buildings of Science departments was found to increase with the increase in base electricity consumption whereas in the case of buildings of non-science department no such relation was observed. Base energy consumption reduction in the case of buildings of the engineering department was found higher whereas, peak energy consumption reduction in the case of buildings of Administrative Buildings and buildings of Business Administration department is found higher.

Peak load reduction is also another important issue for the management of energy in university buildings. In the year 2018, in the air-conditioning peak days, the electricity consumption in the West Campus of Chubu University exceeded the contracted value of peak demand. This encouraged us for the development of the forecasting method which produce better forecasting result than the existing methods. In this paper, we proposed a hybrid model, a combination of clustering and ARIMA model for the electricity load forecasting in the university buildings. The novel approach discussed in this paper includes clustering one whole year data including the forecasting day using the K-means clustering method and using the clustering result for forecasting the electricity load in the buildings. The combination of clustering and the ARIMA model has proved to increase the performance of forecasting rather than the ARIMA model alone. Forecasting electricity load with appreciable accuracy several hours before peak hours can provide the management authorities with sufficient time for making a strategy for peak load reduction. This method can also be implemented in the demand response for reducing the electricity bills by avoiding electricity usage during the high electricity rate hours. As the application of the forecasting model, we developed a visualization platform using Shiny App. Using Shiny we can interactively analyze the energy consumption of the desired building of the university in the form of trend graph, heatmap, cluster centers and calendar plot of the clustering result and the forecasting result. Viewing interactively the forecasted value of electricity consumption of different buildings using web-based server such as AWS by the energy conservation group members and sending a request mail to act for the peak load reduction depending on the status of the forecasted value is believed to help in the peak load reduction in university buildings.

## 概要

建築部門のエネルギー需要は、総エネルギー使用量の 30%以上、世界の電力需要の 55%以上を占めている。建築部門のエネルギー需要は、発展途上国におけるエネルギーへのアクセスの改善、エネルギー消費機器の所有と使用の拡大、および建物の床面積の急速な成長により、急速に増加している。持続可能な建築物を開発する取り組みは進んでいるが、建築部門の成長とエネルギー需要の増加に追いついていない。

建物のエネルギー使用パターンを分析し、省エネルギーの計画を立てることは、建物の省エネルギーに重要である。本研究では、建物のエネルギー消費パターンを分析するためクラスタリングの K 平均法を使用した。K 平均法の初期重心はランダムに選択されるため、デフォルトのプロセスではクラスタリングの番号が毎回変化する。この不安定性を取り除くために、累積度数分布に基づくパーセントイル法を使用して初期重心を決定した。その結果、最初の重心は試行回数に関係がなくなるように修正された。この方法で生成された結果の精度は高く、クラスタ間の内的結合と分離はランダム初期化方法よりも優れていた。大学のスケジュールと比較すると、6つのクラスタが適切であることがわかったため、1年間の電力消費データ全体を6つのクラスタに分類した。

建物の省エネルギーには要因別エネルギー消費を分類することも重要である。そこで本研究では、クラスタリング手法を用いて、大学の建物のエネルギー消費量を、ベースエネルギー消費量、人間活動分と空調因分に分類した。この分類は、建物のどの要素に省エネを優先させるかを決定するのに重要であると考えられる。

工学部、事務建物、経営経営情報学部の建物を分析したところ、工学部のベースエネルギー消費量が大きく、クラスタ毎のベースエネルギーの差が大きいことがわかった。一方、事務建物と経営情報学部の建物の場合ベースエネルギー消費は小さく、ベースエネルギーを示すクラスタ 1 とクラスタ 6 の差は非常に小さいことがわかった。ピーク日のピークとベースを分析すると、理系建物のピーク電力消費量はベース電力消費量の増加とともに増加することがわかったが、文系建物の場合はそのような関係は観察されなかった。工学部の建物はベースエネルギー消費の削減の可能性が高く、事務系建物と経営情報学部の建物はピークエネルギー消費の削減の可能性が高いことがわかった。

ピーク負荷の削減は、大学の建物のエネルギー管理には重要な問題である。2018年には、空調のピーク時に、中部大学西キャンパスの電力消費量がピーク需要の契約値を超えていた。これにより、既存の方法よりも優れた予測結果を生成する予測方法の開発が必要であ

る。本論文では、大学の建物の電力負荷予測のために、クラスタリングと ARIMA モデルを組み合わせたハイブリッドモデルを提案した。この方法では、K-means クラスタリング手法を使用して予測日を含む 1 つの年間データをクラスタリングし、クラスタリング結果を使用して建物の電力負荷を予測するものである。クラスタリングと ARIMA モデルの組み合わせにより、ARIMA モデルのみではなく、予測の精度が向上することが確認された。ピーク時間の数時間前にかかなりの精度で電力負荷を予測することにより、ピーク負荷削減の戦略を立てることができる。この方法は、電気料金が低い時間帯に電力の使用を回避することにより、電気料金を削減するためのデマンドレスポンスにも応用できる。予測モデルのアプリケーションとして、Shiny App を使用した視覚化プラットフォームを開発した。Shiny を使用すると、クラスタリング結果と予測結果のトレンドグラフ、ヒートマップ、クラスタセンター、カレンダープロットのグラフを作成し、大学の目的の建物のエネルギー消費量をインタラクティブに分析できる。エネルギー管理部署が AWS などの Web ベースのサーバーを使用して、さまざまな建物の消費電力の予測値をインタラクティブに表示し、予測値に応じてピーク負荷削減に対処するための方法をメールを送信することによって大学の建物のピーク負荷削減に役立つと考えられる。

## Abbreviations

BEMS:	Building Energy Management System
PAM:	Partitioning Around Medoids
CLARA:	Clustering Large Applications
STING:	Statistical Information Grid
CLIQUE:	Clustering In QUEst
KPCA:	Kernel Principal Component Analysis
PCA:	Principal Component Analysis
ARIMA:	Autoregressive Integrated Moving Average
SARIMA:	Seasonal Autoregressive Integrated Moving Average
FOM:	Figure Of Merit
EHP:	Electric Heat Pump
RMSE:	Root Mean Square Error
MAPE:	Mean Absolute Percentage Error
MAE:	Mean Absolute Error
DBN:	Deep Belief Network
SVR:	Support Vector Regression
LR:	Linear Regression
MA:	Moving Average
ADF:	Augmented Dickey Fuller
ACF:	Autocorrelation Function
PACF:	Partial Autocorrelation Function
AIC:	Akaike Information Criteria

# Table of Contents

Acknowledgement.....	II
Abstract.....	III
Abbreviation.....	V
List of figures.....	IX
List of Tables.....	XIV
Outline.....	XV

## Chapters

<b>Chapter 1. Introduction.....</b>	<b>1</b>
1.1 World energy.....	1
1.2 Energy scenario of Japan.....	2
1.3 Building Energy consumption.....	3
1.4 Efforts taken by Chubu University for Energy conservation.....	3
1.5 The reason for using clustering method for electricity consumption analysis.....	8
<b>Chapter 2. Clustering and proposed method.....</b>	<b>10</b>
2.1 Clustering.....	10
2.1.1 Partitioning methods.....	10
2.1.2 Hierarchical methods.....	10
2.1.3 Density Based Clustering.....	11
2.1.4 Grid Based Clustering.....	11
2.2 Literature review of clustering.....	12
2.3 Outlier detection and imputation of missing values.....	13
2.4 Overview of K-means Clustering Algorithm.....	15
2.5 Proposed method.....	25
2.6 Experiments and Results.....	19
2.6.1 Accuracy verification.....	19
2.6.1.1 Accuracy verification using Real World datasets.....	19
2.6.2 Cluster quality comparison using university data.....	20



2.7 Determination of number of clusters.....	22
2.8 Statistical method to find the number of clusters.....	28
<b>Chapter 3. Energy pattern analysis using proposed k-means clustering.....</b>	<b>30</b>
3.1 Estimating energy consumption by application.....	30
3.1.1 Data used for the analysis.....	31
3.2 Clustering result.....	31
3.3 Application of Energy Management.....	36
3.3.1 Estimation energy consumption by factor.....	36
3.4 Accuracy measurement of the Clustering analysis result.....	39
3.5 Utilization of energy management.....	43
3.6 Energy consumption pattern in different buildings.....	44
3.6.1 Office.....	44
3.6.2 Hospitals .....	45
3.6.3 Departmental stores .....	45
3.6.4 Hotels.....	46
3.7 Analysis of energy consumption pattern of university buildings using clustering technique.....	46
3.8 Yearly change in energy consumption pattern of Chubu University.....	56
3.9 Comparison between gas air-conditioning usage building and electricity usage air-conditioning building.....	58
3.9.1 Analysis of annual change in energy consumption pattern.....	58
<b>Chapter 4. Electricity load forecasting.....</b>	<b>66</b>
4. Introduction.....	66
4.1 Motivation for electricity load forecasting.....	67
4.2 Literature review.....	68
4.3 ARIMA model.....	69
4.4 Target of the research.....	70
4.5 Methodology.....	71
4.6 Fitting of the ARIMA model.....	75

4.7 Results.....	78
4.8 Discussion .....	86
4.9 Measures for electricity peak reduction .....	86
4.10 Measures for base energy reduction.....	87
<b>Chapter 5. Data analysis and visualization using Shiny.....</b>	<b>88</b>
5.1 R programming language.....	88
5.2 Shiny.....	88
5.2.1 Structure of the Shiny App.....	89
5.3 Shiny as a Visualization platform.....	91
<b>Chapter 6. Conclusion.....</b>	<b>97</b>
<b>References.....</b>	<b>99</b>

## List of Figures

<b>Figure 1.1</b> Share of Renewables in Net Annual Additions of Power Generating Capacity, 2008-2018.....	1
<b>Figure 1.2</b> World electricity generation by fuel and Share of net electricity generation by fuel.....	2
<b>Figure 1.3</b> 1.3(a) Status of Japan`s energy scenario from the year 2000 to 2015. ....	2
1.3(b) Pie chart showing the ratio of energy source of the year 2015.....	3
<b>Figure 1.4</b> Overview of different departments of Chubu University.....	4
<b>Figure 1.5</b> Amount of CO2 emissions from educational facilities (in Tokyo) and the position of Chubu University in the year 2010. ....	4
<b>Figure 1.6</b> Overview of power saving by smart grid system and power generation.....	5
<b>Figure 1.7</b> Smart grid monitor displaying the status of energy consumption in Chubu University in the real time.....	6
<b>Figure 1.8</b> Map of Chubu University with the location of departments in East campus grid and West campus grid.....	7
<b>Figure 1.9</b> The control system for smart grid.....	7
<b>Figure 1.10</b> Amount of CO2 emissions from educational facilities (in Tokyo) and the position of Chubu University in the year 2016. ....	8
<b>Figure 1.11</b> Classification of electricity consumption of different departments of Chubu University based on the representative weeks. ....	9
<b>Figure 2.1</b> Boxplot with outliers.....	14
<b>Figure 2.2</b> Boxplot after outliers were removed.....	15
<b>Figure 2.3</b> Flow chart of the K-means algorithm.....	16
<b>Figure 2.4</b> Electricity use in Chubu University in different electricity use periods.....	17
(a) Sunday and holidays, (b) Saturday, (c) Lecture day without air-conditioning	
(d) Lecture day with air-conditioning, (e) Lecture day with air-conditioning peak	
(f) Non-lecture days with air-conditioning	
<b>Figure 2.5</b> Selection of the initial centroid by using the percentile method.....	18
<b>Figure 2.6</b> Model of $366 \times 24$ order matrix.....	21

<b>Figure 2.7</b>	Part of the data frame of the $366 \times 24$ order matrix.....	21
<b>Figure 2.8</b>	Clustering result of Chubu University.....	23
	(a) Cluster center for $k = 3$ , (b) Cluster centers for $k = 4$	
	(c) Calendar plot for $k = 3$ , (d) Calendar plot for $k = 4$	
<b>Figure 2.9</b>	Clustering result of Chubu University.....	25
	(a) Cluster center for $k = 5$ , (b) Cluster centers for $k = 6$	
	(c) Calendar plot for $k = 5$ , (d) Calendar plot for $k = 6$	
<b>Figure 2.10</b>	Actual calendar plot of Chubu University.....	26
<b>Figure 2.11</b>	The number of days per cluster by the day of the week.....	27
<b>Figure 2.12</b>	Internal validation measure using Dunn index.....	28
<b>Figure 2.13</b>	Stability validation measure using AD.....	29
<b>Figure 2.14</b>	Stability validation measure using FOM.....	29
<b>Figure 3.1</b>	Electricity consumption pattern of peak period by conventional method (superposition of electric consumption for representative week) .....	30
<b>Figure 3.2</b>	electricity consumption pattern during peak summer determined by clustering.....	31
<b>Figure 3.3</b>	Clustering result: whole university.....	32
<b>Figure 3.4</b>	Clustering result: Department of Engineering.....	33
<b>Figure 3.5</b>	Clustering result: Business administration.....	33
<b>Figure 3.6</b>	Clustering result: International Studies.....	33
<b>Figure 3.7</b>	Clustering result: Humanities.....	34
<b>Figure 3.8</b>	Clustering result: Contemporary Education.....	34
<b>Figure 3.9</b>	Clustering result: Offices.....	34
<b>Figure 3.10</b>	Clustering result: Bioscience and Biotechnology.....	35
<b>Figure 3.11</b>	Clustering result: Life and Health Science.....	35
<b>Figure 3.12</b>	Whole university, relation between cluster and schedule.....	36
<b>Figure 3.13</b>	Estimated primary energy consumption factor.....	39
<b>Figure 3.14</b>	cluster centers: building no. 52.....	40
<b>Figure 3.15</b>	Calendar plot: building no. 52.....	40
<b>Figure 3.16</b>	One-year average energy consumption of consent of building no. 52.....	41

<b>Figure 3.17</b> One-year average hourly-energy consumption of lighting of building no.52.....	42
<b>Figure 3.18</b> Ratio of energy consumption by factor for measured and clustering result for building no. 52.....	42
<b>Figure 3.19</b> Energy consumption in Offices.....	44
<b>Figure 3.20</b> Energy consumption in Hospitals.....	45
<b>Figure 3.21</b> Energy consumption in departmental stores.....	45
<b>Figure 3.22</b> Energy consumption in hotels.....	46
<b>Figure 3.23</b> Calendar plot for building no. 3-7-8 of College of Engineering.....	47
<b>Figure 3.24</b> Cluster center for building no. 3-7-8 of College of Engineering.....	48
<b>Figure 3.25</b> Percentage of energy consumption per cluster of building no. 3-7-8 of College of Engineering.....	48
<b>Figure 3.26</b> Calendar plot for building no.1of the office building.....	49
<b>Figure 3.27</b> Cluster centers for building no. 1 of the office building.....	49
<b>Figure 3.28</b> Percentage of energy consumption per cluster in building no. 1 of the office building.....	50
<b>Figure 3.29</b> Calendar plot for building no. 21 of College of Business.....	51
<b>Figure 3.30</b> Cluster centers for building no. 21 of College of Business.....	51
<b>Figure 3.31</b> Percentage of energy consumption per cluster of building no. 21 of the College of Business.....	52
<b>Figure 3.32</b> Classification of energy consumption by usage in three different buildings of Chubu University.....	53
<b>Figure 3.33</b> Bar graph indicating the yearly base energy consumption per unit area in the three different buildings of Chubu University.....	54
<b>Figure 3.34</b> Boxplot showing the ratio between peak and base energy for three buildings as Chubu University.....	55
<b>Figure 3.35</b> Scatter plot showing the relation between base and peak energy consumption in buildings of Science and Non-science departments.....	55
<b>Figure 3.36</b> Change in yearly energy consumption by factor in whole university.....	57
<b>Figure 3.37</b> Change in yearly gas consumption (m <sup>3</sup> ) in whole university.....	57
<b>Figure 3.38</b> Yearly change in energy consumption by factor in building no. 20.....	58

<b>Figure 3.39</b>	Yearly change in energy consumption by factor in building no. 52.....	59
<b>Figure 3.40</b>	Peak air-conditioning energy consumption of cluster 6 of building no. 20 of the year 2014 and 2015.....	59
<b>Figure 3.41</b>	Cluster centers of building no. 20 of the year 2014.....	60
<b>Figure 3.42</b>	Cluster centers of building no. 20 of the year 2015.....	60
<b>Figure 3.43</b>	Calendar plot of building no. 20 of the year 2014.....	61
<b>Figure 3.44</b>	Calendar plot of building no. 20 of the year 2015.....	62
<b>Figure 3.45</b>	Cluster centers of building no. 52 of the year 2015.....	63
<b>Figure 3.46</b>	Cluster centers of building no. 52 of the year 2017.....	63
<b>Figure 3.47</b>	Calendar plot of building no. 20 of the year 2015.....	64
<b>Figure 3.48</b>	Calendar plot of building no. 20 of the year 2017.....	65
<b>Figure 4.1</b>	A typical example load forecasting and its forecasting error.....	67
<b>Figure 4.2</b>	Electricity load graph with peak load crossing the demand electricity load in West Campus of Chubu University.....	67
<b>Figure 4.3</b>	Map of Chubu University with East-campus in red boundaries.....	70
<b>Figure 4.4</b>	Line graph of whole electricity load of academic year 2018.....	71
<b>Figure 4.5</b>	Flow chart of Proposed method.....	72
<b>Figure 4.6</b>	Electricity load of cluster center 1 to 6.....	73
<b>Figure 4.7</b>	Electricity load distribution of cluster centers using box plot .....	74
<b>Figure 4.8</b>	ACF of the training data.....	75
<b>Figure 4.9</b>	PACF of the training data.....	76
<b>Figure 4.10</b>	Residual analysis for the ARIMA (2, 0, 1) (2, 0, 1) <sub>24</sub> fit to the data.....	77
<b>Figure 4.11</b>	Calendar plot of clustering result with number of clusters k = 6.....	78
<b>Figure 4.12.</b>	Forecasting result ARIMA model.....	79
<b>Figure 4.13</b>	Forecasting result proposed method.....	79
<b>Figure 4.14</b>	Forecasting result of ARIMA and proposed method in comparison to actual data on 27 August 2018.....	80
<b>Figure 4.15</b>	MAPE value for ARIMA and Proposed method.....	82
<b>Figure 4.16</b>	RMSE value for ARIMA and Proposed method.....	83
<b>Figure 4.17</b>	MAE value for ARIMA and Proposed method.....	83

<b>Figure 4.18</b> Forecasting using ARIMA model.....	84
<b>Figure 4.19</b> Forecasting using proposed model.....	84
<b>Figure 4.20</b> Forecasting result of ARIMA and proposed method removing Saturday and Sunday on the training data. ....	85
<b>Figure 5.1</b> Example of graphical visualization of data in the form of histogram using Shiny.....	89
<b>Figure 5.2</b> Image of Shiny Building Energy Analysis tool.....	91
<b>Figure 5.3</b> The date selection feature of Shiny App.....	92
<b>Figure 5.4</b> Building selection feature of Shiny App.....	92
<b>Figure 5.5</b> Producing line graph and heatmap using Shiny App.....	93
<b>Figure 5.6</b> Producing multiple building line graph using Shiny App.....	93
<b>Figure 5.7</b> Selection of the range of Y-axis feature of Shiny App.....	94
<b>Figure 5.8</b> Line graph after the range of Y-axis changed.....	94
<b>Figure 5.9</b> Production of cluster centers and calendar plot using Shiny App.....	95
<b>Figure 5.10</b> Forecasting result using Shiny App.....	96

## List of Tables

<b>Table 2.1</b> Resulting accuracy of different dataset.....	19
<b>Table 2.2</b> Comparison of K-means algorithm with different datasets.....	20
<b>Table 2.3</b> Comparison of K-means algorithm with random initialization and the proposed method.....	21
<b>Table 2.4</b> Characteristics of electricity use in Chubu University.....	22
<b>Table 3.1</b> Correspondence between departments and buildings.....	32
<b>Table 3.2</b> Correspondence between clusters and university schedules.....	37
<b>Table 3.3</b> Cluster used to calculate energy consumption by factor.....	38
<b>Table 3.4</b> Possible measures for CO <sub>2</sub> emission reduction.....	43
<b>Table 3.5</b> Energy consumption by factor in whole university from the year 2015-2018.....	56
<b>Table 4.1</b> AIC values of suggested ARIMA models.....	76
<b>Table 4.2</b> Forecasting result suggested ARIMA model and proposed method.....	81
<b>Table 4.3</b> Error values of forecasting result by ARIMA model and proposed method.....	82



# Outline

## Chapter 1

This chapter is mainly the introduction related to energy. Present energy status of the world and Japan, as well as the energy-related challenges, has been discussed. The efforts taken by Chubu University for energy conservation has been discussed. The reason for using the clustering method for electricity consumption analysis has been explained.

## Chapter 2

This chapter explains the various types of clustering algorithms. The literature review section explains about different works done in the field of clustering and cluster center initialization. Method for the detection of outliers and their imputation method is explained. The proposed method for the determination of better initial centroid is discussed. It is shown that the accuracy of classification using the proposed method is better than the random initialization method and this method can solve the problem of empty clusters. Considering the presence or absence of air-conditioning, presence or absence of lectures, energy consumption on Saturdays and peak energy consumption, six clusters in the case of Chubu University is found to be appropriate.

## Chapter 3

This chapter is the application part of the proposed method explained in chapter 2. Using proposed K-means clustering the electricity consumption data of the whole university as well as departments are divided into six clusters and their cluster centers and calendar plot is drawn. Based on the university schedule, and from the difference of the energy of the clusters, the energy consumption by factor is calculated which includes base energy consumption, human activity factor energy consumption and air-conditioning factor energy consumption. The accuracy of classification is compared with the actual measurement at the building no.52 of College of Life and Health Sciences. As a result, energy consumption by factor using K-means clustering was found similar to the actual measurement. The energy consumption pattern of offices, departmental stores, hotels, and hospitals is described. The energy consumption of three different buildings belonging to the College of Engineering, College of Business and Administrative building is analyzed and the relationship between their base and the peak energy consumption is analyzed. The annual change in energy consumption pattern of gas air-conditioning usage building and electricity air-conditioning usage building is studied and the effect of EPH renewal on the peak energy consumption is analyzed

## **Chapter 4**

This chapter is related to electricity load forecasting in university buildings. We proposed a method, a hybrid model of clustering and the ARIMA model that produces better forecasting than the existing forecasting models. Electricity consumption data of the East campus of Chubu University is taken as the target for the analysis. To make the selection of parameters  $(p, d, q)$  of the ARIMA model automatic, `auto.arima` function of R programming language is used. Clustering of one whole year data including the forecasting day from 6 to 9 AM is done and the result of the clustering is used in the ARIMA model to produce a 24 hour ahead forecasting. The performance of forecasting was calculated using RMSE, MAPE, and MAE. Forecasting of electricity load data from 1<sup>st</sup> April 2018 to 31<sup>st</sup> March 2019 was done and the average value of MAPE, RMSE and MAE is calculated. The forecasting result using the proposed hybrid model was found to be more accurate than the ARIMA model.

## **Chapter 5**

This chapter is related to data analysis and visualization using Shiny App. Shiny is an R package that makes it easy to build interactive web-based apps straight from R. Using Shiny, the result of the analyzed data can be viewed in the form of the line graph, heatmap, cluster centers, and calendar plot. Once the data is deployed into the web-based server such as AWS, real-time forecasting of the analyzed result can be viewed and real time comparison between the actual and forecasted value can be done. Shiny App can be used for energy conservation in university buildings.

## **Chapter 6**

This chapter is the conclusion of this thesis.

# Chapter 1

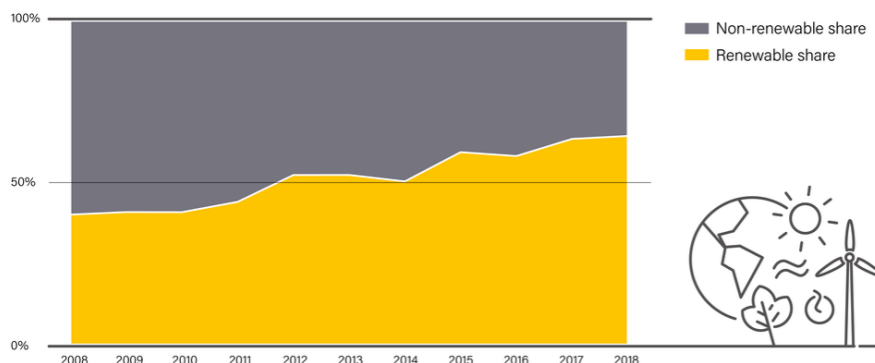
## Introduction

### 1.1 World energy

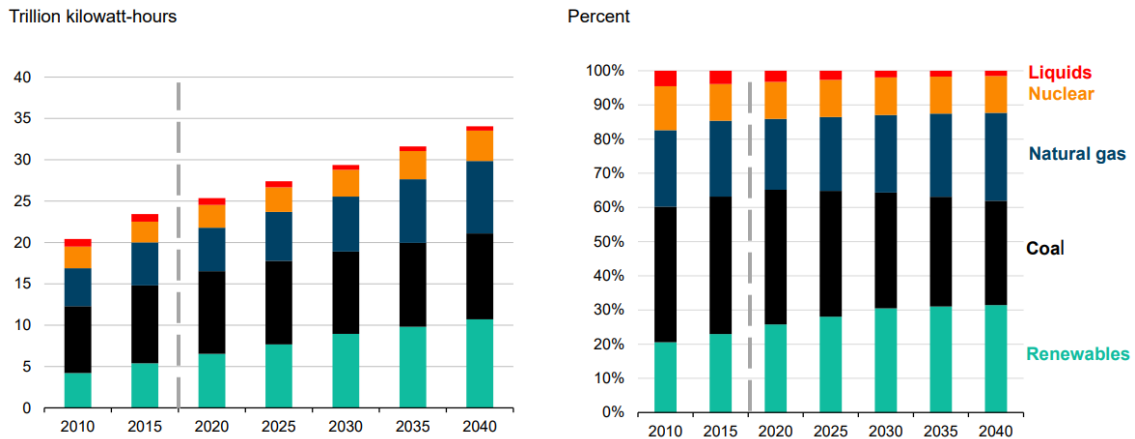
Present time the world has two major energy related challenges. One of the major challenge is to meet the rising energy demand of the world and the next is the reduction of carbon emission from the fossil fuels. The center of gravity of the energy demand is shifting, with the expanding middle class in Asia accounting for much of the growth in the global GDP. Meanwhile, the way in which the energy is being consumed is changing in real time, as the world electrifies and energy increasingly becomes part of broader service that are bought and sold in even more competitive and effective digital markets.

According to 2018 report [1] in 2018 global energy demand and carbon emission use grew at their fastest rate since 2010/11, moving even further away from the accelerated transition envisaged by the Paris climate goals. The report stated that much of the rise in energy growth was due to the increased demand for cooling and heating due to weather related effects. China, the United States and India together accounted for almost 70% of the total increase in demand. Due to rise in fossil fuel consumption, global energy related carbon dioxide emissions grew an estimated 1.7% during the year [2].

One positive aspect of the world energy is that increasing invest in the renewables. Figure 1.1 shows that around 2/3 of investment in the recent years are in the renewable sector. Solar and wind power are now the main stream option in the power sector, with an increasing number of countries generating more than 20% of their electricity with solar and wind power. But, current trends show that boulder policy decision are needed across all sector of energy end-use to make our energy system sustainable [2].



**Figure 1.1** Share of Renewables in Net Annual Additions of Power Generating Capacity, 2008-2018. Figure adopted from [2].

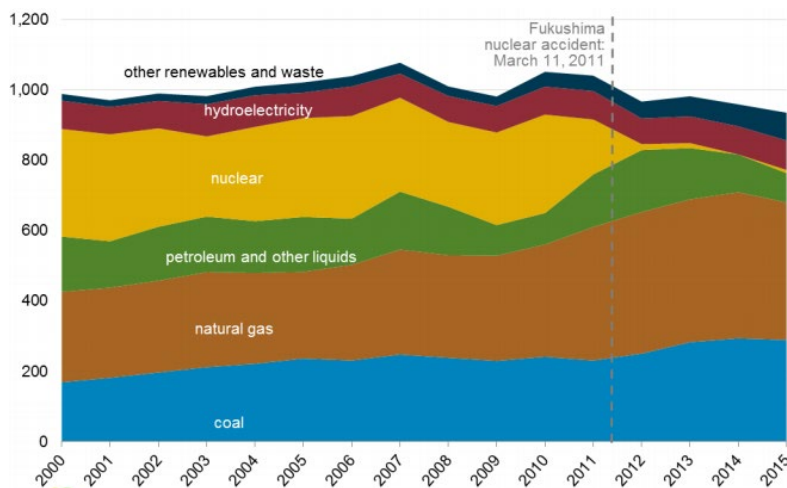


**Figure 1.2** World electricity generation by fuel (left) and Share of net electricity generation by fuel (right). Figure adopted from [3].

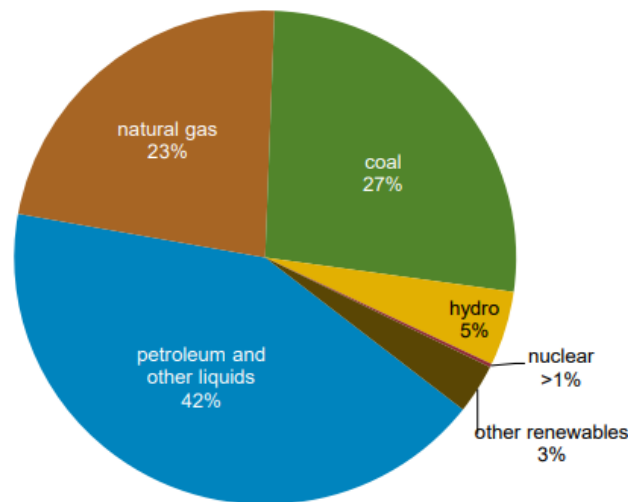
Figure 1.2 show that despite of rise in renewable energy and decrease in the dependence in fossil fuels, fossil fuels will still constitute about 50% of the energy market until 2040.

### 1.2 Energy scenario of Japan

Up until 2011, Japan was generating about 30% of electricity from nuclear power and this was expected to increase to at least 40% by 2017. The plan is now for at least 20% by 2030, form a depleted fleet. Recent Japan needs to imports about 90% of its energy requirement. Currently 37 reactors are operable. The first two restarted in August and October 2015, with a further seven having restarted. 17 reactors are currently in the process of restart approval [4].



**Figure 1.3(a):** Status of Japan`s energy scenario from the year 2000 to 2015.



**Figure 1.3(b).** Pie chart showing the ratio of energy sources of the year 2015

Figure 1.3 (a) states Japan`s total energy consumption, from 2000-2015 and figure 1.3 (b) is the pie-chart showing energy consumption of 2015. Figure adopted from [5]

### 1.3 Building Energy consumption

The energy demand of the building sector accounts for more than 30% of the total energy consumption and more than 55% of the global electricity demand. Efforts to develop sustainable buildings are progressing but are still not keeping pace with the increasing size of the building sector and the rising demand for energy. Energy demand from buildings and building construction will continue to rise, driven by improved access to energy in developing countries, greater ownership and use of energy-consuming devices, and rapid growth in global building floor area, at nearly 3% per year [6].

Thus, to overcome the problem of energy crisis in the future due to the depletion of fossil fuels and protect the environment from global warming due to excessive emissions of CO<sub>2</sub>, transition from fossil fuels to renewable energy as well as energy conservation is equally important. In this paper, we discuss methods that is connected with the energy conservation in buildings using energy pattern analysis using clustering technique.

### 1.4 Effort taken by Chubu University for Energy conservation

Chubu University is a private university of Japan located in the Aichi prefecture. The university consists of seven different departments. After East Japan earthquake most of the power plants were shut down and Japan suddenly faced the problem of energy crisis. Since then the awareness in Japanese government and public as well as private companies and universities raised. In December 2015, Paris Climate Conference (COP21) took place where

195 countries adopted the first-ever universally, legally binding global climate deal.

Japan made a commitment to reduce the domestic emission reduction by 26.0% by the year 2030 in comparison to 2013.

In 2010 the electricity consumption in Chubu university exceeded the demand electricity load, followed by the request from government in 2011 for energy conservation, Chubu university strengthened the energy conservation activities.

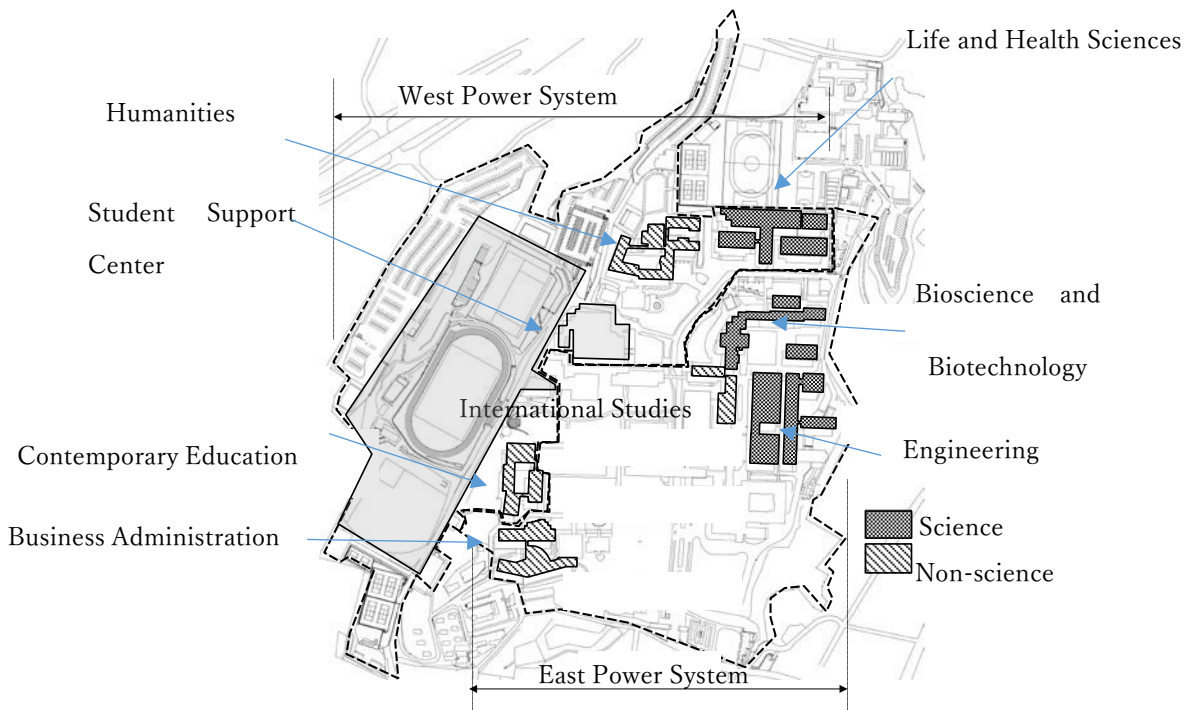


Figure 1.4 Overview of different departments of Chubu University

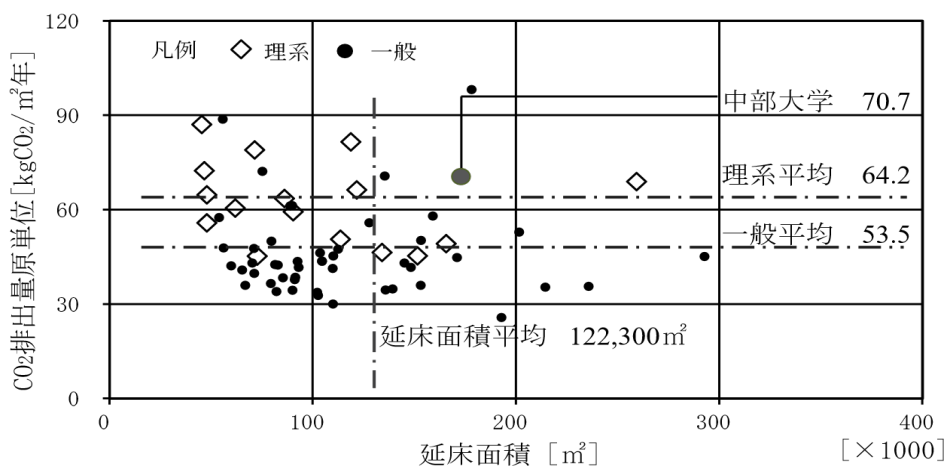
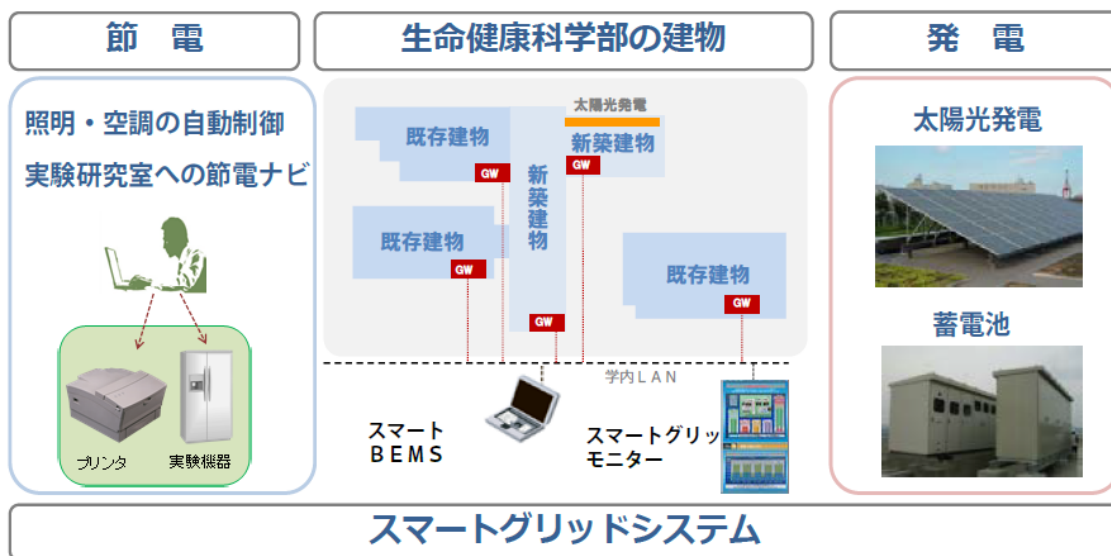


Figure 1.5 Amount of CO<sub>2</sub> emissions from educational facilities (in Tokyo) and the position of Chubu University in the year 2010. Figure adopted from [7]

As shown in figure 1.5, In 2010 the CO<sub>2</sub> emission ratio of Chubu University was 70.7 kgCO<sub>2</sub>/m<sup>2</sup>/year, which is around the energy consumption of Science Universities of Japan.

To reduce the amount of energy consumption, a number of energy conservation measures has been carried out. Chubu University introduced a system called BEMS, which measures the electricity in every building of Chubu University, manages energy consumption in each grid connected to the intra, and deliver power saving mails. One major contributor for the energy conservation is the construction of smart grid as shown in figure 1.6. Micro grid was constructed in the department of Life and Health Sciences. The grid consists of the combination of solar energy plants of capacity 20 kW, cogeneration of capacity 144 kW and storage batteries. The power saving part consists of “automatic control of lighting and air-conditioning” and “power saving navigation”. Power saving navigation is a system used for the energy conservation where mails related to the energy consumption status at the present time and possible energy consumption status of the day based on the energy forecasting is sent to the energy conservation related members of the university and requested for the possible measures of energy consumption reduction. In the year 2012 there was a reduction in electricity consumption by 115 kW which is 24.4% lesser than the last year, also 6.3% lower carbon emission was achieved.



**Figure 1.6** Overview of power saving by smart grid system and power generation. Figure adopted from [7].



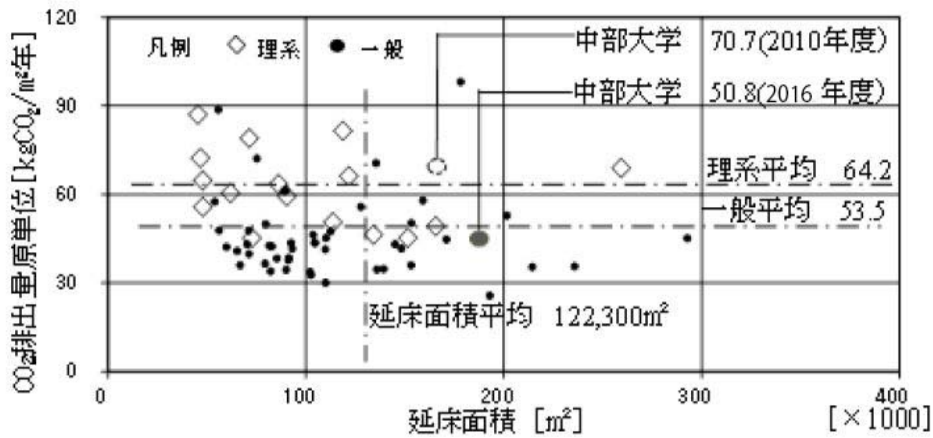
**Figure 1.7** Smart grid monitor displaying the status of energy consumption in Chubu University in the real time. Figure adopted from [7].

The successful result in the College of Life and Health Science led to the expansion of smart grid in the whole University. As shown in the figure 1.8, the University grid can be divided into the East-Campus grid and the West-Campus grid. Smart BEMS is connected to the university's energy metering system enabling centralized management of energy consumption in each grid and delivering of power-saving mails. In the control system as shown in figure 1.9, the energy forecast for faculties and Campus are calculated from weather forecast and past result. The BEMS in the center of the figure calculates the power saving target amount from the difference between the power management value and the predicted value and performs automatic control of the air-conditioning and lighting and discharge battery of the micro-grid and start the use of co-generation. At the same time, the facility users are sent a power-saving email regarding the forecast of the power consumption as well as request for power-saving for experimental research equipment and laboratory equipment that does not rely on automatic control.





saving mail is delivered in two stages, warning message and final alarm at 95% level of the power saving target. The East and West campus grids manage their own consumption forecast and the combined power consumption of each department grid. When it is predicted that the target value will be exceeded, the above-mentioned power-saving mail is sent and automatic control are executed to all the department grids for the power conservation. On the power generation side, the discharge of stored electricity is done at the time of delivery of first alarm. The amount of CO<sub>2</sub> emission was reduced by 965 tons, 14% lower in 2011 comparison to 2016 and the reduction in the peak energy consumption was 11% (528 kW) in 2016 in comparison to 2011.



**Figure 1.10** Amount of CO<sub>2</sub> emissions from educational facilities (in Tokyo) and the position of Chubu University in the year 2016. Figure adopted from [7]

As shown in the figure 1.10, CO<sub>2</sub> emission of 70.7 kgCO<sub>2</sub>/m<sup>2</sup>/year in the 2011 was reduced to 50.8 kgCO<sub>2</sub>/m<sup>2</sup>/year in 2016. The achievement in the CO<sub>2</sub> emission led to the shifting of the position of Chubu University near to the non-science Universities of Japan.

### 1.5 The reason for using clustering method for electricity consumption analysis

Different measures for the reduction of energy consumption in Chubu University is stated in [7]. As shown in the figure 1.11, to classify the electricity consumption of Chubu University using smart BEMS, electricity consumption of whole university as well as the departments were classified into seasons and New Year`s Holiday (i.e. summer, winter, spring, autumn and New Year`s Holiday) representative weeks. From the representative weeks the base energy consumption as well as the peak energy consumption were determined. But this is a lengthy process and same week of each season may not represent the representative days for each building.

In reference [8], it was found that the university calendar was matching with the clustering result of the university. A method that classifies the energy consumption pattern of the building by analyzing all the electricity consumption data of each day of the building is essential and in this case, clustering technique comes to be handy. Thus in this research, we use clustering method to classify the energy consumption pattern of the university.

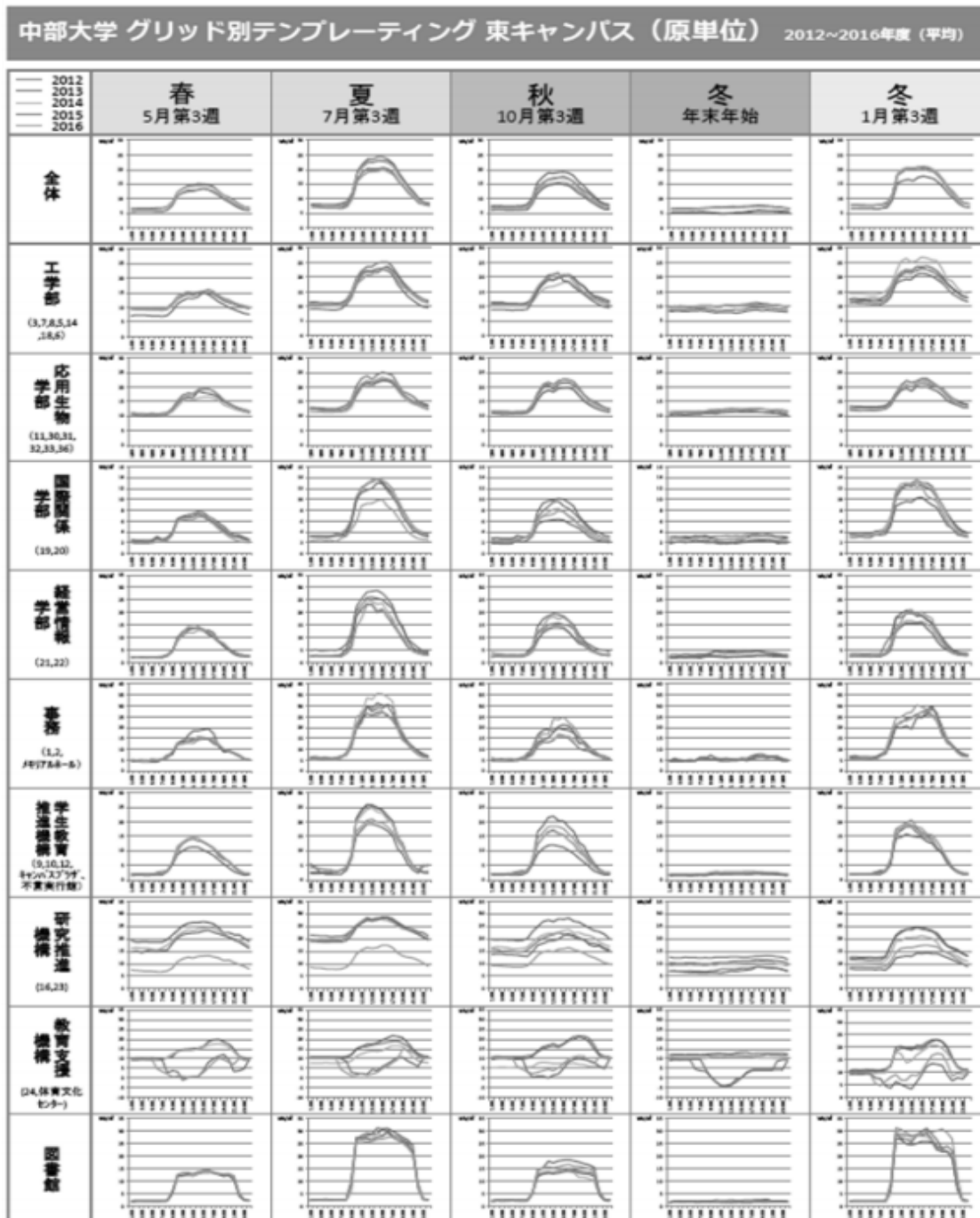


Figure 1.11 Classification of electricity consumption of different departments of Chubu University based on the representative weeks. Figure adopted from [7].

## Chapter 2

### Clustering and the proposed method

#### 2.1 Clustering

Clustering is the process of partitioning data objects into groups, or clusters, so that the objects within a cluster are similar to one another and dissimilar from the objects on other clusters [9]. Cluster analysis is an unsupervised learning method that acts as a cornerstone in intelligent data analysis processes. It is used for the exploration of relationship among a collection of patterns by organizing them into homogeneous clusters [10]. Clustering has variety of applications in different domains, visualization, optimization, finance, manufacturing, and medical organizations [11]. Due to the improvement in the sensors, data loggers, detection, and storing technology, and the remarkable progress in internet searches, digital imaging, and video surveillance, an enormous amount of data are generated on a daily basis. The rapid increase in both the volume and the types of data has necessitated the development of methodologies that can automatically understand, process, and summarize data. To deal with this problem, clustering technique are helpful. Clustering algorithm can be categorized into partitional clustering, hierarchical clustering, density based clustering and grid based clustering [12-14].

##### 2.1.1 Partitional clustering

Given a set of  $n$  objects, a partitioning clustering constructs  $k$  partitions of the data, where each partition represents a cluster and  $k \leq n$ . That is, it divides the data into  $k$  groups such that each group must contain at least one object. In other words, partitioning methods conduct one-level partitioning on data sets. The basic partitioning methods typically adopt exclusive cluster separation. That is, each object must belong to exactly one group. Most partitioning methods are distance-based. Given  $k$ , the number of partitions to construct, a partitioning method creates an initial partitioning. It then uses an iterative relocation technique that attempts to improve the partitioning by moving objects from one group to another. The general criterion of a good partitioning is that objects in the same cluster are “close” or related to each other, whereas objects in different clusters are “far apart” or very different.

There are many types of partitioning algorithms; they are K-means, K-medoids, PAM (Partitioning Around Medoids) and CLARA (Clustering Large Applications).

### **2.1.2 Hierarchical clustering**

A hierarchical clustering creates a hierarchical decomposition of the given set of objects. A hierarchical clustering can be classified as being either agglomerative or divisive, based on how the hierarchical decomposition is formed. The agglomerative approach, also called the bottom-up approach, starts with each object forming a separate group. It successively merges the objects or groups close to one another, until all of the groups are merged into one (the topmost level of the hierarchy), or a termination condition holds. The divisive approach, also called the top-down approach, starts with all the objects in the same cluster. In each successive iteration, a cluster is split up into smaller clusters, until eventually each object is in one cluster, or a termination condition holds.

### **2.1.3 Density Based Clustering**

Density based clustering are devised to create arbitrary shaped clusters. In this approach, a cluster is regarded as a region in which the density of data objects exceeds a threshold [15]. For example, for each data point within a given cluster, the neighborhood of a given radius has to contain at least a minimum number of points. Such a method can be used to filter out noise (outlier) and discover cluster of arbitrary shape.

Density based methods can divide a set of objects into multiple exclusive clusters, or a hierarchy of clusters. Typically, density based methods consider exclusive clusters only, and do not consider fuzzy clusters. Moreover, density-based methods can be extended from full space to subspace clustering.

### **2.1.4 Grid based clustering**

Grid-based algorithm defines a set of grid-cells, it assigns objects to the appropriate grid cell and compute the density of each cell and eliminate cells, whose density is below a defined threshold “ $t$ ”. Form clusters from contiguous (adjacent) groups of dense cells (usually minimizing a given objective function). Grid-based algorithm uses multiresolution grid data structure. Clustering complexity depends on the number of populated grid cells and not on the number of objects in the dataset. Some popular Grid Based clustering methods are: • STING (STatistical INformation Grid approach) • CLIQUE (Clustering In QUEst)[16].

Among various clustering algorithm, K-means developed by MacQueen is most widely used. In this paper, we use K-means algorithm, a famous partitional algorithm because it had the ability to group large amount of data with computational time relatively fast and efficiently [17]. The simplicity of K-means means that the algorithm has been adopted in many fields. It is popular because it is able to quickly and efficiently cluster amount of data, including

outliers. It remains a basic framework for developing numerical or conceptual clustering through various possibilities of distance and prototype choices.

## 2.2 Literature review of clustering

Several attempts have been made to try and solve the cluster centroid initialization problem. Amri et al. used K-means clustering to analyze the electricity use patterns of 370 clients collected from 2011 through 2014. Using a dimension reduction technique, whole data were reduced to four attributes. Each attribute represented the sum of the values of the four seasons: spring, summer, autumn, and winter. Then, the data were classified into five clusters using K-means clustering. As a result, the highest and the lowest energy use in summer and spring were determined [18]. Damayanti et al. used the K-harmonic clustering technique for grouping a one-year electric load profile. One whole year of electrical use data were divided into two clusters. An electrical load profile was generated for both clusters. The first cluster pattern had an irregular load pattern dominated by holidays, whereas the second load profiles described the burden on weekdays due to the higher load demand of offices and industries [19]. Santamouris et al. used the intelligent fuzzy clustering technique to classify the energy performance of school buildings. Five energy clusters for both the total and the heating energy use were calculated. The clustering method was compared with the frequency rating procedure. The fuzzy clustering technique was found to produce more robust classes and classify the buildings according to existing similarities [20]. Arai and Barakbah proposed a hierarchical method to optimize the initial centroids for the K-means algorithm. This algorithm uses the clustering result of the K-means algorithm and then transforms all the centroids of the clustering result by combining with the hierarchical algorithm to determine the initial centroids for K-means. This method is better for complex clustering with large data sets and many attributes. However, this method takes advantage of the K-means algorithm for speed and precision [21]. Yedla et al. proposed a new method for finding better initial centroids and provided an efficient process of assigning data points to suitable clusters with reduced time complexity. In this method, the initial centroid is chosen by calculating the distance of each data point from the origin, then the data points are sorted and divided into k equal parts. Then, the middle point of each data point is taken as the initial centroid. This method was found to produce better initial centroids and provides an efficient method of assigning the data points to the suitable cluster [22]. Shakti and Thanamani proposed a method Kernel Principal Component Analysis (KPCA) for reducing the time complexity and improving the accuracy of K-means algorithm. For the dimension reduction technique, principle component analysis (PCA) is used and KPCA is an extended form of PCA. The time complexity of the K-means algorithm is high if a large dataset is used for clustering. Thus, before applying the K-means algorithm, the dimensions of the data are reduced using KPCA.

This method is able to overcome the problem of the K-means algorithm for high-dimensional data [23]. Huang et al. proposed an automatic variable weighing method in the K-means algorithm that can automatically estimate the variable weights. The variable weights created by this approach estimate the significance of the variables in clustering and can be deployed in variable selection in various data mining applications where large and complex real data are often used. Identification and removal of insignificant variables according to the weight values was found to improve the clustering result [24]. Fahim et al. proposed a method for enhancing the performance of the K-means algorithm. In the original K-means algorithm, the distance between each point and cluster centroid is calculated in each iteration so the computational time is high. In this method, from the second iteration, distance is measured for only those points whose distance increased, decreasing computation time. However, the initial centroids are selected randomly so this method is sensitive to the initial centroid and also does not produce a unique result [25]. Prahastono et al. explained various clustering techniques (e.g., hierarchical, K-means, fuzzy K-means, follow the leader, and fuzzy relation) and their characteristics for the classification of customers and the generation of electric load profiles. Each clustering algorithm has its own peculiarities, so selection of the clustering algorithm depending on the data is essential [26]. Moliana-Solana et al. reviewed developments in information technologies and their influence on building energy management. Various aspects of data science for building energy management and techniques, like smart metering, the Internet of Things, and cloud computing that have been applied or could be applied to solve the energy problem, were discussed [27].

The literature summarized above proposed new methodologies for the selection of better initial centroids; however, regarding the selection of initial centroids for time series energy use data, little research has been conducted. The methodology used in this research for the determination of better initial centroids for building electricity use data has two major benefits: (1) the uncertainty of the K-means algorithm is removed because the results produced by the proposed method are the same irrespective of the number of trials and (2) the accuracy of proposed method is better than that of the K-means random initialization method.

### **2.3 Outlier detection and imputation of missing values**

An outlier is an observation that deviates from other observations to arouse suspicion that it was generated by a different mechanism; it may also be a noisy observation that does not fit the assumed model of the data [28]. Outliers commonly occur in building electricity use measurements. The presence of an outlier in clustering data may lead to an inappropriate result. If there are missing values in the clustering data, then clustering becomes impossible. However, simply dropping the fixed number of values as outliers may inadvertently cause the

loss of important observations. Thus, in this research, to detect the outliers, the data were normalized to  $[0,1]$  using a min-max normalization technique. Outliers present in the data used in this research contain some extreme values thus, Tukey's method, which uses quartiles for the detection of outliers, was preferred in this research. Normalization of the data using the min-max normalization technique was performed as shown in Equation (1):

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} (x'_{max} - x'_{min}) + x'_{min} \quad (1)$$

where  $x_{max}$  and  $x_{min}$  are the original maximum and minimum values of  $x$ , respectively. By using min-max normalization, the original values ( $x$ ) are transformed to  $x'$  in the range  $[x'_{min}, x'_{max}] = [0,1]$ .

Tukey's method [29] is less sensitive to extreme values because it uses quartiles that are resistant to extreme values. The interquartile range (IQR) is the distance between the lower (Q1) and upper (Q3) quartiles. The inner fence =  $[Q1 - 1.5 IQR, Q3 + 1.5 IQR]$ , and the outer fence =  $[Q1 - 3 IQR, Q3 + 3 IQR]$ . The values that lie beyond the inner and outer fences are considered possible outliers, whereas the extreme values that lie beyond the outer fence are considered probable outliers. The data used in this research contained some extreme values compared to remaining data; as a result, not all were detected by inner fence so the outer fence was used for outlier detection. The outliers beyond the outer fence were converted into not available (NA) values. Then, the NA values were imputed with appropriate values using linear interpolation using the zoo package of R programming language.

Figure 2.1 depicts a boxplot with outliers. Due to the presence of extreme values, the box plot is in the form of a line. Figure 2.2 depicts the boxplot after outlier removal using Tukey's method. Some values are seen above the inner fence, which are not outliers but actual values that occur in electricity use data during summer and winter air-conditioning peak days.

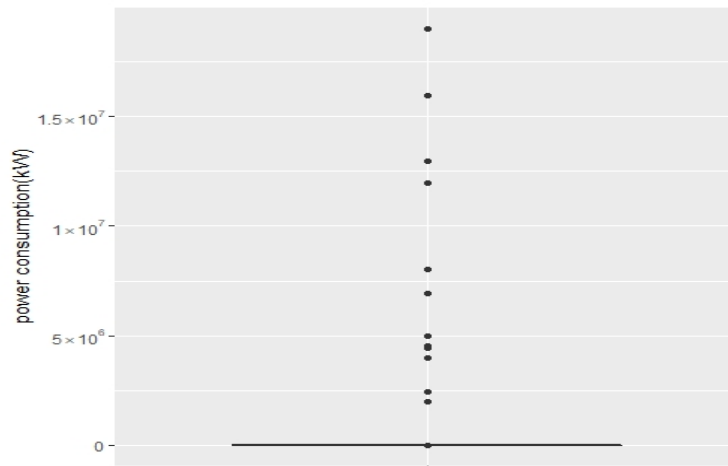
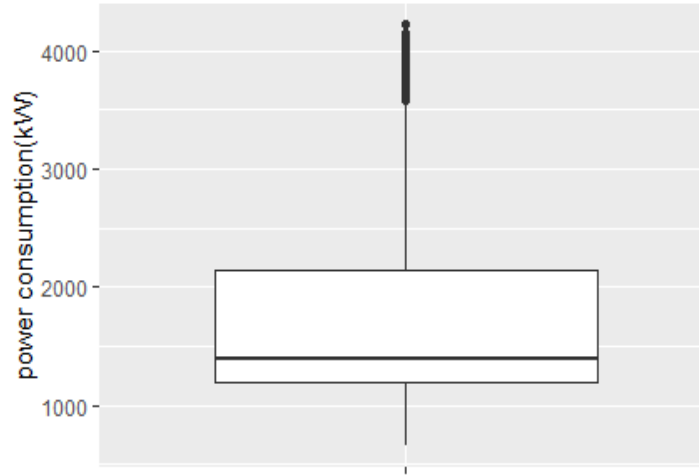


Figure 2.1 Boxplot with outliers.





**Figure 2.2** Boxplot after outliers were removed.

#### 2.4 Overview of K-Means Clustering Algorithm

K-means uses Euclidean distance formula to find the correlation between two objects:

$$\text{dist}(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2)$$

where  $x_i$  and  $y_i$  are the attributes of a given object, and  $i$  varies from 1 to  $n$ .

Initial centroids are determined randomly using K-means clustering. The steps taken are as follows:

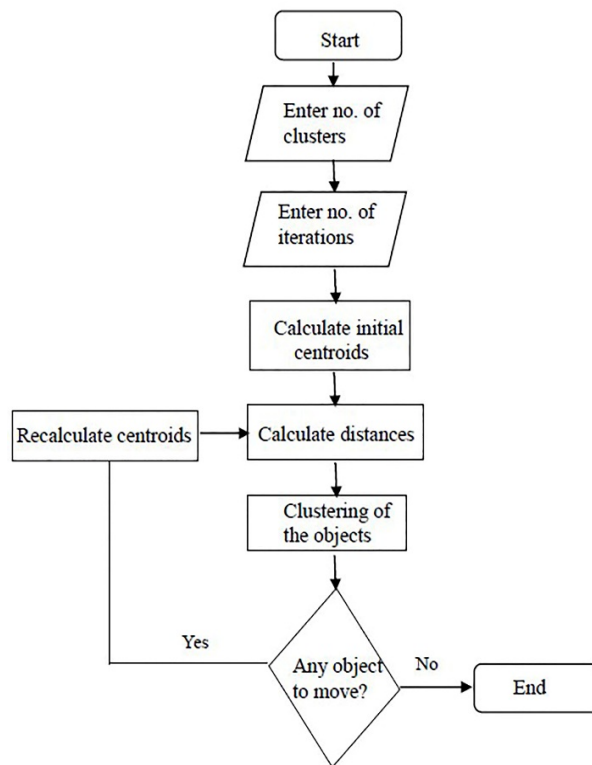
- i) Determine the desired value of  $k$ , where the value of  $k$  is the number of the desired clusters.
- ii) Determine the initial centroids. The initial centroid is assigned randomly from the existing data, and the number of clusters is equal to the number of initial centroids.
- iii) Find the nearest centroid of each data point by calculating the distance to each centroid using the Euclidean distance formula.
- iv) Group the data by minimum distance. A data point will be part of a cluster if it has the closest distance from its cluster center.
- v) Find new centroids based on the average of the data for each cluster.
- vi) Return to step (iii).
- vii) Stop if there are no data changes in the cluster assignment.

The data used in this research were a time series of building electricity data, and the shape of the input vectors had features that were arranged by time. Time series data are commonly used for forecasting. Ricardo et al. [30] and Michelangelo et al. [31] used time series data for solar power forecasting and renewable energy forecasting respectively.

Time series clustering is divided into two groups:

- a) Feature-based or model-based
- b) Raw data-based

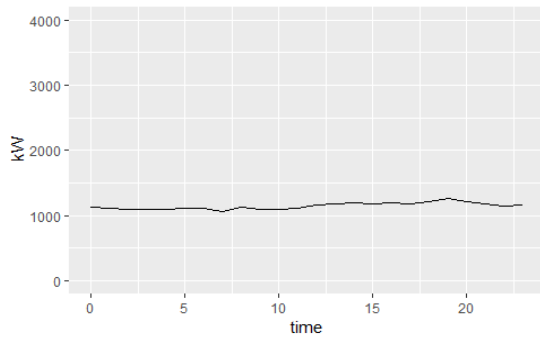
In the case of feature-based clustering, raw data are summarized or transformed by means of feature extraction or parametric models, e.g., dynamic regression, autoregressive integrated moving average (ARIMA), and neural networks, to make the clustering work more easily. On the other hand, raw data-based clustering is directly applied over time series vectors without any space-transformation prior to the clustering phase [32]. In the method reported here, we deal with raw data-based clustering. Figure 2.3 is a flow chart indicating the process of a K-means algorithm.



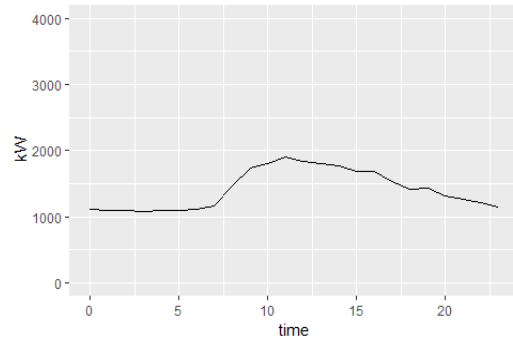
**Figure 2.3** Flow chart of the K-means algorithm

## 2.5 Proposed method

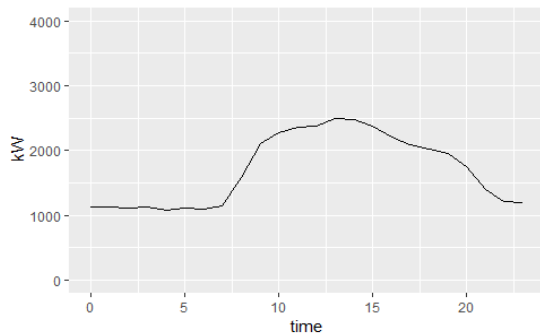
K-means clustering selects initial centroids randomly, so the result of the clustering algorithm changes every time. It is desirable to have a clustering algorithm that produces a unique result irrespective of the clustering run time. We propose a method for enhancing the performance of the K-means algorithm by selecting better initial centroids. This method produces a unique clustering result, and the clustering accuracy is better than the K-means random initialization method.



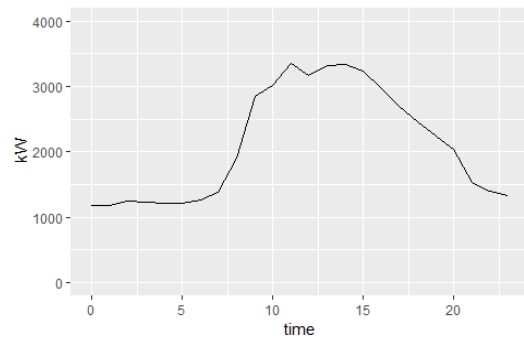
(a) Sunday and holidays



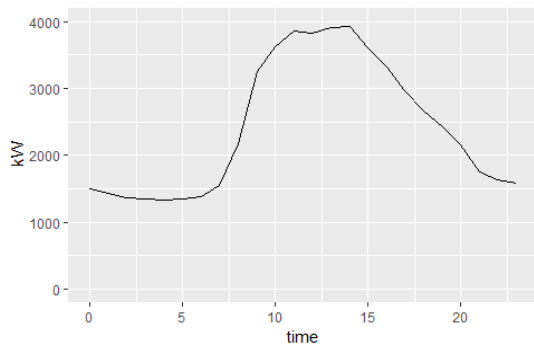
(b) Saturday



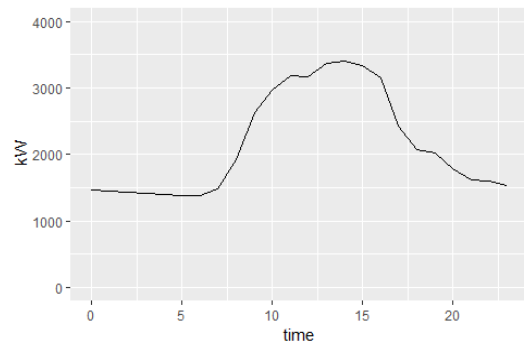
(c) Lecture day without air conditioning



(d) Lecture day with air-conditioning



(e) Lecture day with air-conditioning peak



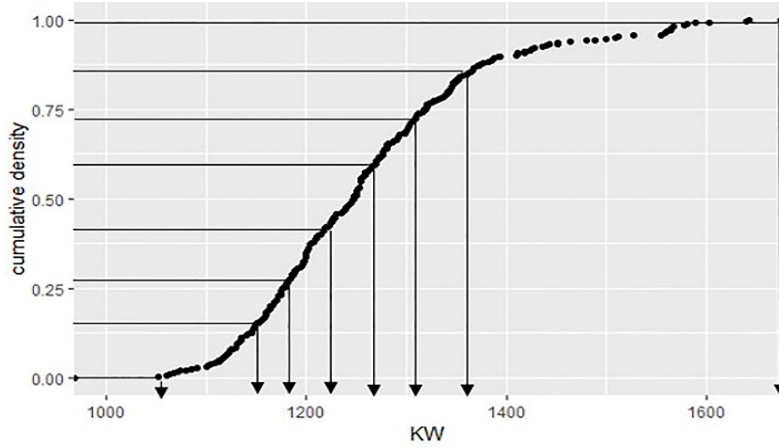
(f) Non-lecture days with air-conditioning

**Figure 2.4** Electricity use in Chubu University, with (a)~(f) representing different electricity use periods.

To show the variation in the electricity use pattern of Chubu University, six sets of graphs in different electricity use period were plotted as shown in figure 2.4. Except for Sundays and holidays, the patterns of electricity use in Chubu University were almost similar. Electricity use increases sharply around 8 AM and reaches to its peak in the noon around 12PM-1PM. In the evening and early morning, electricity use was low, representing the base electricity

consumption of the university. A difference was mainly observed in terms of the hourly electricity use. Thus, in this research, we decided to select the initial centroids based on the hourly distribution of one-year electricity use.

The initial cluster centroids of the K-means algorithm were determined by using the percentile method based on the empirical cumulative distribution function. In the case of  $k$  clusters, cumulative density was divided into  $(k+2)$  values according to the percentiles. Then, the value of electricity consumption data corresponding to the percentiles in empirical cumulative distribution was chosen. The minimum and maximum initial values were neglected in order to reduce the effect of maximum and minimum values and possible outliers, and also to prevent an empty cluster. Thus, the final  $k$  values out of  $k+2$  values were taken as the initial centroids for K-means.



**Figure. 2.5** Selection of the initial centroid by using the percentile method

Figure 2.5 shows the determination of the initial centroids in an empirical cumulative distribution using the percentile method. In Figure 2.5, the distribution of data at 23:00 is shown. Other initial centroids for remaining time stamps were also calculated in similar ways. For  $k$  clusters,  $(k+2)$  equally separated percentile values were chosen, and  $(k+2)$  values from the empirical cumulative distribution corresponding to the percentile values were obtained.

The requirements of the approach are as follows:

$Dat = \{d_1, d_2, d_3, \dots, d_b, \dots, d_n\}$ , a set of  $n$  data points.

$d_i = \{x_1, x_2, x_3, \dots, x_b, \dots, x_m\}$ , a set of attributes of one data point.

$F = \{f_1, f_2, \dots, f_b, \dots, f_k\}$ , a set of  $k$  centroids

The following steps are required to ensure a set of  $k$  desired clusters:

- 1: For  $k$  clusters, divide the cumulative probability equally into  $(k+2)$  parts using the percentile method.

- 2: Find  $(k+2)$  values corresponding to the percentile in an empirical cumulative distribution.
- 3: Compute the distance between each data point  $d_i$  ( $1 \leq i \leq n$ ) to all the initial centroids.
- 4: Find the data points  $d_i$  closest to the centroid  $f_l$  and assign the data  $d_i$  to cluster  $l$ .
- 5: For each cluster  $l$ , recalculate the new centroids.
- 6: Compute the distance between the new centroids and each data  $d_i$ .
- 7: Continue this process until the data points change in the cluster assignment.

The “kmeans” function built in the R programming language is used for K-means algorithm. Libraries of R used include dplyr, plyr, readr, reshape2 and ggplot2.

## 2.6 Experiments and Results

### 2.6.1 Accuracy verification

#### 2.6.1.1 Accuracy verification using Real World datasets

To find the accuracy of the proposed method, an experiment was conducted on four different Real World datasets: namely Iris data, Wine data, Ruspini data and New Thyroid data. Iris data, Wine data and New-Thyroid datasets were downloaded from the UCI Machine Learning Repository [33]. Ruspini data is included in the R package “cluster”. These datasets are typical tests for many classification techniques.

We represent the accuracy percentage ( $r$ ) as a performance measure of the experiment. This is calculated as:

$$r = 1 - e$$

where  $e$  is the clustering error and is defined as [7]

$$e = \frac{\text{Number of misclassified pattern}}{\text{Total no.of patterns}} \times 100\% \quad (3)$$

**Table 2.1** Resulting accuracy of different dataset

Name of dataset	No. of cluster	K-means (100 trials)			Proposed method Accuracy (%)
		Max(%)	Min(%)	Average(%)	
Iris	3	89.3	58	80.9	89.3
Wine	3	70.39	56	65.8	70.39
Ruspini	4	94	79	89.2	100
New Thyroid	3	86.04	79.1	82.13	86

The maximum, minimum and average accuracy of 100 trials obtained by using K-means

and the proposed method are shown in table 2.1. It can be seen that the accuracy of the proposed method is noticeably higher than the average accuracy for 100 trials using the K-means algorithm with random initialization.

The within-cluster sum of the square distance of the proposed method as shown in table 2.2 was smaller, whereas between-cluster sum of the square distance of the proposed method was greater than the K-means with random initialization method for all four datasets. This shows that the proposed algorithm produces better clustering than the K-means random initialization method.

**Table 2. 2** Comparison of K-means algorithms with different datasets

Name of dataset	No. of clusters	K-means (100 trials)		Proposed method	
		Average		Between-cluster sum of square dist.	Within-cluster sum of square dist.
Iris	3	585.8	95.46	602.5	78.85
Wine	3	15146222	2445691	15221607	2370690
Ruspini	4	218108	26226	231493	12881
New Thyroid	3	34679.2	29009.8	35204.8	28876.7

### 2.6.2 Cluster quality comparison using university data

- **Description of university data**

The proposed method was used to analyze the building electricity time series data. The data used in this research were the electric use data from Chubu University. The electricity consumption in each building of the university was measured using the BEMS (Building Energy Management System), and the data were collected by the BEMS server every minute. These data were summed to create hourly electricity use data; thus, 1 year of data from each building consists of 8,760 data points. The electricity use data represent the entire electricity use data of the buildings of the university

The raw data were arranged in the order of days starting from April 1, 2015, until March 31, 2016. Each day consists of 24 hours of data from 0:00 to 23:00. The data in this form were not suitable for clustering, so it was necessary to convert the data into a  $366 \times 24$  order matrix as shown in figure 2.6. Figure 2.7 represents the part of the data frame of the  $366 \times 24$  order matrix used in this research.

$$\begin{bmatrix} x_{1,1} & x_{1,2} & x_{1,3} & \dots & x_{1,24} \\ x_{2,1} & x_{2,2} & x_{2,3} & \dots & x_{2,24} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ x_{366,1} & \dots & \dots & \dots & x_{366,24} \end{bmatrix}$$

	00:00	01:00	02:00	03:00	04:00	05:00
2015/04/01	1078	1089	1055	1083	1065	1084
2015/04/02	1097	1086	1081	1063	1049	1044
2015/04/03	1051	1024	1036	1035	1026	1064
2015/04/04	1119	1122	1108	1109	1111	1109
2015/04/05	1112	1098	1105	1095	1104	1114
2015/04/06	1106	1105	1105	1094	1098	1114
2015/04/07	1189	1172	1171	1174	1180	1178
2015/04/08	1129	1122	1123	1114	1120	1138

**Figure 2.6** Model of  $366 \times 24$  order matrix **Figure 2.7** Part of the data frame of the  $366 \times 24$  order matrix

● **Cluster quality comparison using proposed method and K-means**

To determine the quality of the clusters, we determined the total within-cluster and between-cluster sum of the squared distances for the K-means algorithm and the proposed method. Small within-cluster distances and a large between-cluster sum of squared distance are preferred for better clustering. The values obtained from K-means with random initialization, as shown in table 2.3, were the average of the values obtained from 10 run times of K-means clustering.

**Table 2.3** Comparison of K-means clustering with random initialization and the proposed method

No. of clusters (K)	Within-cluster sum of squared distance		Between-cluster sum of squared distance	
	K-means with random initialization	Proposed method	K-means with random initialization	Proposed Method
4	$2.28 \times 10^8$	$2.08 \times 10^8$	$1.66 \times 10^9$	$1.68 \times 10^9$
5	$1.65 \times 10^8$	$1.65 \times 10^8$	$1.72 \times 10^9$	$1.72 \times 10^9$
6	$1.36 \times 10^8$	$1.30 \times 10^8$	$1.75 \times 10^9$	$1.76 \times 10^9$
7	$1.16 \times 10^8$	$1.09 \times 10^8$	$1.59 \times 10^9$	$1.82 \times 10^9$
<b>Total</b>	<b><math>6.45 \times 10^8</math></b>	<b><math>6.14 \times 10^8</math></b>	<b><math>6.74 \times 10^9</math></b>	<b><math>6.95 \times 10^9</math></b>

The number of trials of the proposed method was only one because it produces the same result. As seen in table 2.3, in each of the cases from cluster 4 to cluster 7, the total within-cluster sum of squared distance of the proposed method was smaller than K-means with random initialization, whereas the sum of squared distance between the cluster centroids of the proposed method was greater than the random initialization method. The proposed method was in agreement with the definition of a good clustering algorithm, i.e., small within-

cluster sum squared distances and large sum squared distances between cluster centroids. Thus, the proposed algorithm was better than the K-means with random initialization method.

## 2.7 Determination of number of cluster

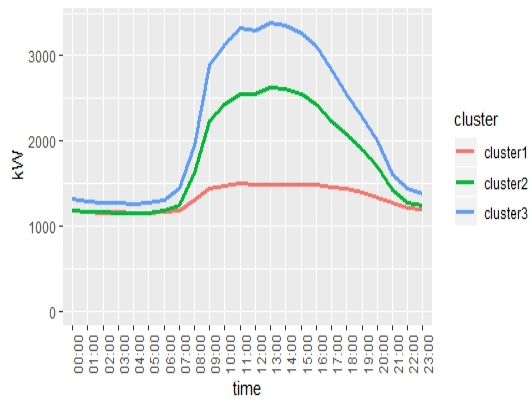
There are many factors that affect the use of electricity in a university. On Sundays and holidays, there are no lectures and no lighting, no air-conditioning or OA equipment are used; thus, the electricity use can be considered to be mainly due to the base energy consumption of the university. Base energy represents the electrical energy of equipment that runs constantly, such as servers, refrigerators, emergency exit lights, research equipment, etc. In Chubu University, Saturday is not a holiday, but the number of students present and the number of lectures and activities in the university are very low compared to normal weekdays; thus, electricity use patterns on Saturday are different to Sundays, holidays and weekdays. The usage of air-conditioning in Chubu University is managed by period. Excluding exceptional condition such as sudden changes in weather condition, usage of air-conditioning in the interim period is not allowed. Air-conditioning on non-lecture days usually occurs in the university vacation period. Air-conditioning in lectures can be divided into two parts: normal air-conditioning on lecture days and lecture days with air-conditioning peaks. In the case of universities in Japan, usually contracts are made with the electric power company, and the contracted value of electricity power demand is fixed for universities. In summer and winter with air-conditioning peaks, if the electricity use exceeds the contracted electricity power demand, the university needs to pay a high amount for their electricity bill; thus, universities authorities make efforts not to exceed the electric consumption beyond the contracted value. Thus, while analyzing the electricity use pattern of universities, air-conditioning peak periods should also be considered.

**Table 2.4** Characteristics of electricity use in Chubu University

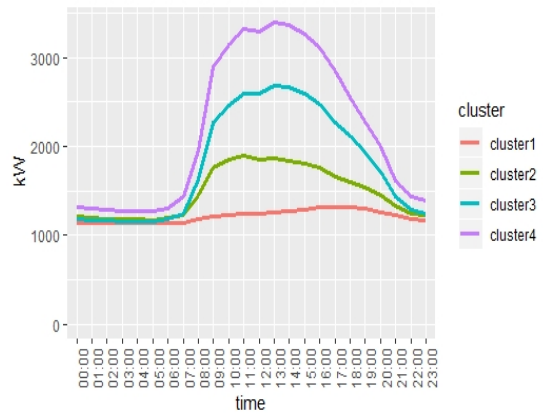
Cluster no.	Lecture	Air-conditioning	University	General occurrences
1	—	—	Closed	Sundays, holidays
2	—	—	Partially open	Saturdays, holidays with events
3	—	✓	Open	Vacations
4	✓	—	Open	Spring and autumn
5	✓	✓	Open	Summer and winter
6	✓	✓	Open	Summer and winter peak periods

\* — represents absence and ✓ represents presence

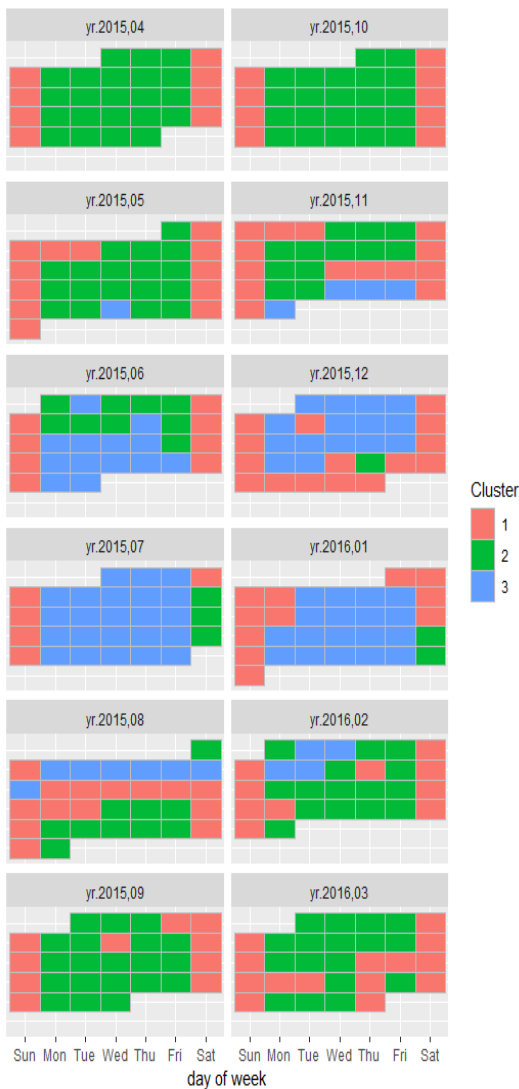




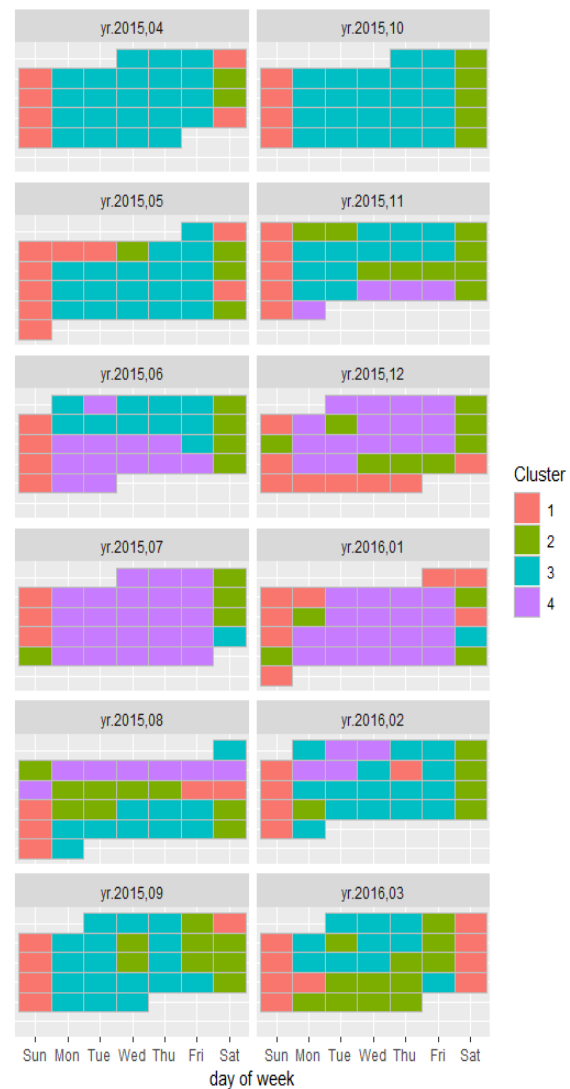
(a) Cluster centers for  $k = 3$



(b) Cluster centers for  $k = 4$

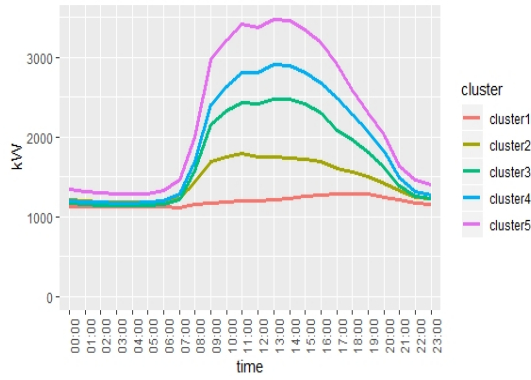


(c) Calendar plot for  $k = 3$

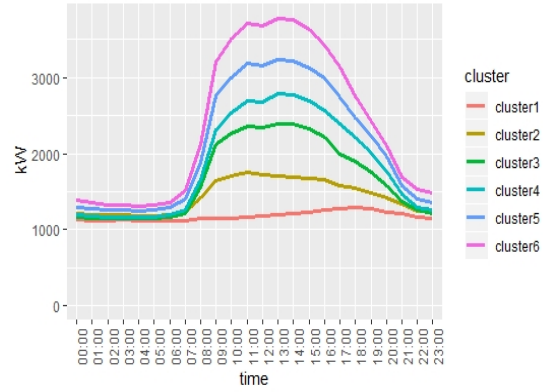


(d) Calendar plot for  $k = 4$

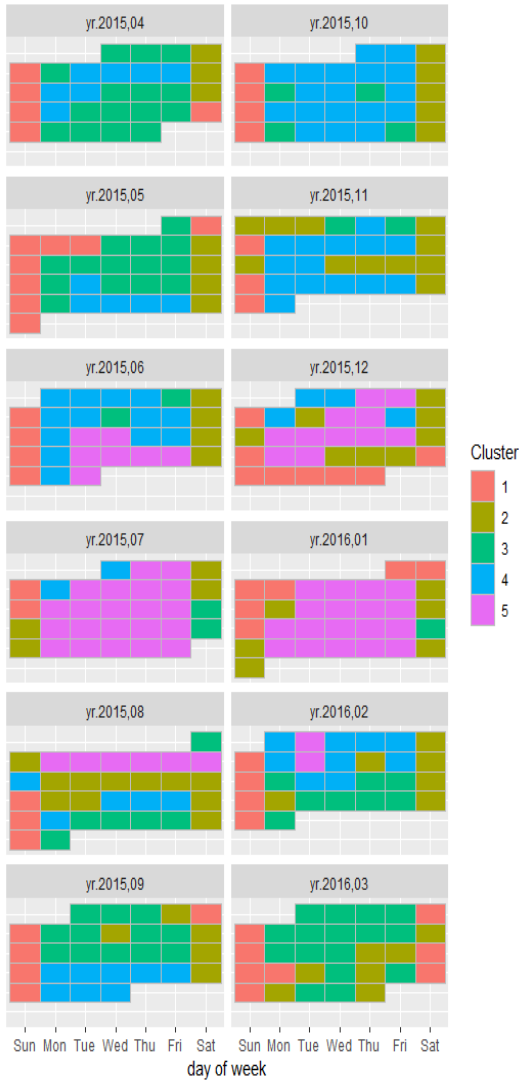
**Figure 2.8** Clustering result of Chubu University. (a) and (b) represent the cluster centers for  $k=3$  and  $4$ , respectively. (c) and (d) represent the calendar plot for  $k=3$  and  $4$ , respectively



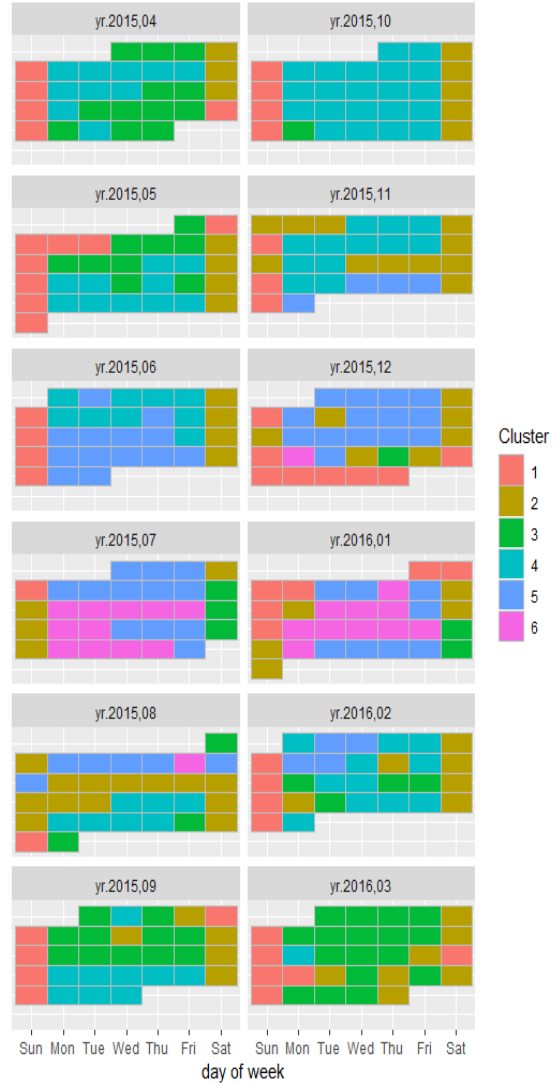
(e) Cluster centers for  $k = 5$



(f) Cluster centers for  $k = 6$



(g) Calendar plot for  $k = 5$



(h) Calendar plot for  $k = 6$

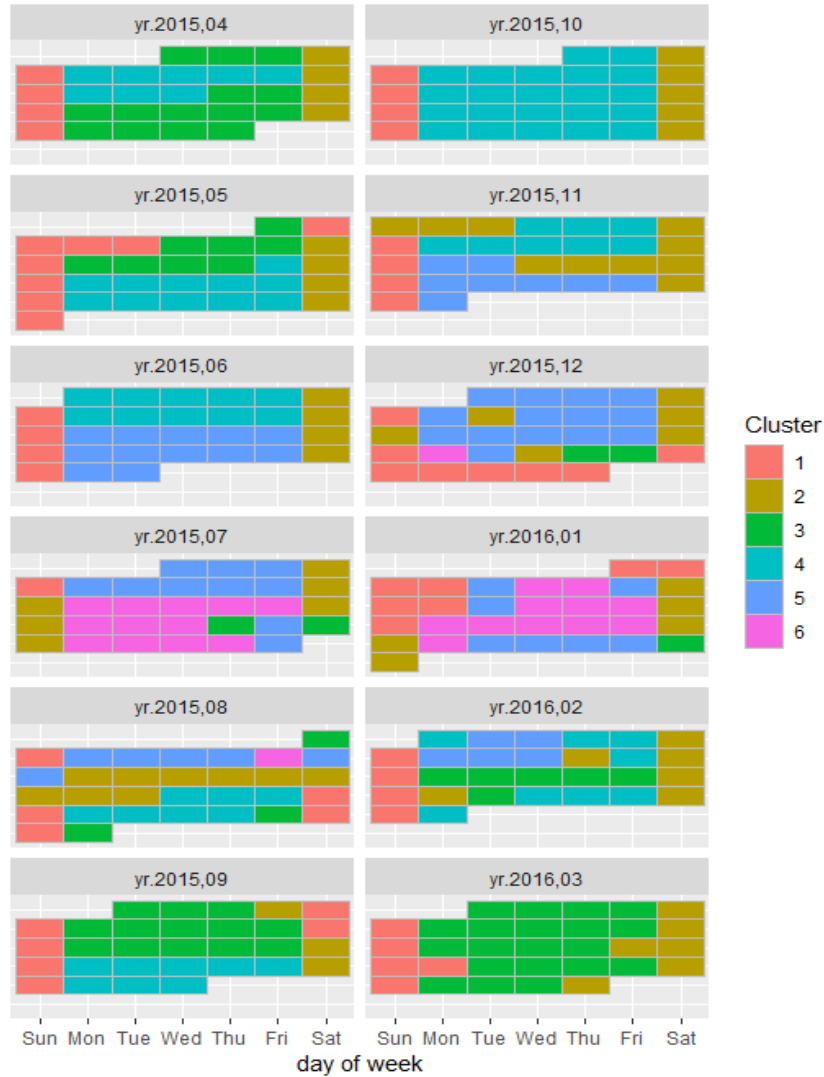
**Figure 2.9** Clustering result of Chubu University. (e) and (f) represent the cluster centers for  $k = 5$  and  $6$  respectively. (g) and (h) represent the calendar plot for  $k = 5$  and  $6$  respectively.

As shown in table 2.4, the electricity use in Chubu University can be categorized into six types depending on the presence or absence of lectures, presence or absence of air-conditioning and presence or absence of holidays.

To analyze the building electricity use patterns, it is necessary to know how much electricity is being used and on which day. So, it is essential to select the proper numbers of clusters ( $k$ ) that properly describe the electricity use patterns. For this purpose, we performed K-means clustering analysis by selecting cluster numbers from 3 to 6 and compared the accuracy with the actual calendar plot created using the university schedule. Figure 2.8 and 2.9 represent the clustering result of Chubu University for 3 to 6 clusters.

In the case of Chubu University in the year 2015 the air-conditioning cooling period was from 15<sup>th</sup> June to 15<sup>th</sup> September, whereas the heating period was from 15<sup>th</sup> November 2015 to 15<sup>th</sup> April 2016. It is possible that, on some of the days, the cluster number stated in table 2.4 changed. For example, most likely, Sunday falls in cluster 1, but occurrences of events such as open campus, university festivals, etc. can change the cluster number on the particular day from cluster 1 to 2, 3 or even higher clusters; thus, in addition to table 2.4, it is necessary to analyze the electricity use of each day to decide the cluster number of each day. Thus, by considering university schedule, table 2.4 and analyzing the electricity use of each day, a calendar plot of Chubu University is created. We consider this calendar plot as the actual calendar plot and use this to compare the accuracy of the clustering result.

On comparing the actual calendar plot of Chubu University with the calendar plot of clustering results for 3 clusters, Sunday and Saturday were not separated; also, air-conditioning and non-air-conditioning periods were not separated. For 4 clusters, the air-conditioning and interim non-air-conditioning periods were not separated. In the case of 5 clusters, the air-conditioning peak started prior to the actual period compared to the university schedule. Thus, clusters 3 to 5 could not represent the actual electricity use patterns in Chubu university. In the case of 6 clusters, air-conditioning and non-air-conditioning interim periods were separated; also, the peak electricity use period was separated which matched with the university schedule.



**Figure 2.10** Actual calendar plot of Chubu University

The accuracy of the clustering result for 6 clusters was determined by comparing the calendar plot produced by the clustering with the university schedule. The accuracy ( $r$ ) of the clustering result was defined as the total no. of accurately classified patterns per total number of patterns; i.e.,

$$r = \frac{A_p}{T_p} \times 100[\%] \quad (4)$$

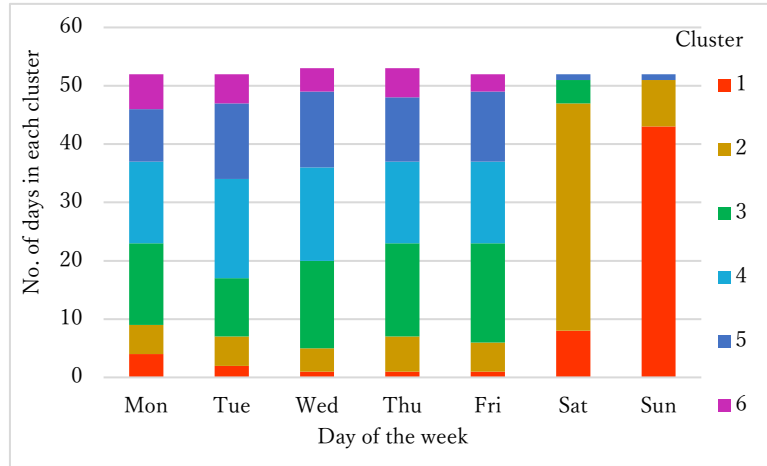
where,

$A_p$  = number of accurately classified patterns

$T_p$  = total no. of patterns

As there were 366 days, we consider the total number of patterns to be 366. Out of the total number of patterns, 327 patterns in the clustering result were found to match with the

university schedule, and so the total number of accurately classified patterns was 327. Using equation 4, the accuracy of the clustering result was 89.34%. Thus, 6 clusters were appropriate for Chubu University. This method can also be used in other universities in Japan which follow a similar electricity use pattern to Chubu University.



**Figure 2.11** The number of days per cluster by day of the week.

Figure 2.11 represents the clustering result of all the Chubu University data with six clusters in the bar graph with x- and y-axes representing the day the of week and number of days in each cluster, respectively. We found that cluster 1 was mainly concentrated on Sundays, Saturdays, and days without lectures. Cluster 2, which represented the electricity use on non-lecture days, was mainly due to electricity consumption of lighting and OA equipment and was mainly concentrated on Saturdays. Air conditioning days were almost always concentrated on weekdays.

## 2.8 Statistical method to find the number of clusters

To find the appropriate number of clusters, the Dunn index was used as an internal validity measure and Average Distance (AD) and Figure of Merit (FOM) as stability measures to find the optimum number of clusters.

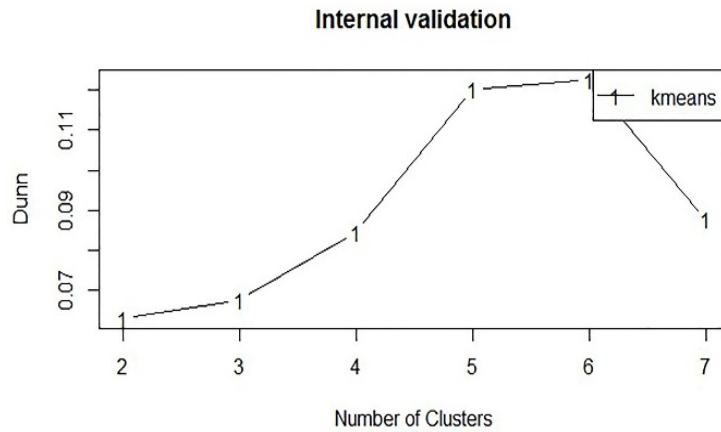
Dunn's validation index ( $D$ ) is an internal validity measure to identify sets of clusters that are compact, with a small variance between members of the cluster, and are well separated, where the means of different clusters are sufficiently far apart, as compared to the means within a cluster. A higher value of  $D$  indicates a better clustering result, and  $D$  can be calculated using Eq. (4): [34]

$$D = \min_{1 \leq i \leq k} \left[ \min_{i+1 \leq j \leq k} \left[ \frac{\text{dist}(c_i, c_j)}{\max_{1 \leq l \leq k} [\text{diam}(c_l)]} \right] \right] \quad (5)$$

where  $\text{dist}(\mathbf{c}_i, \mathbf{c}_j)$  is the distance between cluster  $\mathbf{c}_i$  and  $\mathbf{c}_j$ , and  $\text{dist}(\mathbf{c}_i, \mathbf{c}_j) = \min_{x_i \in \mathbf{c}_i, x_j \in \mathbf{c}_j} \text{dist}(x_i, x_j)$ ,

$\text{dist}(x_i, x_j)$  = is the distance between data points  $x_i \in \mathbf{c}_i, x_j \in \mathbf{c}_j$ ,

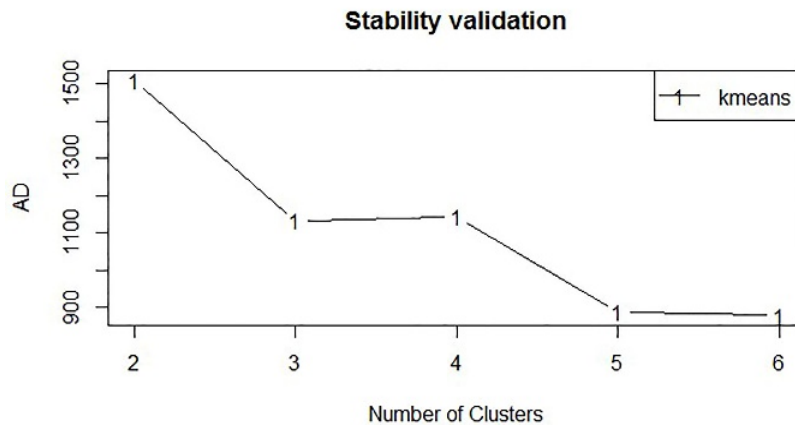
and  $\text{diam}(\mathbf{c}_l)$  is the diameter of cluster  $\mathbf{c}_l$  where  $\text{diam}(\mathbf{c}_l) = \max_{x_{l_1}, x_{l_2} \in \mathbf{c}_l} \text{dist}(x_{l_1}, x_{l_2})$ .



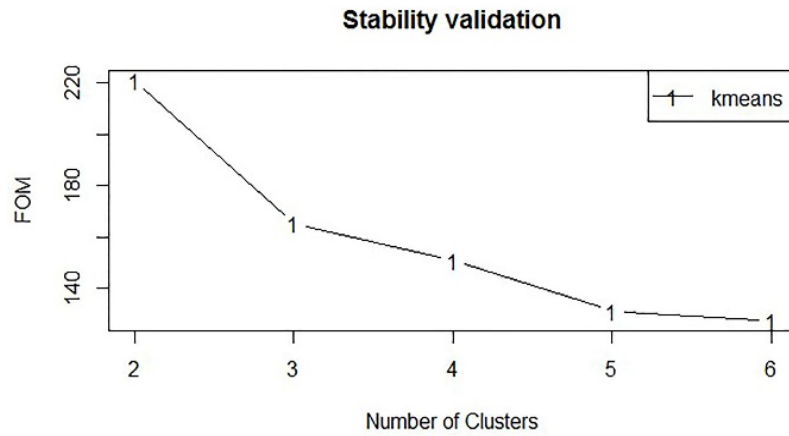
**Fig. 2.12** Internal validation measure using Dunn index

As shown in figure 2.12, when the Dunn index is plotted against the number of clusters, the maximum value of the Dunn index is obtained when the number of clusters is six.

AD is the average distance between observations placed in the same cluster based on the full data and when the clustering is based on the data with a single column removed. AD has a value between zero and infinity, and smaller values are preferred.



**Figure 2.13** Stability validation measure using AD



**Figure 2.14** Stability validation measure using FOM

FOM measures the average intra-cluster variance of the observations in the deleted column, where the clustering is based on the remaining samples. This estimates the mean error using predictions based on the cluster averages. FOM values range between zero and infinity, and smaller values indicate a better performance [35].

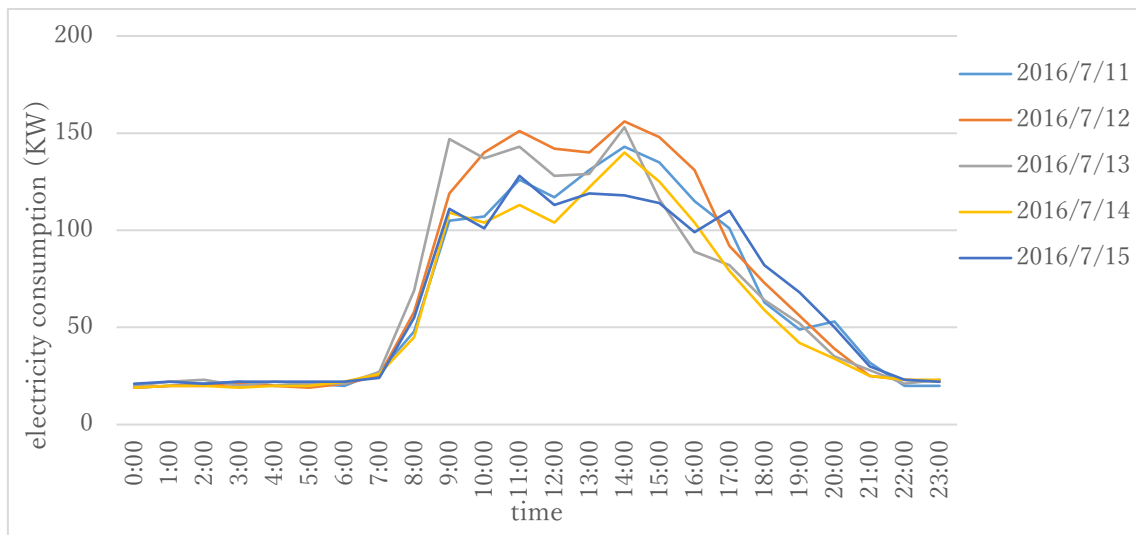
From figures 2.13 and 2.14, the stability validation measures AD and FOM are found to be smallest when the number of clusters is six, so a cluster size of six is appropriate for the data used in this research.

# Chapter 3

## Energy pattern analysis using proposed K-means clustering

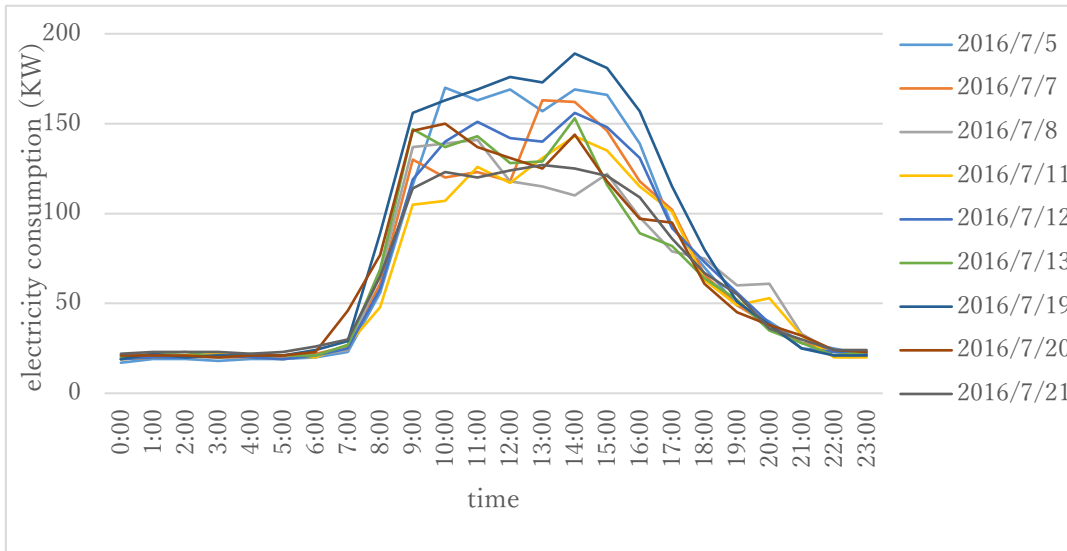
### 3.1 Estimating energy consumption by application

In reference [7], the representative weeks of spring, autumn, winter, year-end and new-year are obtained from actual measurements, and the average values are overlaid to confirm changes in power consumption every year. The representative week is the week that includes the third week of May, July, January and the year-end and new-year holidays. Base and peak energy consumption are extracted and the necessary countermeasures are examined. Figure 3.1 shows an example in the summer, but irregular peak power consumption can be observed. Here the representative week is determined, but the characteristics period may change due to operational changes.



**Figure 3.1** Electricity consumption pattern of peak period by conventional method (superposition of electric consumption for representative week)





**Figure 3.2** electricity consumption pattern during peak summer determined by clustering

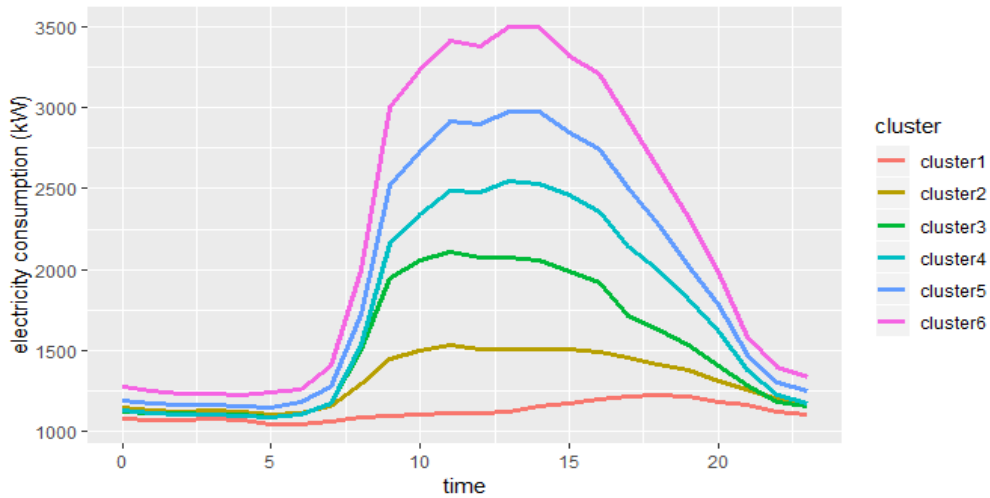
Figure 3.2 shows the result of classification using the K-means clustering method. In this method, the classification is based on power consumption, and figure 3.2 corresponds to the peak period in summer. Since the period is not limited to one week, more days are selected than in figure 3.1, but in terms of peak period extraction, it can be used for analysis in the same way as in figure 3.1.

### 3.1.1 Data used for analysis

The data is obtained by Chubu University Smart BEMS and is collected every minutes at the measurement point of each cubicle of each building. Data is collected on the server, and daily files can be downloaded. The analysis was done using the data of the year 2015. Table 3.1 shows the correspondence between departments and building.

### 3.2 Clustering result

K-means clustering on the electrical consumption data of whole university was conducted. The center of gravity of each cluster is obtained, which is shown in figure 3.3.



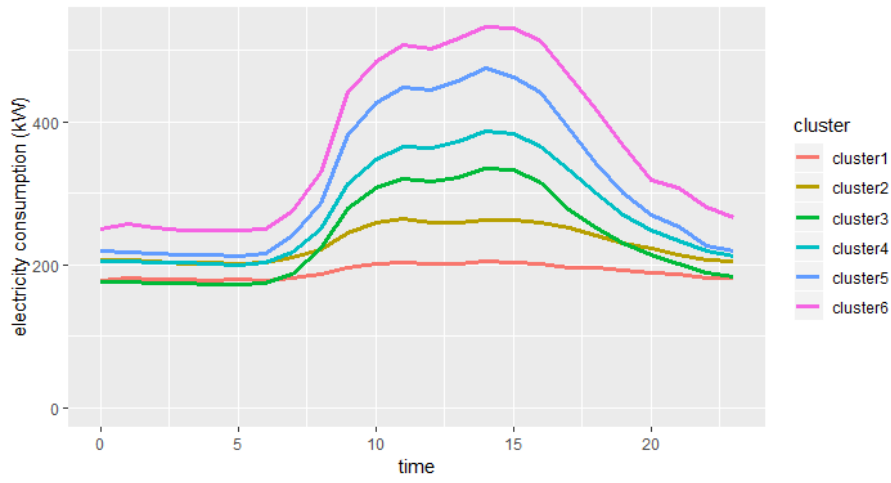
**Figure 3.3** Clustering result: whole university

The electrical consumption data is classified into six clusters on the basis of size of hourly electricity consumption. Shows 3.4 to 3.11 shows the clustering result at each department. Similar to the whole university, the clusters are arranged in the order of low to high energy consumption from bottom to top. The energy consumption in the night time from 10 PM to

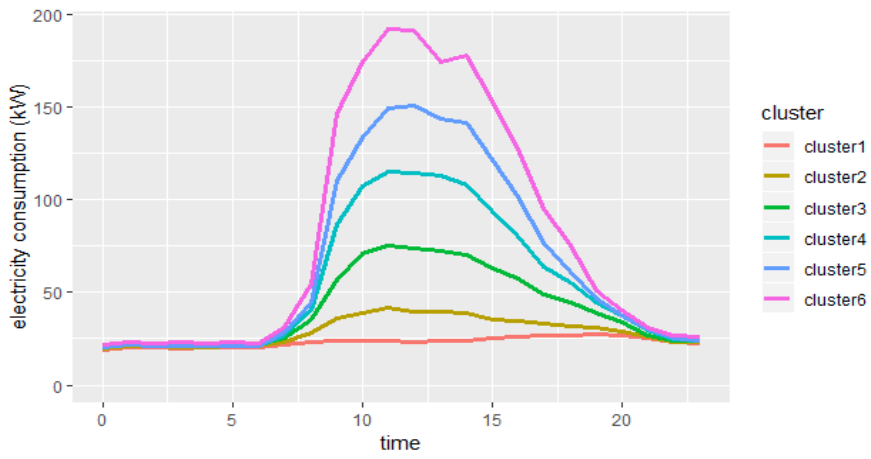
Table 3.1 Correspondence between departments and buildings

departments	Buildings number
Engineering	3 · 7 · 8, 5, 6
Business Administration	21, 22
International Studies	19, 20
Humanities	25, 26, 27, 28
Boiscience and Biotechnology	30, 31, 32 33 · 36, 11, 17
Life and Health Sciences	50 · 51 , 52, 55
Contemporary Education	70 · 71 72
Offices	1 · 2 , Memorial Hall

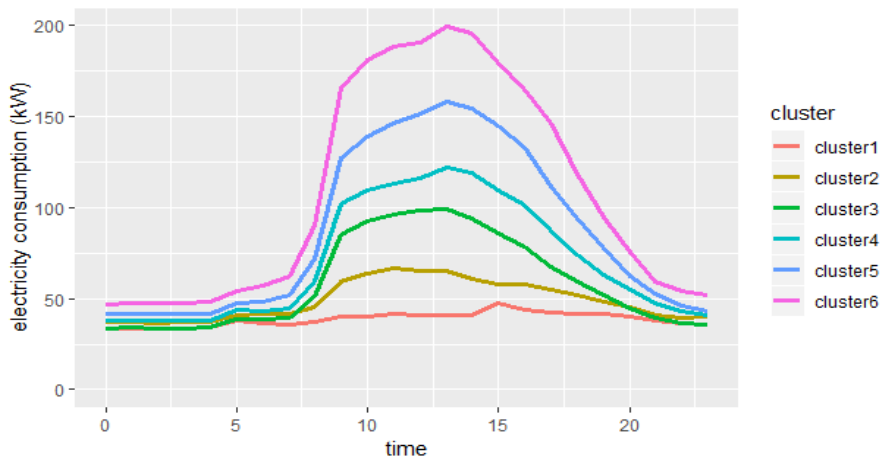
8 AM is due to the energy consumption of servers, refrigerators etc. In figure 3.4, the base energy portion of the engineering department is found different. The difference between base energy consumption of cluster 1 and cluster 6 is also found high. This effect can be considered as the influence of usage of air-conditioner at night and research facilities.



**Figure 3.4** Clustering result: Engineering



**Figure 3.5** Clustering result: Business administration



**Figure 3.6** Clustering result: International Studies

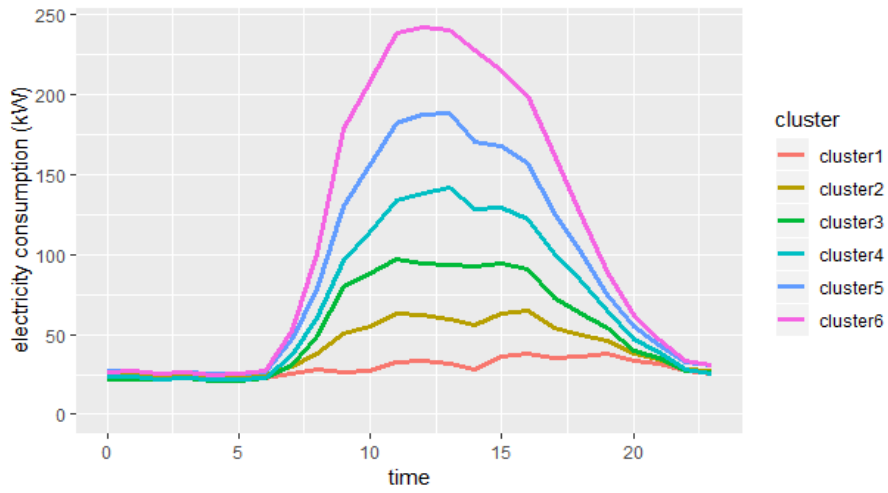


Figure 3.7 Clustering result: Humanities

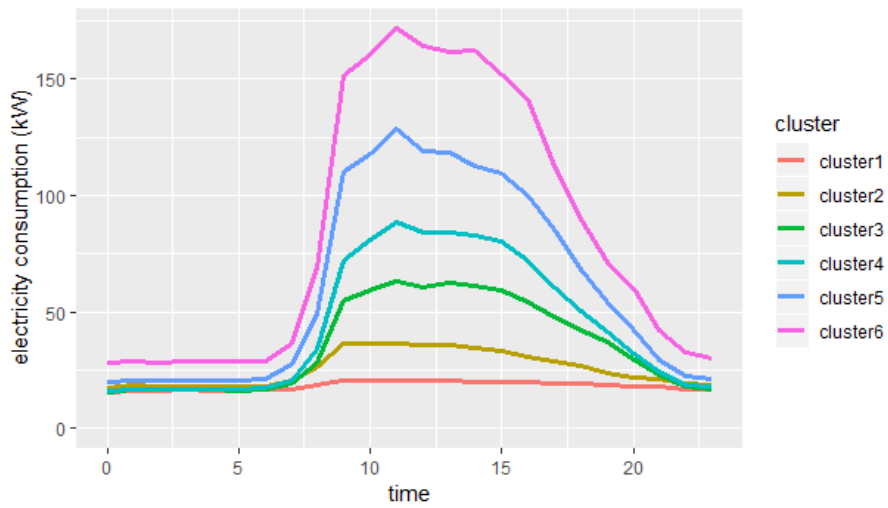


Figure 3.8 Clustering result: Contemporary Education

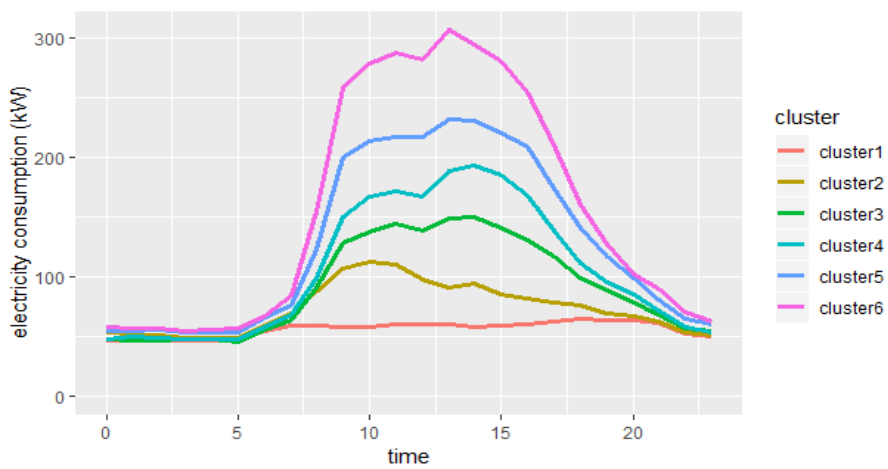
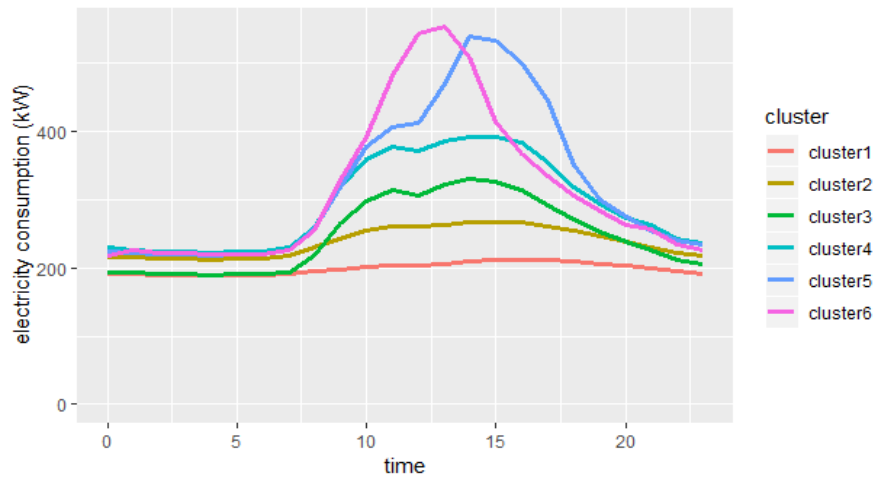
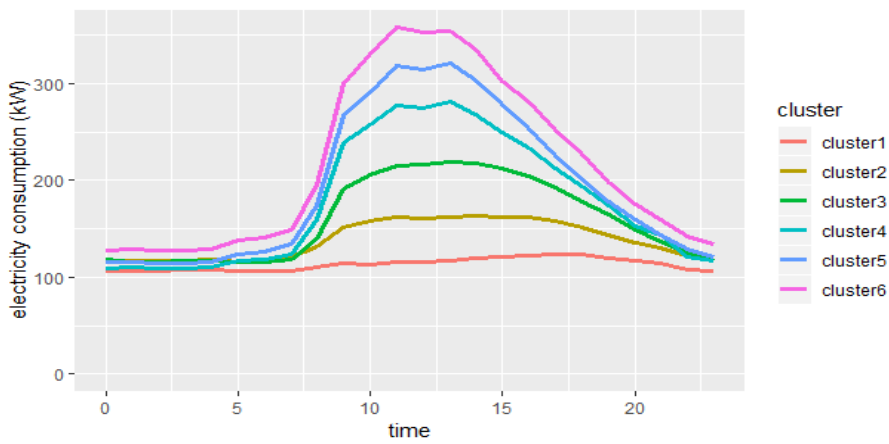


Figure 3.9 Clustering result: Offices



**Figure 3.10** Clustering result: Bioscience and Biotechnology



**Figure 3.11** Clustering result: Life and Health Sciences

### 3.3 Application of Energy Management

#### 3.3.1 Estimating energy consumption by factor

In Chubu university the air-conditioning heating and cooling is controlled by the period. In 2015, the cooling period was from June 15 to September 15, and the heating period was from November 16, 2015 to April 15, 2016. However, it may be operated beyond this period depending on application in the laboratory and weather conditions in the lecture room. Taking these into consideration, the relation between each cluster and university schedule is shown in table 3.2.



**Figure 3.12** Whole university, relation between cluster and schedule

Table 3.2 shows the correspondence between the cluster and the university schedule. Table 3.3 shows is the estimation of energy consumption by factor for each department. Most of the air-conditioning is a multi-package type for buildings, which is both electricity and gas usage type. The fifth column of table 3.3 shows the distinction between electric and gas air-conditioning. Depending on the university schedule, the presence or absence of air-conditioning is determined from the air-conditioning period. But, the clustering data is electricity consumption data and no gas data is used. The energy consumption by factor is estimated from the classification of clusters and energy usage. From the relationship with the university schedule, cluster 1 is mainly on the holidays, when there is very less human activities and air-conditioning. So, cluster 1 can be considered as the base energy consumption of the university. Cluster 2 and 3 usually falls on the days without the use of air-conditioning so, subtracting cluster 2 and 3 from 1 gives the human activity energy consumption. The air-conditioning factor is obtained by subtracting the energy consumption of cluster 4~6 with cluster 3 and adding the primary energy consumption of gas used for air-conditioning.

Figure 3.13 shows the estimated primary energy consumption ratio of different departments and office buildings of Chubu University. The ratio of base energy is found largest in the faculty of engineering. Office buildings with few experimental facilities, humanities department, Business information department, Department of Contemporary education etc. have a small ratio of base. High ratio of base energy in the entire university is due to the influence of refrigerators, servers, experimental facilities etc. in the buildings of science departments.

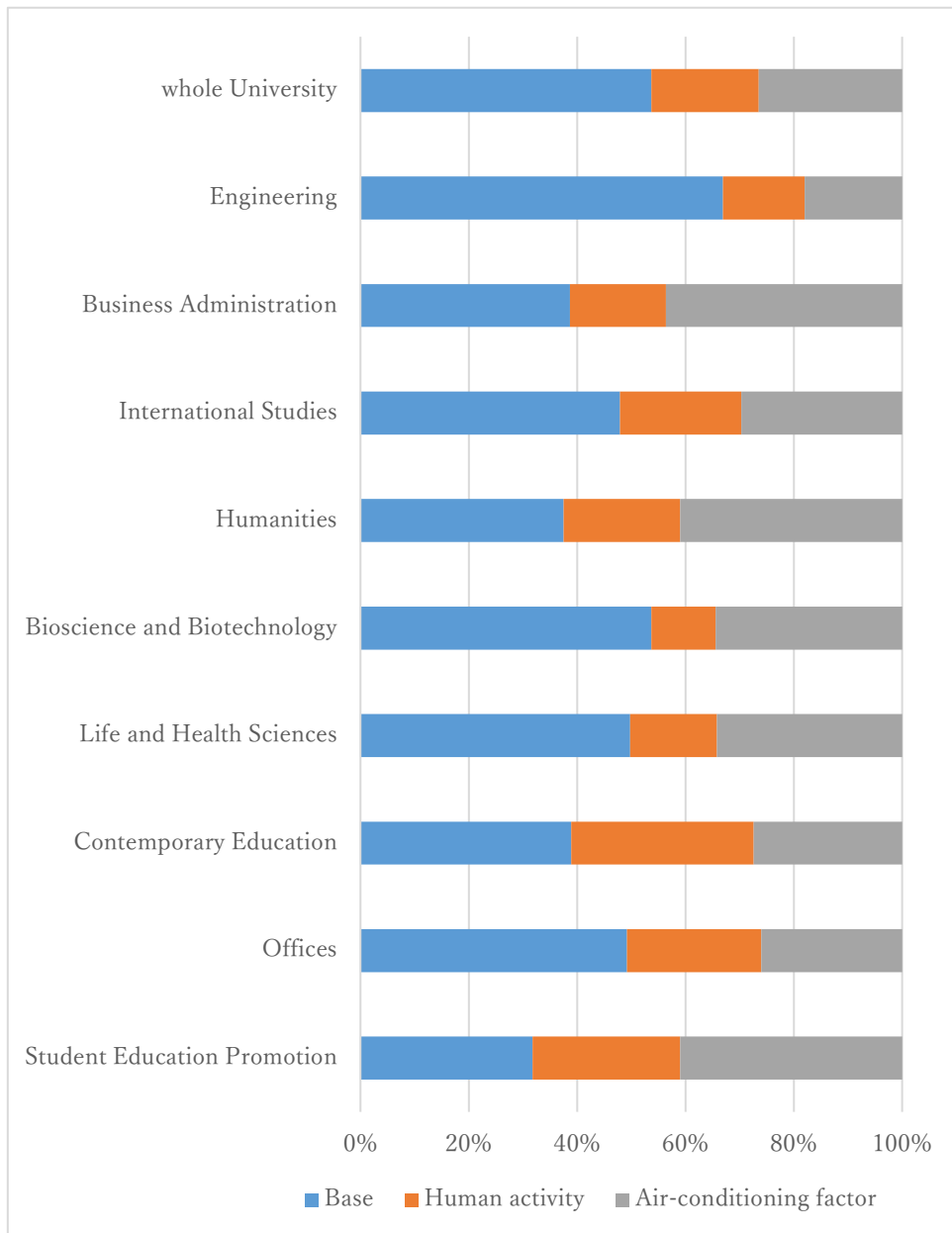
**Table 3.2** Correspondence between clusters and university schedules

cluster	period • day of week	Air-conditioning
6	July • January lecture day	○
5	spring • autumn lecture day	○
4	spring • autumn lecture day spring break entrance and final exams	○
3	spring • autumn lecture day	×
2	Saturday, summer vacation	×
1	Sunday, Year-end and new-year, Holidays	×

**Table 3.3** Cluster used for calculating energy consumption by factor

	Base	Human activity energy consumption factor	Air-conditioning energy consumption factor	Gas air-conditioning main usage
Engineering	①	②-① ③-①	④-③ ⑤-③ ⑥-③	
Bioscience and Biotechnology	①	②-① ③-①	④-③ ⑤-③ ⑥-③	○
Life and Health Sciences	①	②-① ③-①	④-③ ⑤-③ ⑥-③	○
Business Administration	①	②-① ③-①	④-③ ⑤-③ ⑥-③	○
International Studies	①	②-① ③-①	④-③ ⑤-③ ⑥-③	
Humanities	①	②-① ③-①	④-③ ⑤-③ ⑥-③	○
Contemporary Education	①	②-① ③-①	④-③ ⑤-③ ⑥-③	
Offices	①	②-① ③-①	④-③ ⑤-③ ⑥-③	○





**Figure 3.13** Estimated primary energy consumption by factor

### 3.4 Accuracy measurement of the Clustering analysis result

From figure 3.13, it is shown that using K-means clustering the energy consumption can be classified into energy consumption by factor. We did the accuracy measurement in building no. 52, where the energy consumption is measured separately by factor. Building no. 52 is the only building in Chubu university where, lighting, air-conditioning and consent is measured separately. In this building, lighting, consent and power are measured separately in the distribution panel on each floor. Lighting included the power consumption of lighting fixtures, ventilation fans and air-conditioning indoor units. Power includes the power consumption of

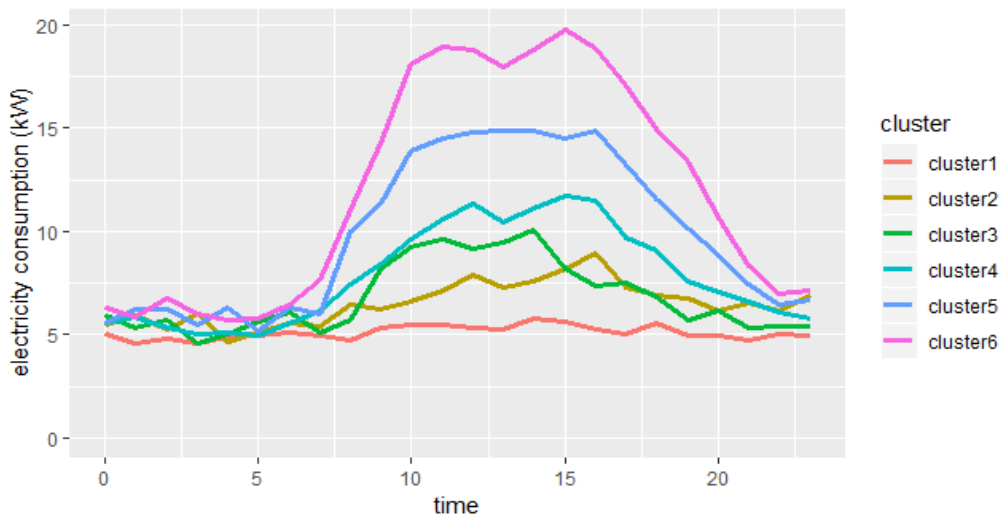


Figure 3.14 cluster centers: building no. 52



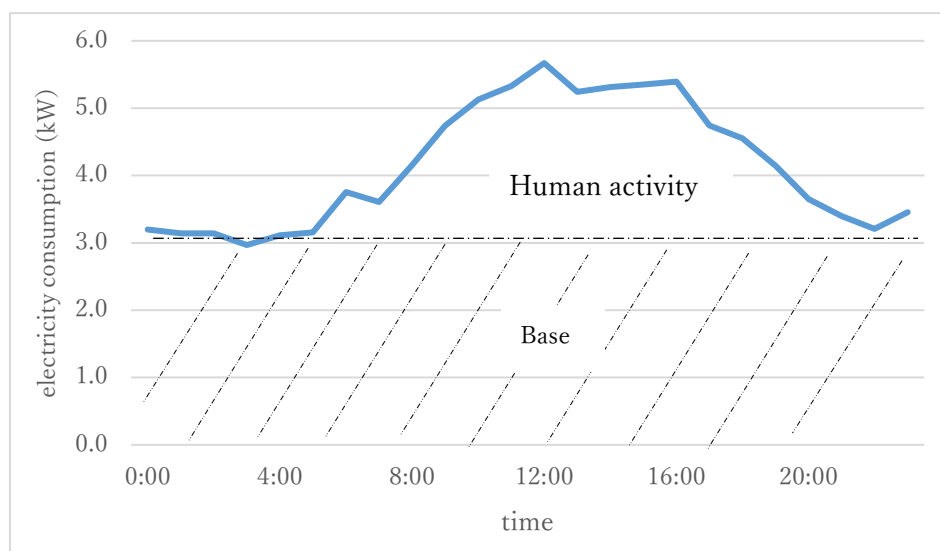
Figure 3.15 Calendar plot: building no. 52

drainage pumps and refrigerators, and does not include elevators.

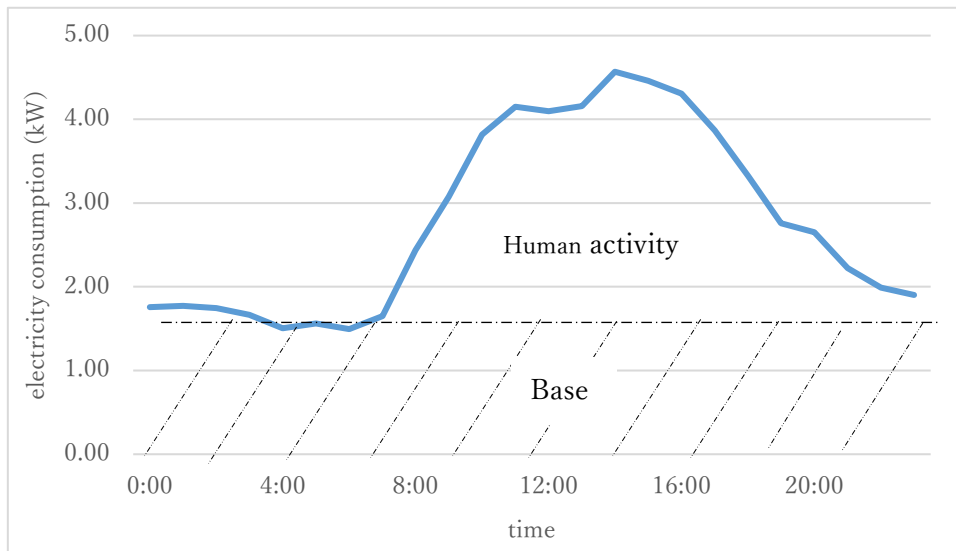
The measurement data is one-hour interval and is measured from April 2015 to March 2016. The power consumption of the lighting, consent, and power was measured individually in building no. 52 and clustering was performed on the entire data to estimate energy consumption by factor. Figure 3.14 shows the cluster center of each cluster of building no. 52. The clustering result was plotted in the form of calendar plot as shown in figure 3.15. Energy consumption in the winter season was found higher in comparison to summer season.

The outdoor unit power panel belongs to a separate building and is not included in the power consumption of building no. 52. On the other hand, there are 44 indoor units in all buildings, and the total power consumption is about 8% of the connected lighting system, the ratio of air conditioning to the power is small, thus the air-conditioning factor is included in human activities energy consumption. The air-conditioning in building no. 52 is a gas type Variable Refrigerant Flow (VRF) air-conditioner, and the gas used for air-conditioning are measured individually. The primary energy consumption for air-conditioning is the gas consumption converted to primary energy consumption. Therefore, the result of clustering and the measured primary energy consumption for air-conditioning are same.

Although, lighting and consent are measured separately, they included both the base and human activity energy consumption. Thus, it is necessary to separate both base and human activity energy consumption.

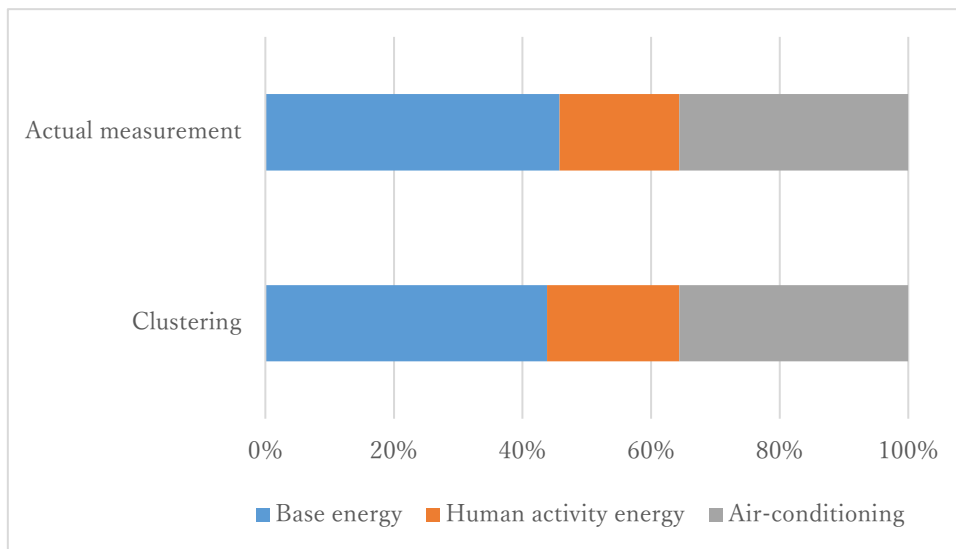


**Figure 3.16** One-year average hourly energy consumption of consent of building no. 52



**Figure 3.17** One-year average hourly energy consumption of lighting of building no. 52

Figure 3.16 is the one-year average of consent. In this figure, the value from 0 o'clock to 5 o'clock is the base and the rest is the human activity. Figure 3.17 is the average of annual lighting system. Similarly, the average value from 0 o'clock to 5 o'clock is used as the base and the rest as human activity energy consumption.



**Figure 3.18** Ratio of energy consumption by factor for measured and clustering result for building no. 52

Figure 3.18 shows that, the ratio of base, human activity energy consumption and air-conditioning for measured values and the clustering result is almost similar. Thus, clustering can be used for the classification of single energy consumption data of the university buildings into base, human activity energy consumption and air-conditioning energy consumption factor.

### 3.5 Utilization of energy management

Up to this section, we have confirmed the content of energy consumption analysis by factor using K-means method. Furthermore, we will examine the goal of low carbonization at the university. In the report [7], we achieved a 14% reduction in carbon emissions in the year 2016 compared to the year 2011. We will consider the reduction target up to the year 2030 from the result based on 2016. The table 3.4 shows the primary energy consumption by factor determined from the result of the entire university in 2016. As for human activities factor energy consumption, we expect to reduce power consumption by 40% by using LED and improving the efficiency of office automation equipment by 2030. Air conditioning is expected to reduce energy consumption by 15% by replacing equipment until 2030. In the energy white paper, the power emission factor is 0.37kgCO<sub>2</sub>/kWh. To achieve a 40% reduction rate in the civilian sector in the global warming countermeasures decided by the government, the base energy consumption needs to be reduced by 17.8% when calculating with a goal of 21% reduction in electricity.

**Table 3.4** Possible measures for CO<sub>2</sub> emission reduction

		Primary energy consumption (MJ/m <sup>2</sup> ·year)	CO <sub>2</sub> emission intensity (kgCO <sub>2</sub> /m <sup>2</sup> · year)	Emissions after measures applied until 2030	Target reduction rate
Base		526	26.8	32	17.8%
Human activity factor		169.4	8.6	5.2	40%
Air-conditioning factor	Electricity	73.5	3.7	3.2	15%
	Gas	213.4	10.8	9.2	15%
Adding solar energy production				-0.8	
Total		982.3	50.1	39.6	

### 3.6 Energy consumption pattern in different buildings

#### 3.6.1 Office building

The energy consumption in the case of office buildings consists of base energy consumption which runs 24 hours and all year-round. The rise in energy consumption in the day time is due to the energy consumption of air-conditioning, lighting, office equipment like printers, computers, fax, etc. and the energy consumption of consent.

The base energy consumption is near to 20% of the peak energy consumption. In comparison to the peak energy consumption the base energy consumption is small which is due to the absence of high energy-consuming devices that runs in off-office hours. The energy consumption is found to increase quite earlier than the office beginning hours. If the lighting is turned off in the day time, then the decrease in electricity consumption in the lunch break is seen. Figure 3.19 is a typical example of energy consumption in offices.

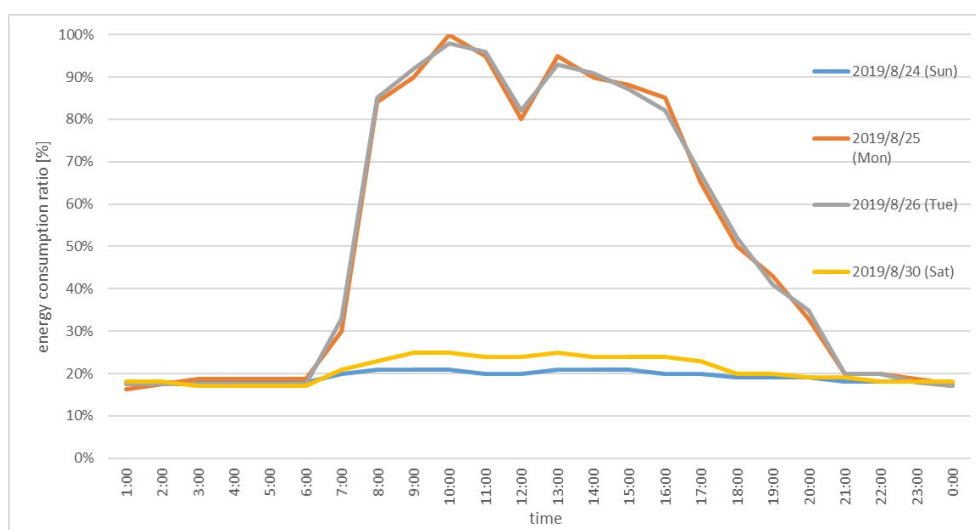


Figure 3.19 Energy consumption in offices. Figure adopted from [7].

#### 3.6.2 Hospitals

The base energy consumption in hospitals is more than 50% of the peak energy consumption. The base energy consumption is mainly due to the energy consumption of medical equipment that runs 24 hours and all the year-round regardless of whether it is weekdays or a weekend. An example of energy consumption in hospitals is shown in figure 3.20.

The peak energy consumption in the daytime is due to the energy consumption of offices of the administrative department and the energy consumption of the outpatient departments.

It is assumed that the base power is largely due to the special equipment of the medical device, but the specific power-saving measures are related to the medical practices, so discussions with relevant parties are necessary.

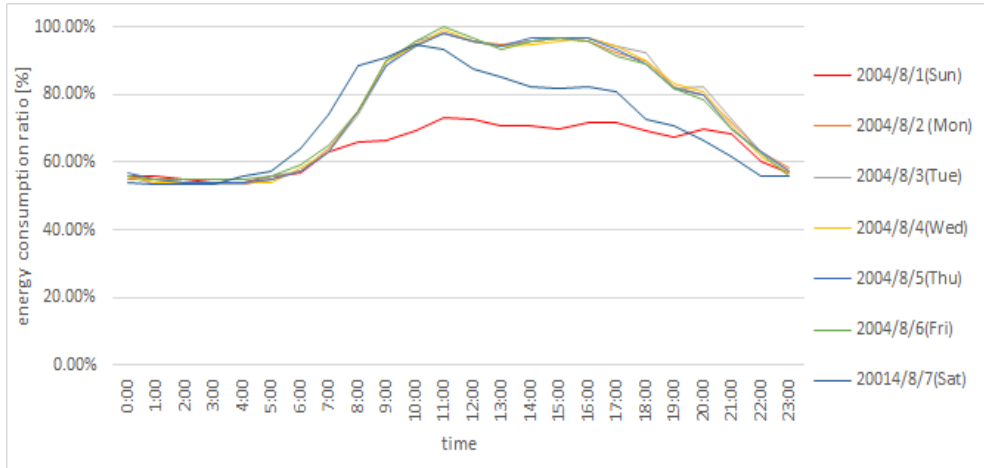


Figure 3.20 Energy consumption pattern in Hospitals. Figure adopted from [7].

### 3.6.3 Departmental store

In the case of departmental stores, the difference between the base and the peak energy consumption is found high. In figure 3.21, the base energy consumption is only 10% of the peak energy consumption. In addition, a sharp rise in energy consumption in the morning during the opening time and a sharp fall in energy consumption during closing time is seen. This characteristic is different from the office buildings as there is no effect of overtime working. This is also characterized by less holiday so; the yearly energy consumption is found high.

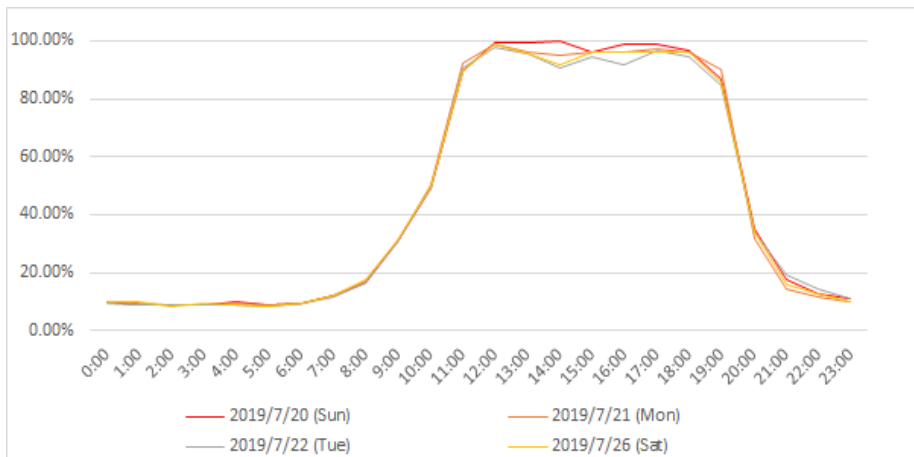


Figure 3.21 Energy consumption pattern in departmental stores. Figure adopted from [7].

### 3.6.4 Hotels

Figure 3.22 shows the energy consumption of hotels in the summer. District heating and cooling is used so the seasonal variation in energy consumption is low. Peak energy consumption during check-in and check-out hours can be observed. However, the variation in energy consumption is low between check-in and check-out. The base energy consumption is above 60% of the peak energy consumption. There are no holidays so the yearly energy consumption is quite higher. Due to the multi-room configuration, turning off the air-conditioning and lighting when the customer is not in the hotel can help in energy conservation. Figure 3.22 represent the energy consumption of typical hotel.

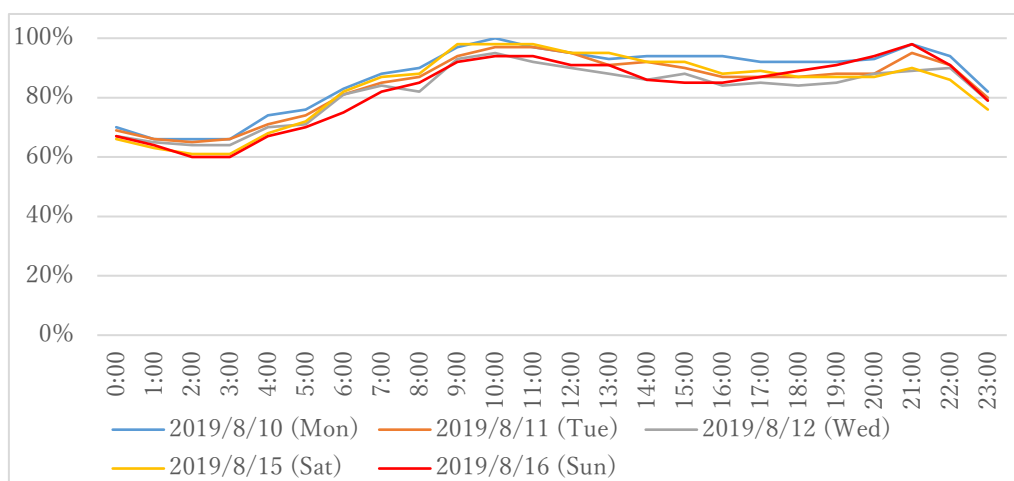


Figure 3.22 Energy consumption pattern of hotels. Figure adopted from [7].

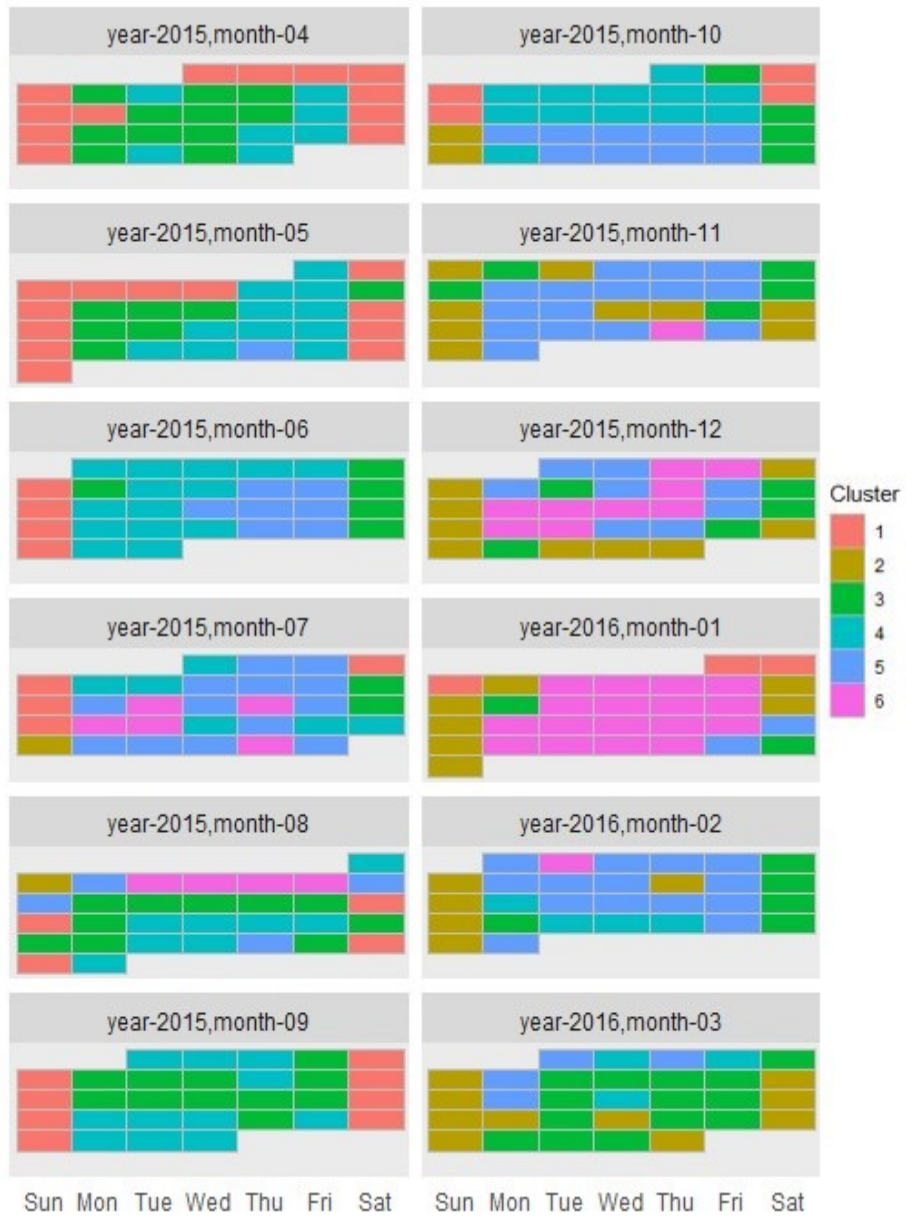
### 3.7 Analysis of Energy Consumption Pattern of University Building Using Clustering Technique

To analyze the energy pattern of different building of Chubu University, we apply the proposed K-means method to find the energy pattern in three buildings of Chubu university namely building no. 3-7-8 belonging to the college of engineering, building no. 21 belonging to College of Business and building no. 1 belonging to office buildings. The college of Engineering contains different research equipment and servers that run continuously. Analyzing the energy pattern of these buildings can clarify the impact of experimental and research equipment on building energy consumption.

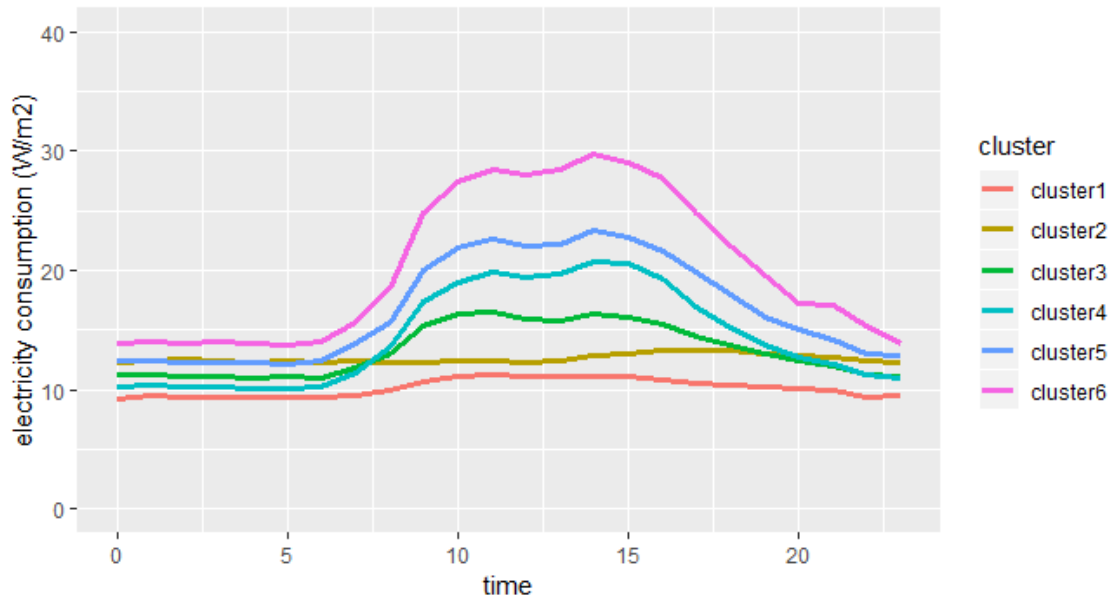
In case of buildings of college of Engineering, base energy consumption, which is represented by cluster 1 in figure 3.24, as well as the separation between cluster 1 and cluster 6 from 0:00 AM to 5:00 AM, was also found to be large. This feature of the College of Engineering building differs from Non-science departments and the Administrative buildings.



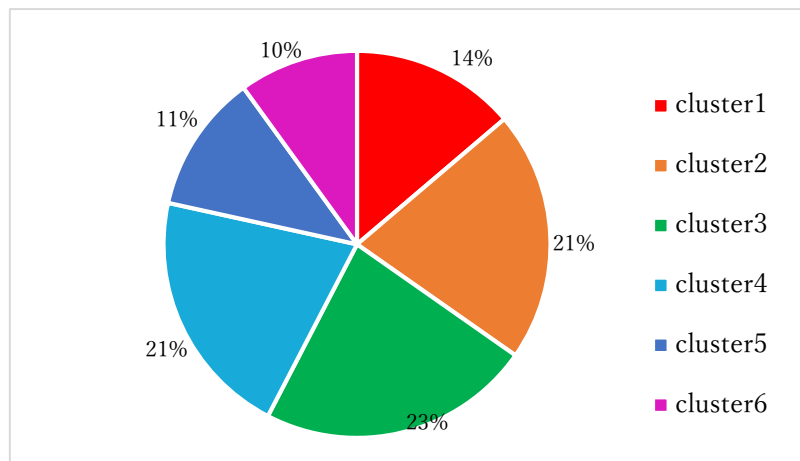
The large consumption of energy on non-air-conditioning days, as represented by clusters 1 to 3 in figure 3.25, indicates the effect of research equipment.



**Figure 3.23** Calendar plot for building no 3-7-8 of College of Engineering



**Figure 3.24** Cluster centers for building no. 3-7-8 of College of Engineering



**Figure 3.25** Percentage of energy consumption per cluster building no. 3-7-8 of College of Engineering



Figure 3.26 Calendar plot for building no. 1 of the office building

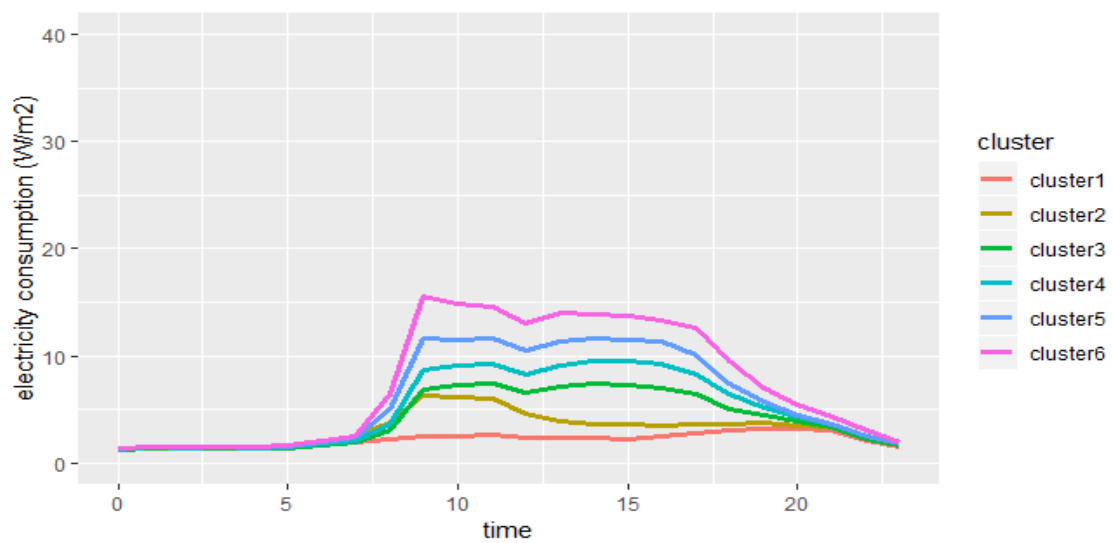
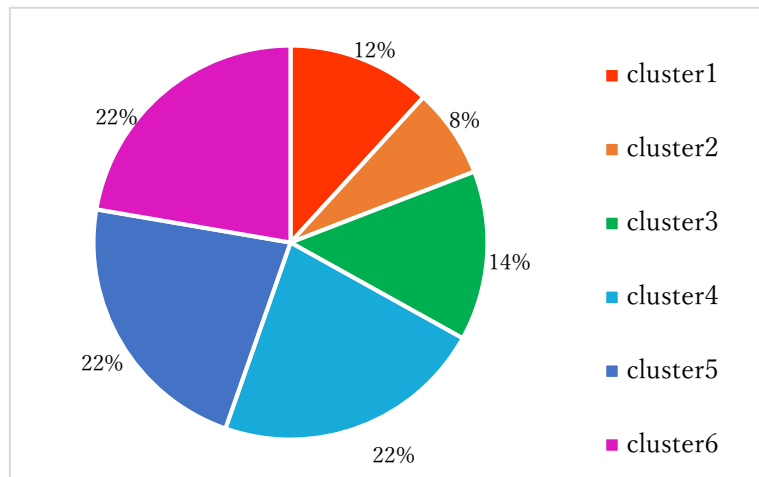


Figure 3.27 Cluster centers for building no. 1 of the office building



**Figure 3.28** Percentage of energy consumption per cluster in building no. 1 of the office building

Figure 3.26 represents the calendar plot of the Administrative building of Chubu University. The Administrative building does not host any research equipment that runs nonstop for 24 hours as found in the College of Engineering. Thus, the base energy consumption, which is represented by cluster 1 in Figure 3.27, is found to be quite small. The low energy consumption of clusters 1 to 3, as represented by figure 3.28, indicates the lack of research equipment in the Administrative building. Most of the energy consumption occurs on air-conditioning days, represented by clusters 4, 5, and 6. The energy consumption increases sharply around 8:00 AM, when the employees start working. Similarly, consumption decreases from 12:00 PM to 1:00 PM, as found from clusters 2 to 6, due to the lunch break. Clusters 1 to 6 coincide with each other from 0:00 AM to 5:00 AM, which also supports the lack of research equipment in office buildings.



Figure 3.29 Calendar plot for building no. 21 of College of Business

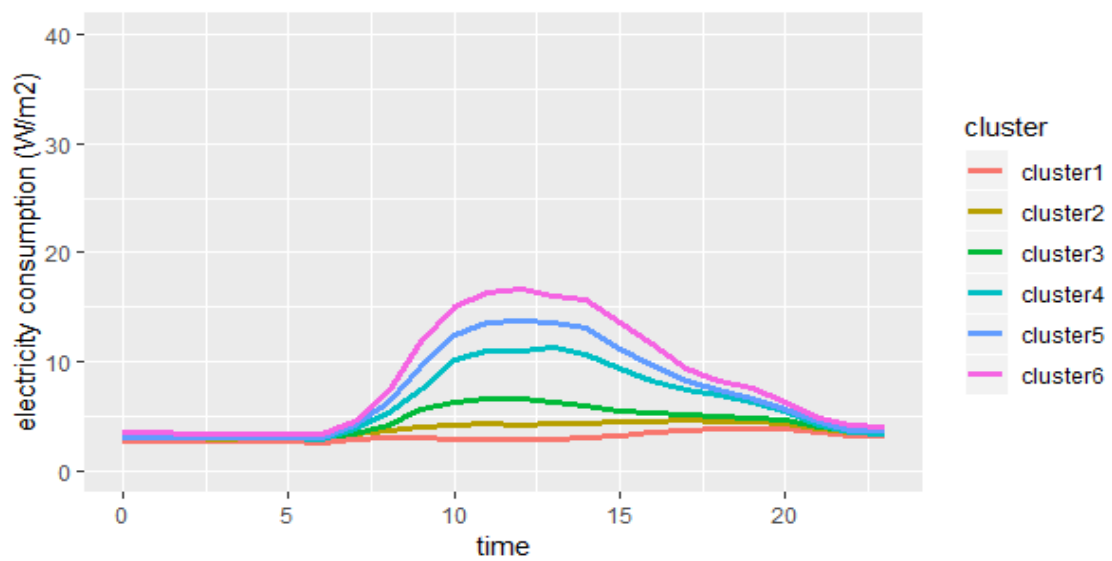
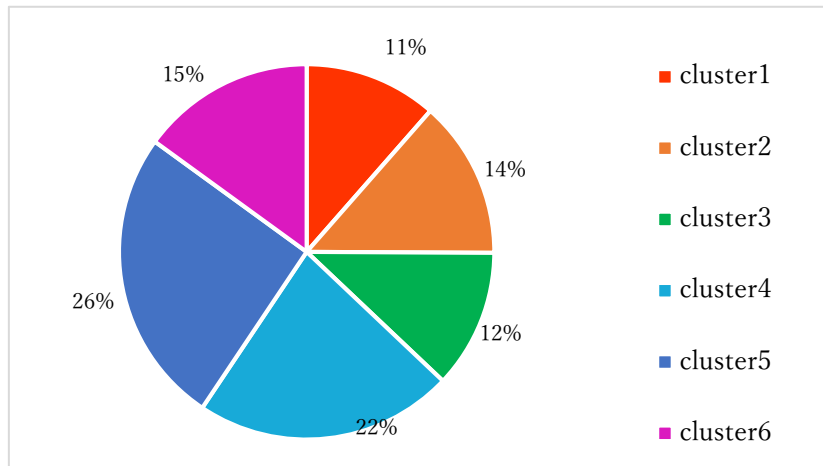


Figure 3.30 Cluster centers for building no. 21 of college of Business



**Figure 3.31** Percentage of energy consumption per cluster of building no. 21 of the College of Business

Figure 3.29 shows calendar plot for the College of Business. The base energy consumption, which is represented by cluster 1 in figure 3.30, is small, and the difference between clusters 1 and 6 between 0:00 AM and 5:00 AM is very small in comparison to the College of Engineering. The low energy consumption of clusters 1 to 3, as represented by figure 3.31, again indicates the lack of research equipment. Most of the energy consumption is on air-conditioning days, represented by clusters 4 to 6. This shows that energy consumption in Non-science departments is quite similar to that of the Administrative building at the university.

In case of Chubu university in the year 2015 the air-conditioning cooling period was from 15<sup>th</sup> June to 15<sup>th</sup> September whereas heating period was from 15<sup>th</sup> November 2015 to 15<sup>th</sup> April 2016. In Chubu university usage of air-condition in the interim period is not allowed except sudden changes to extreme weather condition and some other exceptional cases. Cluster 1 usually falls on Sundays and national and university holidays when no lighting or air-conditioning is being used. The energy consumption on these days is due to sensors, servers, refrigerators and research equipment that run 24 hours a day all year round. This is the least amount of energy consumption that occurs in a day in a year. Thus, cluster 1 yields the base energy consumption of the university if its value is summed for one whole year, i.e.:

$$\text{Yearly base energy consumption} = E \times N \quad (6)$$

Where,

$E$  = energy consumption of cluster 1

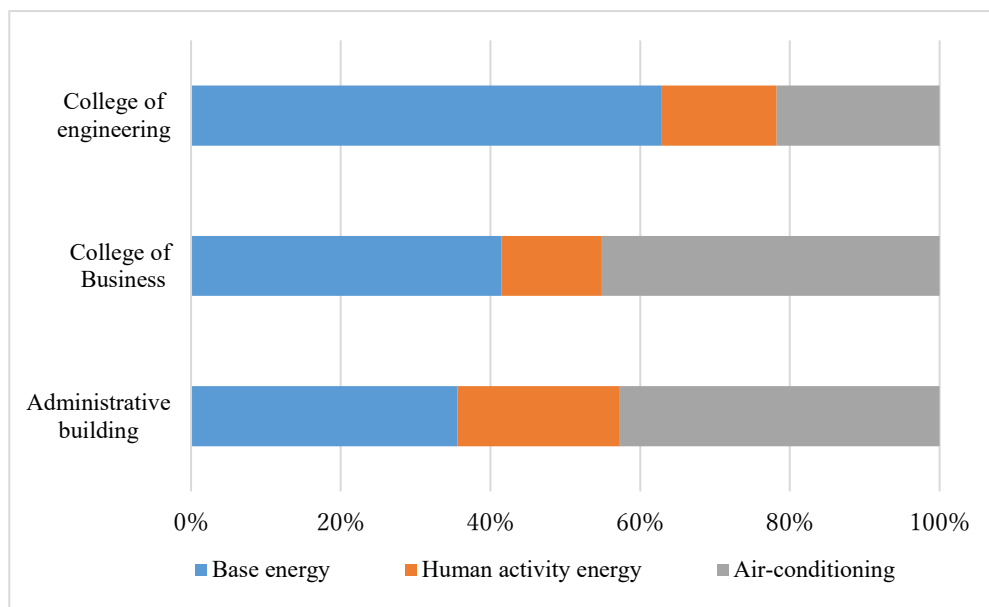
$N$  = number of days of the year

Base energy consumption has always been an important topic for building energy

consumption reduction. Unlike peak energy, which occurs for few hours in the daytime, base energy occurs for whole 24 hours so, reduction of hourly base energy consumption of a building by 1kWh can lead to the yearly base energy consumption reduction by 8,760 kWh.

Energy consumption due to human activities is the energy consumed by the use of lightening, computers, OA equipment, elevators etc., excluding the energy consumption of air-conditioning. Since cluster 2 and 3 mostly falls on the days without the use of air-conditioning, energy consumption due to human activities can be obtained by subtracting energy consumption of cluster 1 from cluster 2 and 3 and summing them.

Cluster 4, 5 and 6 mostly falls on the days with the usage of air-conditioning so, the energy consumption by air-conditioning can be obtained by subtracting energy consumption of cluster 3 from cluster 4, 5 and 6 respectively and summing them. In Chubu university both electricity and gas are used for air-conditioning so, the primary energy consumption of gas is also considered.

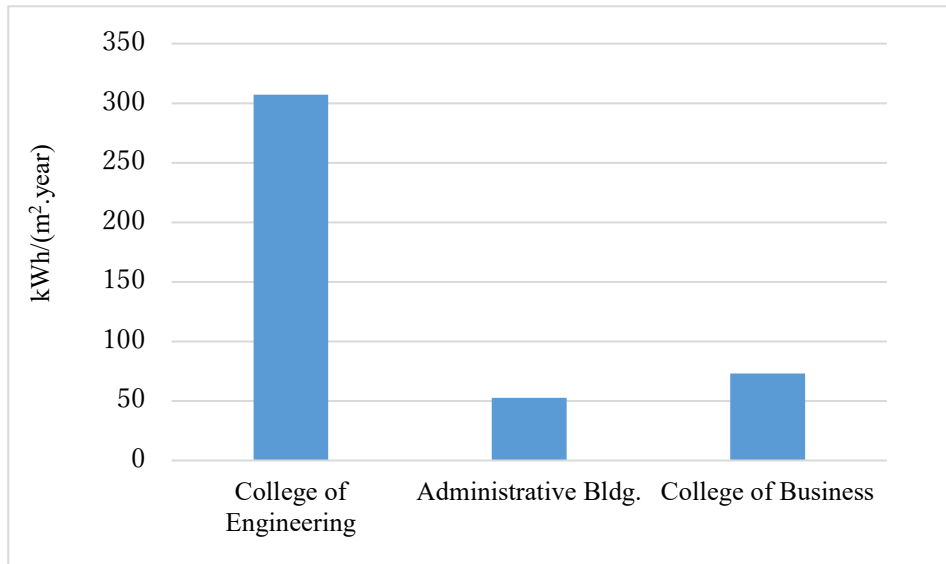


**Figure 3.32** Classification of energy consumption by factor in three different buildings of Chubu University

Figure 3.32 shows the classification of energy consumption by usage in three different buildings of Chubu university. Base energy consumption is found higher for College of Engineering whereas in case of other two buildings air-conditioning energy consumption is found higher.

Using equation (6), the base energy consumption of the college of Engineering, the administrative building, and the college of Business was calculated. Figure 3.33 shows that

the base energy consumption per square meter per year in the College of Engineering is 5.9 times greater than that of the Administrative building and 4.2 times greater than the College of Business. High energy consumption in the College of Engineering is due to research equipment and machines that run 24 hours a day, which are not present in the other two buildings.

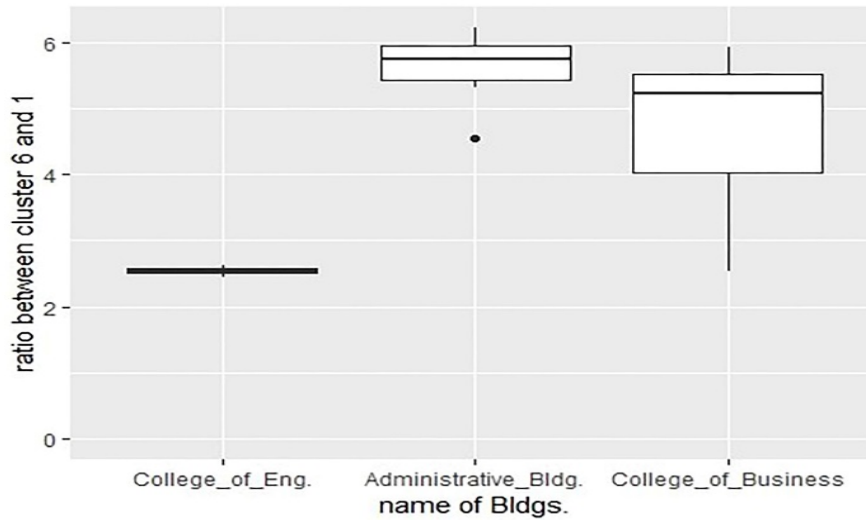


**Figure 3.33** Bar graph indicating the yearly base energy consumption per unit area in the three different buildings of Chubu University

To find the difference between the peak and base energy consumption in the above mentioned buildings, the ratio of the energy consumption of cluster 6 and cluster 1 was calculated from 10:00 AM to 17:00 PM. The boxplot shown in figure 3.34 shows the difference in peak energy consumption between these three buildings. The average ratios of 2.5, 5.62, and 4.70 between cluster 6 and cluster 1 were found for the College of Engineering, the Administrative building, and the College of Business respectively.

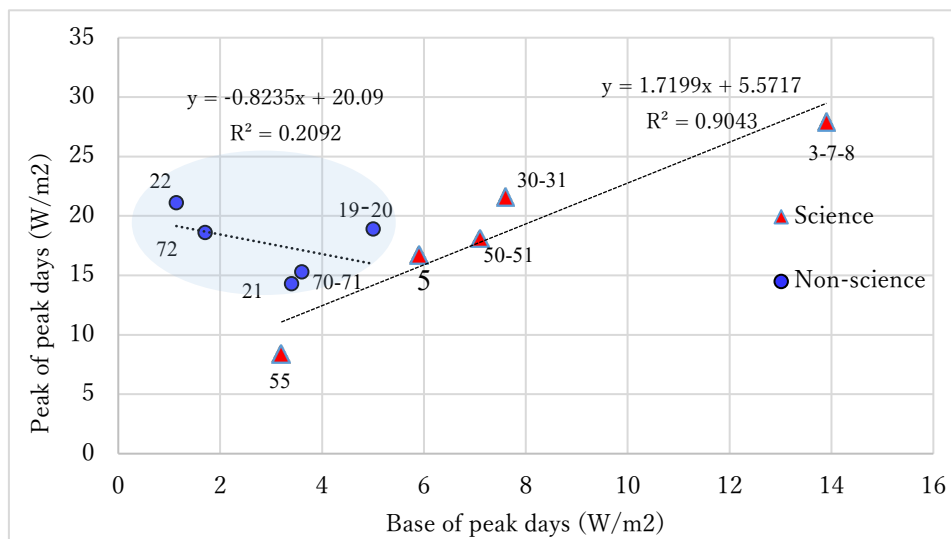


A large base energy but a small ratio between base and peak energy shows that base energy consumption reduction is of primary focus in the College of Engineering, whereas a low base energy consumption but a high ratio between base and peak energy consumption in



**Figure 3.34** Boxplot showing the ratio between peak and base energy for three buildings at Chubu University

the administrative and non-science departments show that there is more potential for the reduction of peak energy consumption than the base energy consumption.



**Figure 3.35** Scatter plot showing the relation between base and peak energy consumption in buildings of Science and Non-science departments

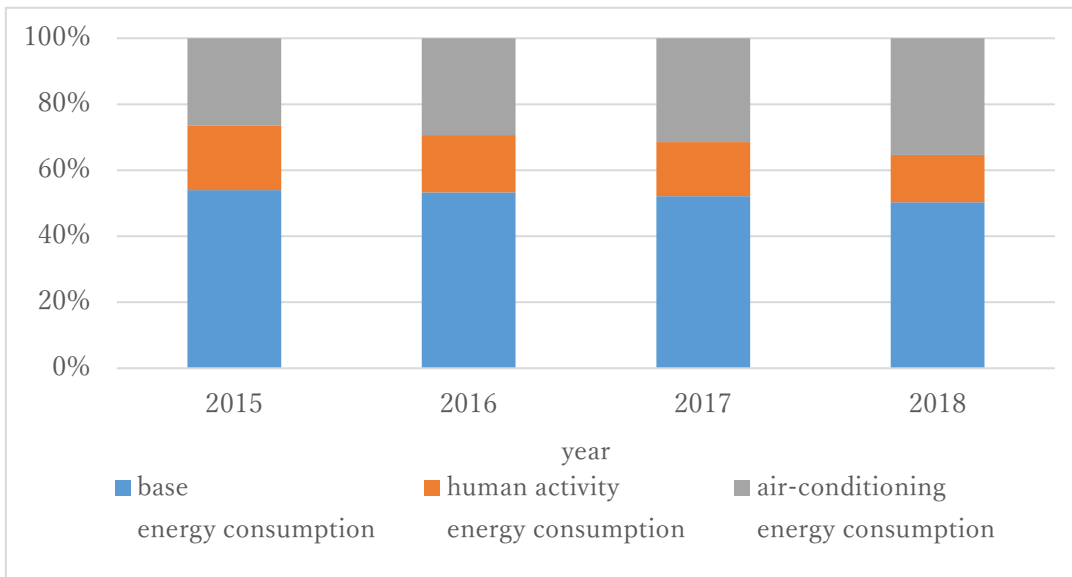
To find the relation between base and peak energy in case of Science and Non-science buildings of Chubu university, additional buildings are included and a scatter plot as shown in figure 3.35 is drawn. Blue points represent the buildings of Non-science departments whereas the red triangle represents the buildings of Science departments. The numbers near the points and triangle are the name of respective buildings of Chubu university. Average base energy in this figure is the average of energy consumption from 0 AM to 5 AM whereas average peak energy consumption is the average of energy consumption from 10 AM to 5 PM. In case of buildings of non-science departments,  $R^2$  value was found small so, the relation between base and peak energy consumption was not found whereas, in case of buildings of Science departments,  $R^2$  value was 0.904 so, peak energy consumption for peak days was found to increase with the increase in base energy consumption.

### 3.8 Yearly change in energy consumption pattern of Chubu University

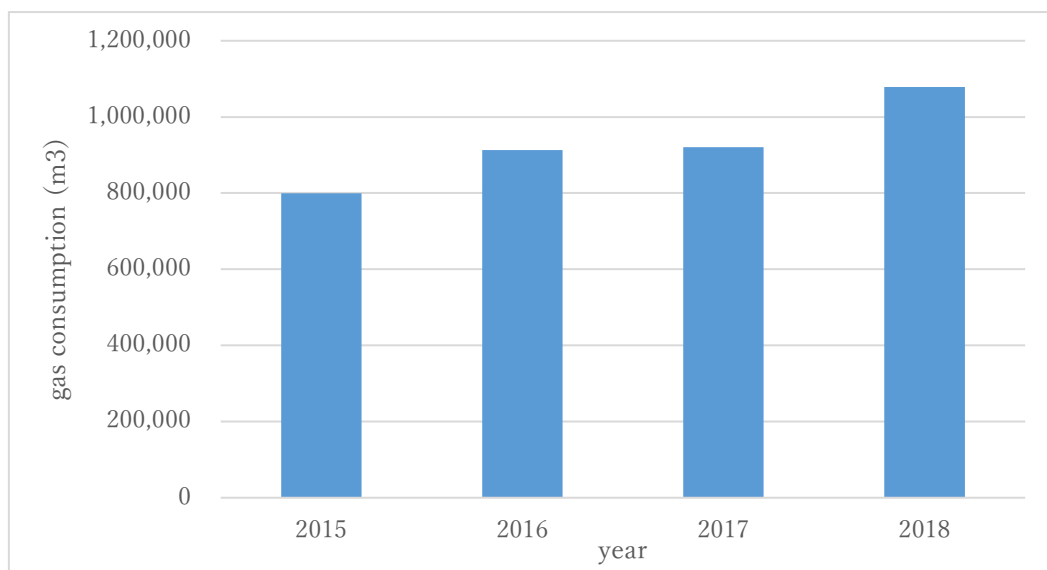
Clustering method can be used for the analysis of yearly energy consumption. For this purpose, the energy consumption by usage is calculated for each year from 2015 to 2018.

Table 3.5 Energy consumption by factor in whole university from the year 2015-2018

Energy consumption by factor	2015	2016	2017	2018
	Energy consumption (kWh)			
Base	28,052,284	28,136,386	26,453,530	26,496,733
Human activity	10,144,616	9,052,489	8,240,745	7,505,409
Air-conditioning	13,743,106	15,337,229	15,849,615	18,355,082



**Figure 3.36** Change in yearly primary energy consumption by factor in whole university



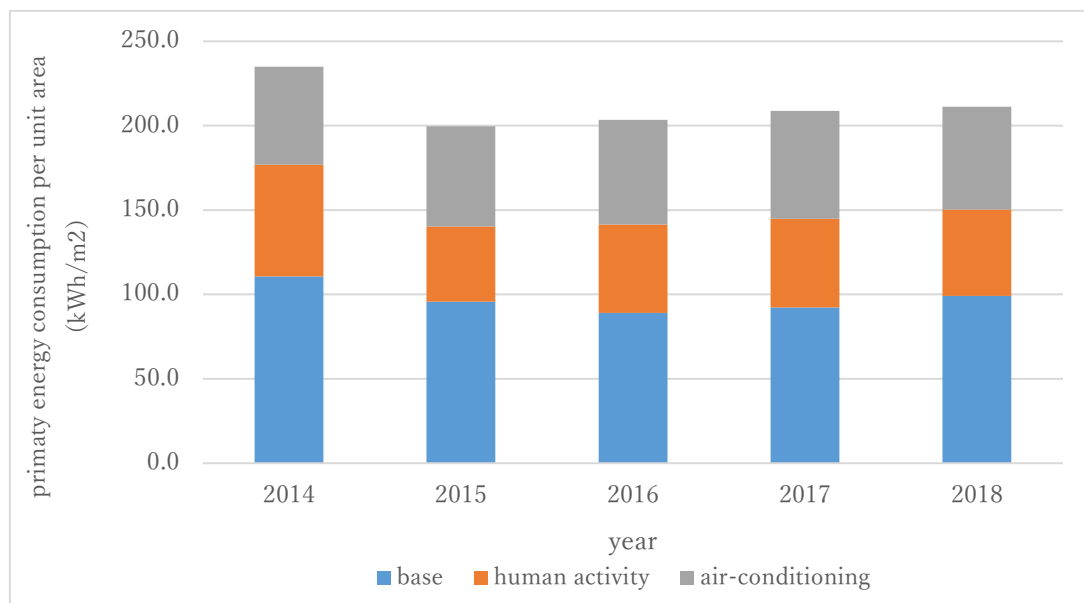
**Figure 3.37** Change in yearly gas consumption (m<sup>3</sup>) in whole university

Air-conditioning energy consumption in the year 2018 is quite higher in comparison to the year 2017. This is also supported by the fact that the gas consumption in this year is quite higher in comparison to the previous year as shown in figure 3.37. The outdoor temperature in the year 2018 was quite higher so, the air-conditioning load was relatively higher compared to the previous year, resulting in the higher air-conditioning energy consumption.

### 3.9 Comparison between gas air-conditioning usage building and electricity air-conditioning usage building

#### 3.9.1 Analysis of annual change in energy consumption pattern

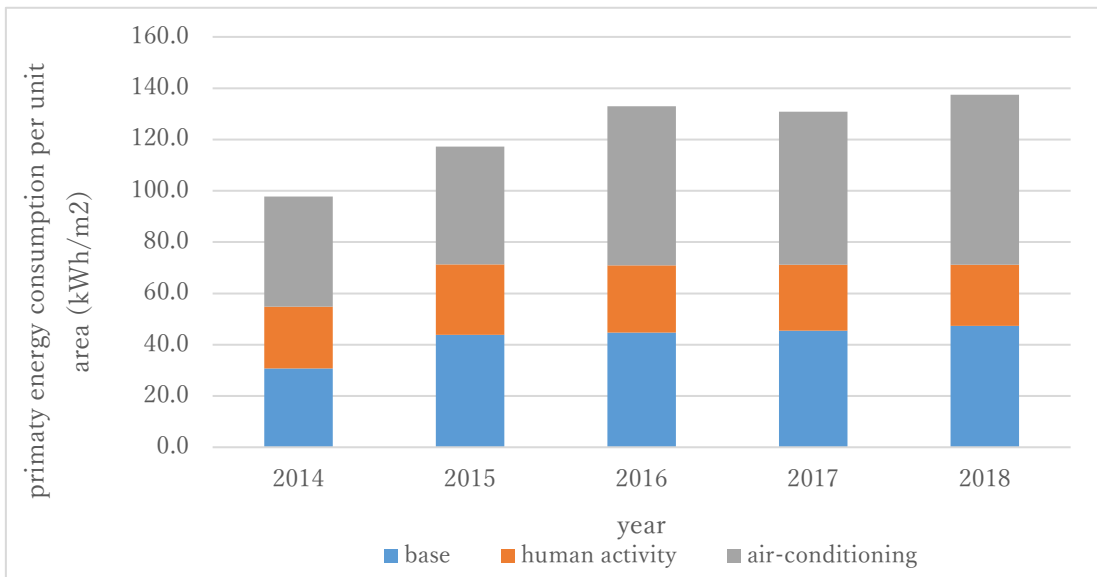
To investigate the characteristics of power consumption for each building we apply clustering method. Building no. 52 which uses gas air conditioning and building no. 20 which uses electric air conditioning are selected as target.



**Figure 3.38** Yearly change in energy consumption by factor in building no. 20

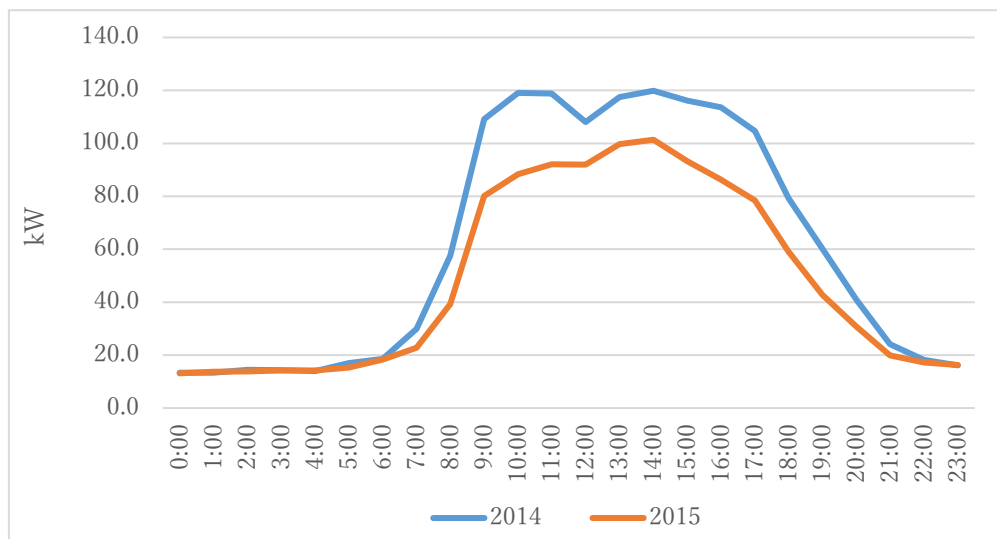
Building no. 52 is a building of Life and Health Science department, which is a four-story building with a total floor area of 2,137 m<sup>2</sup>. Building 20 is a building of International relation department, which is a 12 story building with a total floor area of 5,576 m<sup>2</sup>. Including the area of the related buildings where the electricity is measured combine, the total area is 9,566m<sup>2</sup>. Figure 3.38. and 3.39 represent the energy consumption per unit area estimated using clustering result in a stacked bar graph. Energy consumption per unit area in the building no. 20 in the year 2015 is found to be reduced in comparison to the year 2014. Reduction in the energy consumption is due to the contribution of reduction of base, human activity and air-conditioning energy consumption.

One of the reason for air-conditioning energy consumption reduction is due to the renewal of air-conditioning in the building no. 20. Building no. 20 is a building of Humanities department where no experimental equipment is used. Thus, there is no effect of research and experimental equipment in this building. 190 out of 202 equipment were renewed, which contributed in the energy consumption reduction.



**Figure 3.39** Yearly change in energy consumption by factor in building no. 52

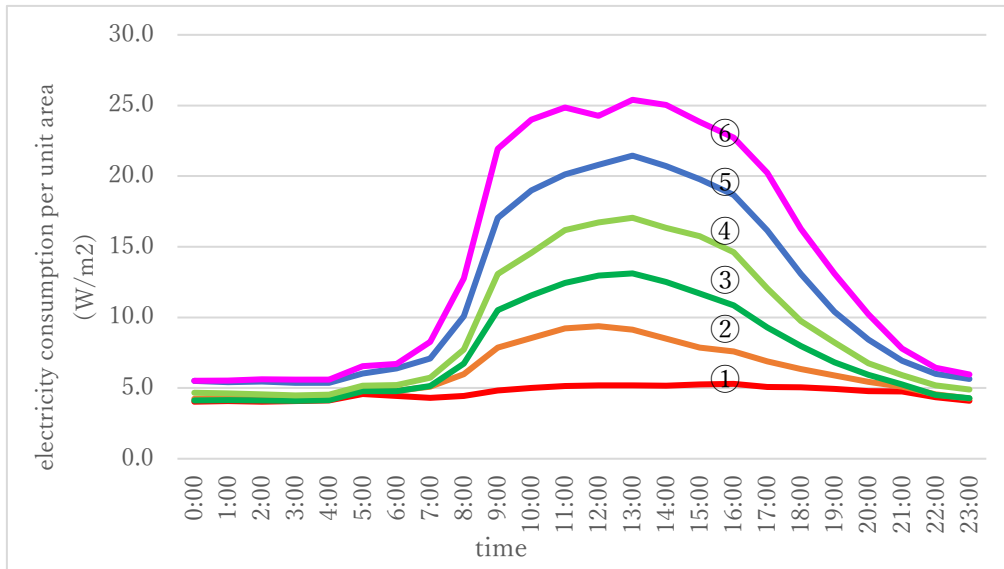
To find the effect of Electric Heat Pump (EHP) renewal on air-conditioning energy consumption reduction cluster analysis was performed from academic year 2014 to 2018 in building no. 20. Cluster 6 represents the energy consumption in the peak air-conditioning days in Chubu University. So, cluster 6 was extracted from the clustering result and was subtracted from cluster 3, which represent the energy consumption due to lighting, computers etc. The difference between cluster 6 and cluster 3 is considered to be the power consumption of air-conditioning during peak load hours.



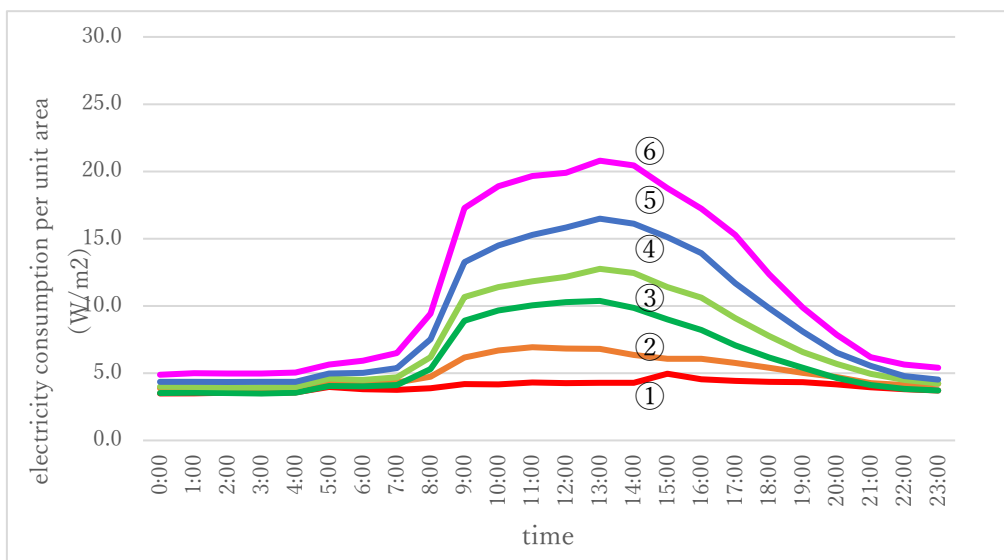
**Figure 3.40** Peak air-conditioning energy consumption of cluster 6 of building no. 20 of the year 2014 and 2015

The energy consumption due to air-conditioning in the peak air-conditioning period is shown in the figure 3.40. It is found that due to the effect of EHP renewal the peak air-conditioning energy consumption is found to reduce by 20.2%.

From figure 3.39 it is found that the energy consumption of building no. 52 is found to increase from the year 2014 to 2018 excluding the year 2017. Increase in energy consumption is mainly found in base and air-conditioning energy consumption.

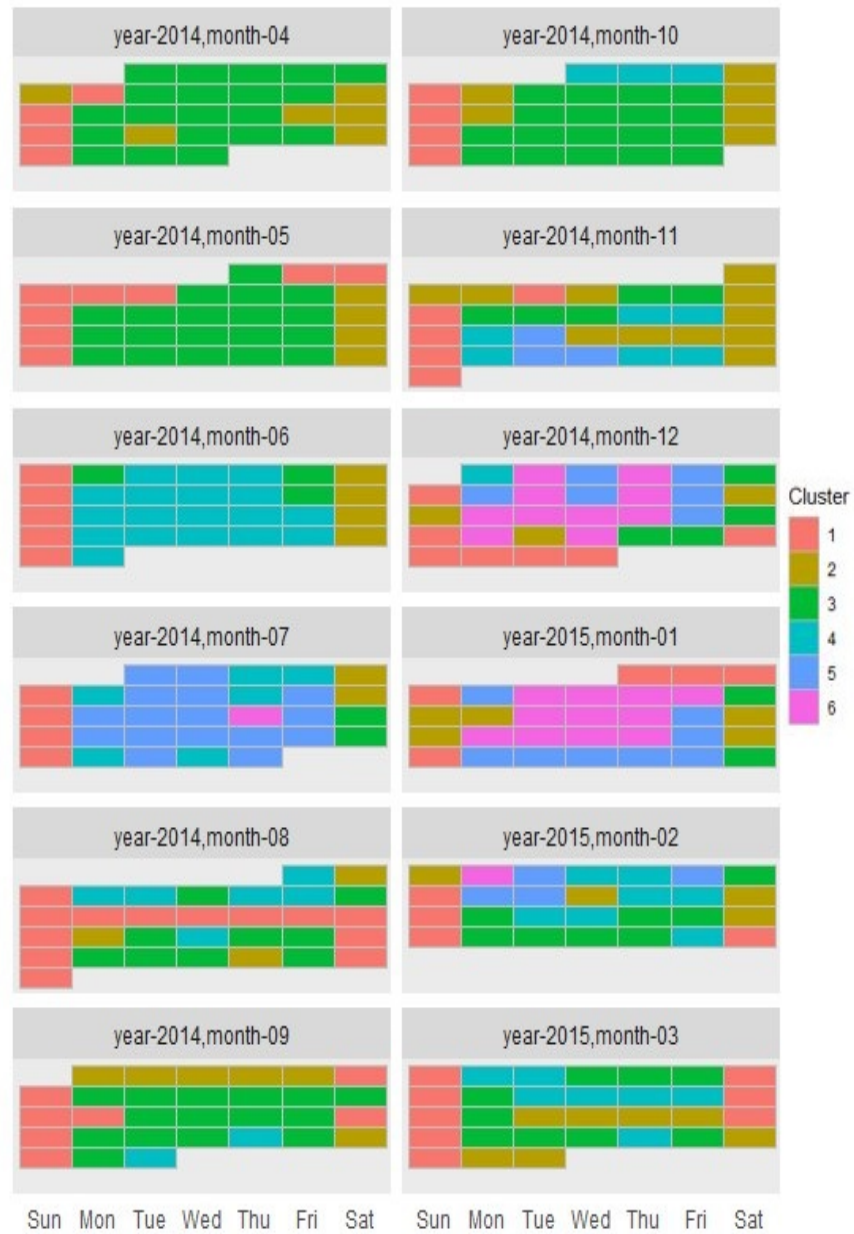


**Figure 3.41** Cluster centers of building no. 20 of the year 2014



**Figure 3.42** Cluster centers of building no. 20 of the year 2015

Figures 3.41 and 3.42 shows the cluster center of the clustering result of building no. 20 of the year 2014 and 2015 respectively with the numbers in circles representing their cluster number. The reduction in the peak electricity consumption as represented by cluster 6 can be observed clearly.



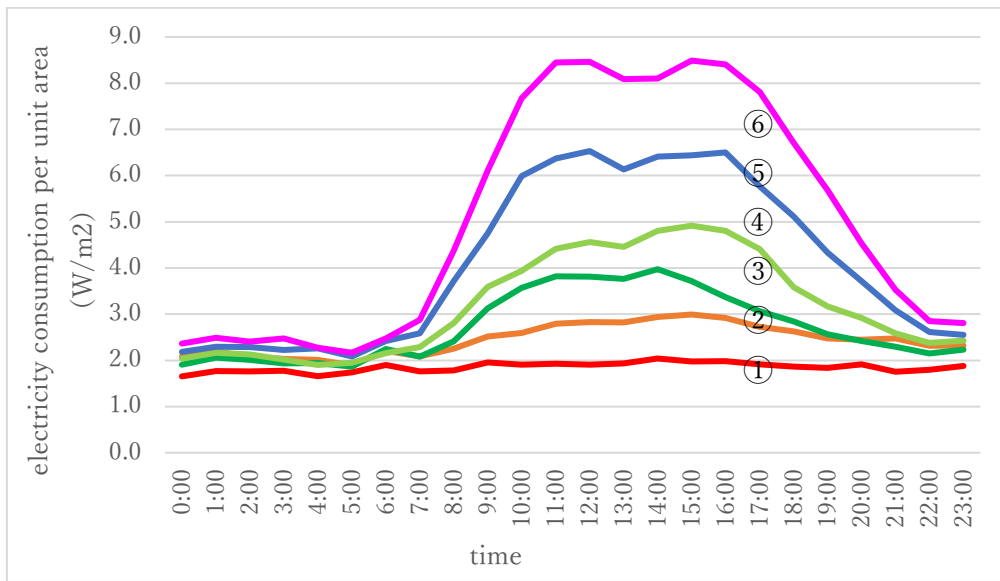
**Figure 3.43** Calendar plot of building no. 20 of the year 2014



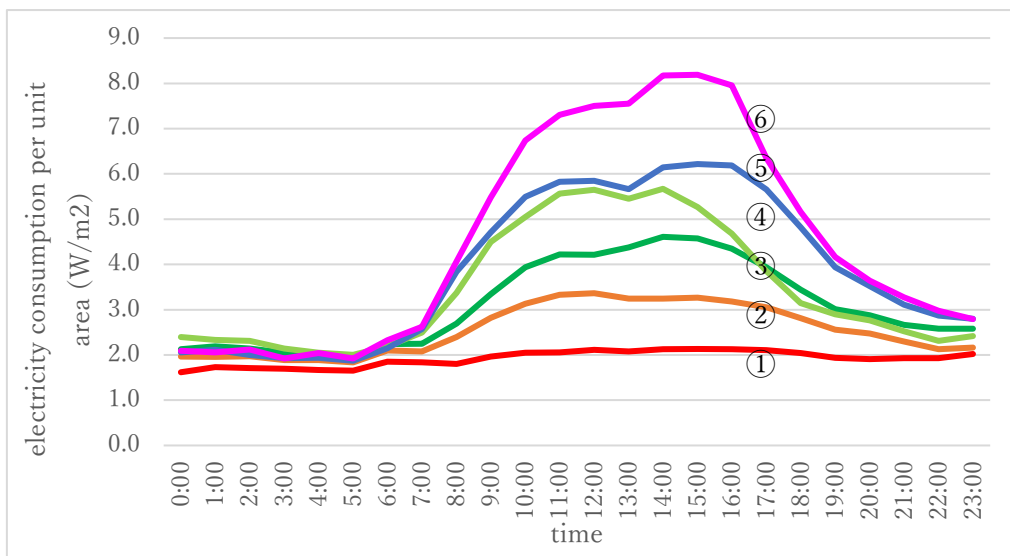
**Figure 3.44** Calendar plot of building no. 20 of the year 2015

Figures 3.43 and 3.44 show the calendar plot of the clustering result of building no. 20 of the year 2014 and 2015. In the year 2014 the days with highest energy consumption represented by cluster 6 is found mostly in the December and January of the winter season. In the year 2015 the days with highest energy consumption is found both in the summer and the winter season.

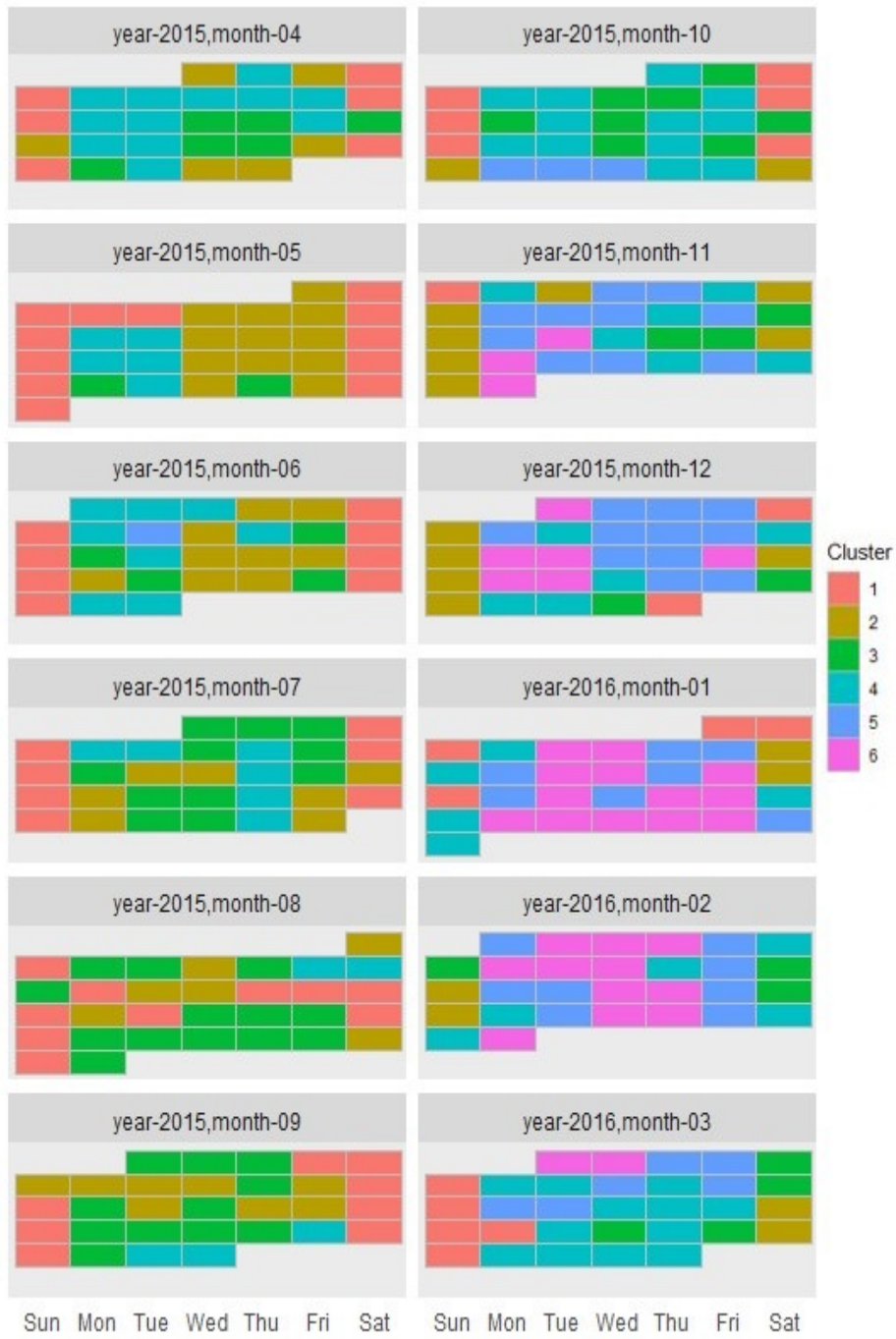




**Figure 3.45** Cluster centers of building no. 52 of the year 2015



**Figure 3.46** Cluster centers of building no. 52 of the year 2017



**Figure 3.47** Calendar plot of building no. 52 of the year 2015



**Figure 3.48** Calendar plot of building no. 52 of the year 2017

Figures 3.45 and 3.46 shows the cluster center of the clustering result which is energy consumption per unit area of building no. 52. It is found that there is not so much difference in the amount of energy consumption in this building in the year 2015 and 2017. But difference in the energy consumption pattern in the energy consumption peak days represented by cluster 6 is observed. Building no. 52 is a building where only gas is used for air-conditioning so, the effect of air-conditioning does not appear in the clustering result. In both of the calendar plots as shown in the figures 3.47 and 3.48, the days with higher energy consumption is found mainly in the winter season.

## Chapter 4

### Electricity load Forecasting

#### 4. Introduction

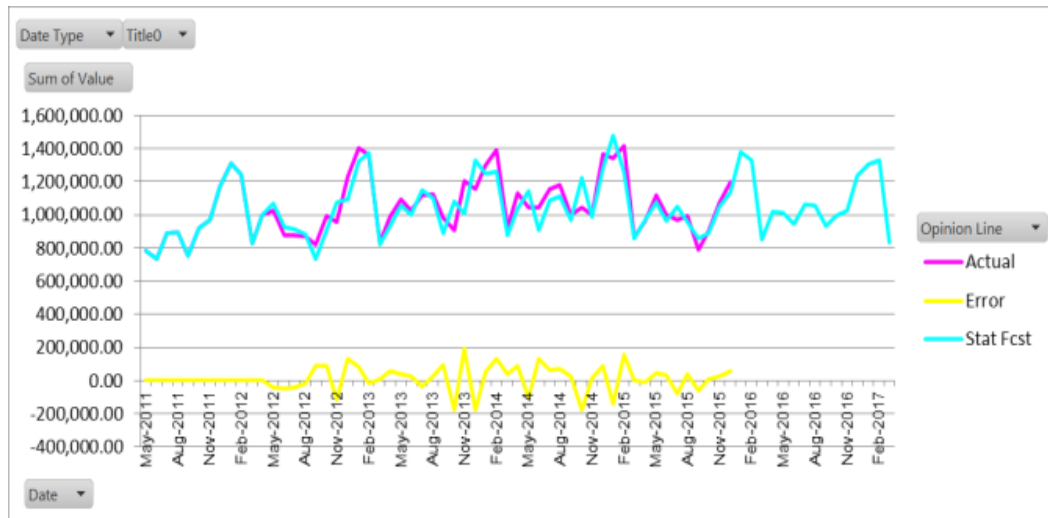
Electricity load forecasting refers to the prediction of load behavior for the future to meet the demand and supply equilibrium. Electricity load forecasting is an important process that can increase the efficiency for the electrical generating and distribution companies, managing authorities of buildings etc. The forecasting helps the utility companies in their operation and management of the supply to their customers.

Some advantages of load forecasting can be listed as below

- Enables the utility and building management authorities to plan well since they have an understanding of the future consumption of load demand
- It minimizes the risk of the utility company as understanding the future long term load helps the company to plan and make economically viable decisions in regard to future generation and transmission investment.
- Maximum utilization of power plants. The forecasting avoids under generation or over generation.
- Helps in deciding and planning for maintenance of the power system. The forecasting avoids under generation or over generation.

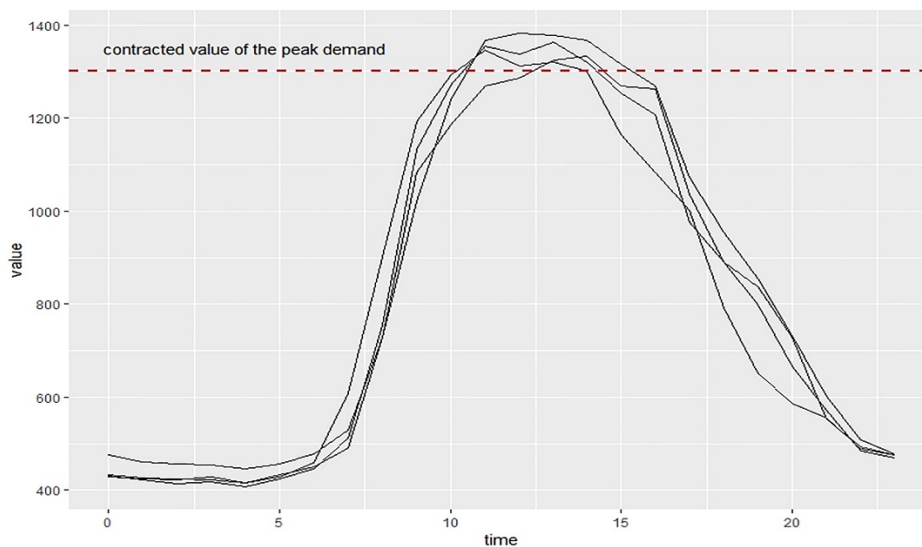
After Great East Japan earthquake, peak load reduction and power saving have become an issue. Many Japanese universities, as well as commercial buildings have made effort for energy conservation and reduction in CO<sub>2</sub> emissions. For energy management and demand response in the buildings, accurate electricity load forecasting is essential. The purpose of this research is to support energy management, such as demand response, peak load reduction etc. in university buildings by the development of a forecasting model, that improves the forecasting accuracy in comparison to existing forecasting methods.

In this paper, we proposed a hybrid model, a combination of clustering and ARIMA model to produce more accurate peak load forecasting of university buildings. Using auto.arima function the selection of ARIMA model has been made automatic so, if we have only time series data, the forecasting of the electricity load with good accuracy can be done without considering the building schedule, presence of holidays, presence of events etc. for forecasting. A typical example of forecasting and its forecasting error is shown in figure 4.1.



**Figure 4.1** A typical example load forecasting and its forecasting error. Figure adopted from [36]

#### 4.1 Motivation for electricity load forecasting



**Figure 4.2** Electricity load graph with peak load crossing the demand electricity load in West Campus of Chubu University

Electricity fee is charged according to the amount of electricity consumed and the contracted value of peak demand. If the electrical load exceeds the contracted value of peak demand, customers need to pay more on the electricity bill and thus avoid exceeding this contracted value.

Figure 4.2 Shows the electricity load on July 17,19,23 and 6<sup>th</sup> August of the academic year 2018 in the West campus of Chubu University. On these days the electricity load of the West

campus of Chubu university exceeded the demand electricity load represented by the dotted red line shown in figure 4.1. This motivated us for the determination of some automatic technique that uses the past data to forecast the electricity load with suitable accuracy. Chubu University is divided into two electricity contract zones i.e. the East Campus zone and the West Campus zone. In this analysis electricity load data of East Campus zone only is used for the analysis.

## 4.2 Literature review

Various machine learning algorithms for short-term and aggregate forecasting of residential electricity consumption for 1 hour and 24 hours ahead residential electricity forecasting is evaluated in [37]. The forecasting accuracy is evaluated using evaluation metrics that are scale independent and robust to a value approaching zero: Normalized root mean square error (NRMSE) and Normalized mean absolute error (NMAE). Classifying the electricity consumption of different households into predefined number of clusters, summing the forecasted value of each cluster's aggregated electricity consumption produces better result than forecasting the electricity consumption of each household individually and summing them. A data processing system to analyzes energy consumption patterns in industrial parks, based on cascade application of Self-Organizing map and the K-means clustering algorithm is presented in [38]. The system is validated in the real load data from an industrial park in Spain. The validation results show that the system adequately finds different behavior patterns that are meaningful and is capable of doing so without supervision and without any prior knowledge about the data. A novel approach for the clustering of electricity consumption behavior dynamics towards large data sets. A density-based clustering technique, CHSFDP, is performed to discover the typical dynamics of electricity consumption and segment customers into different groups is proposed in [39]. A time domain analysis and entropy are conducted on the result of the dynamic clustering to identify the demand response potential of the customers of each group. K-means clustering for time series forecasting in R for predicting the electricity consumption in the residential and industrial sectors of Oman is applied in [40]. Forecasting results of each cluster is analyzed using TSI+RWD, TBATS, ARIMA, etc. and the prediction accuracy is calculated using Mean Absolute Error and Root Mean Squared Error. The TBATS model is found to provide a more accurate result. However, this research was carried with only data for five months, and thus, at least one-year data would be desirable for better forecasting. Dynamic clustering algorithm based on the different criterion such as the season, electricity consumption, and the field where the electricity is consumed is used in [41]. After clustering Deep Belief Network (DBN) based on load demand forecasting is implemented for each cluster and then the multi forecast DBN is implemented

to forecast long-term demand forecasting. Monthly energy consumption forecasting of the year 2005 is performed using data for the year 2004. A comprehensive study of clustering methods for residential electricity demand profile and further applications focused on the creation of more accurate electricity forecast for residential customers is presented in [42]. The determination of an appropriate number of clusters using various distance measures and forecasting results for aggregate models for the entire population is discussed. A method to predict the electricity demand using linear regression (LR) and Support Vector Regression (SVR) is proposed in [43]. The SVR and LR are implemented using weka tool which classifies and clusters the data. Error calculation shows that the proposed method provides improved accuracy as well as improved performances. Clustering for bottom-up short-term load forecasting is used in [44]. The disaggregation strategy uses a nonparametric model to handle forecasting and wavelets to define various notations of similarity between load curves, and it achieves a 16% improvement in forecasting accuracy when applied to French individual consumers.

### 4.3 ARIMA model

Box and Jenkins developed a mathematical model for forecasting a time series by fitting it with the data and using the fitted model for forecasting (i.e. ARIMA model) [45-46]. In general, the ARIMA process is written with the notation ARIMA  $(p, d, q)$ , where  $p$  denotes the number of autoregressive orders in the model. Autoregressive orders specify the previous values from the series are used to predict current values; Difference ( $d$ ) specifies the order of differencing applied to the series before estimating models. Differencing is necessary when trends are present (series with trend are typically nonstationary and ARIMA modeling assuming stationary) and is used to remove their effect. Moving Average ( $q$ ) specifies how deviations from the series mean for previous values are used to predict current values [47]. The parameters  $(p, d, q)$  are always estimated on using the stationary time series which is stationary with respect to its mean and variance.

ARIMA model combines auto regression which fits the current data point to a linear function of some prior data points and moving average adding together several consecutive data points and getting their mean and then using that to compute estimations of the next value. Notation AR( $p$ ) refers to the autoregressive model of order  $p$  and is expressed as [48-49]

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t \quad (7)$$

Where,  $\phi_1, \phi_2, \dots, \phi_p$  are the parameters of the model,  $c$  is a constant and  $\varepsilon_t$  is white noise.

An MA ( $q$ ) model uses past errors as the explanatory variable. The MA ( $q$ ) model is given by

$$x_t = \mu + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (8)$$

where  $\mu$  is the mean of the series and  $\theta_1, \theta_1 \cdots \theta_q$  are the parameters of the model and the  $\varepsilon_t, \varepsilon_{t-q}$  are the white noise error terms. The value of  $q$  is called the order of the MA model. An ARMA consists of two parts, an autoregressive (AR) part, and a moving (MA) part. The model is usually then referred to as the ARMA ( $p, q$ ) model where  $p$  is the order of the autoregressive part and  $q$  is the order of the moving average part.

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (9)$$

An ARIMA ( $p, d, q$ ) model is a generalization of an ARMA model where  $p, d, q$  are non-negative integers that refer to the order of the autoregressive, integrated and moving parts of the model.

To deal with the seasonality of the ARIMA model Box and Jenkins [50] generalized ARIMA model where the seasonal differencing of appropriate order is used to remove non-stationary from the series. A first-order seasonal difference is the difference between an observation and the corresponding observation from the previous year. For a monthly time series  $s=12$ , and for quarterly time series  $s=4$ . The model is generally termed as SARIMA( $p, d, q$ )  $\times$  ( $P, D, Q$ )<sup>s</sup> model.

#### 4.4 Target of the research

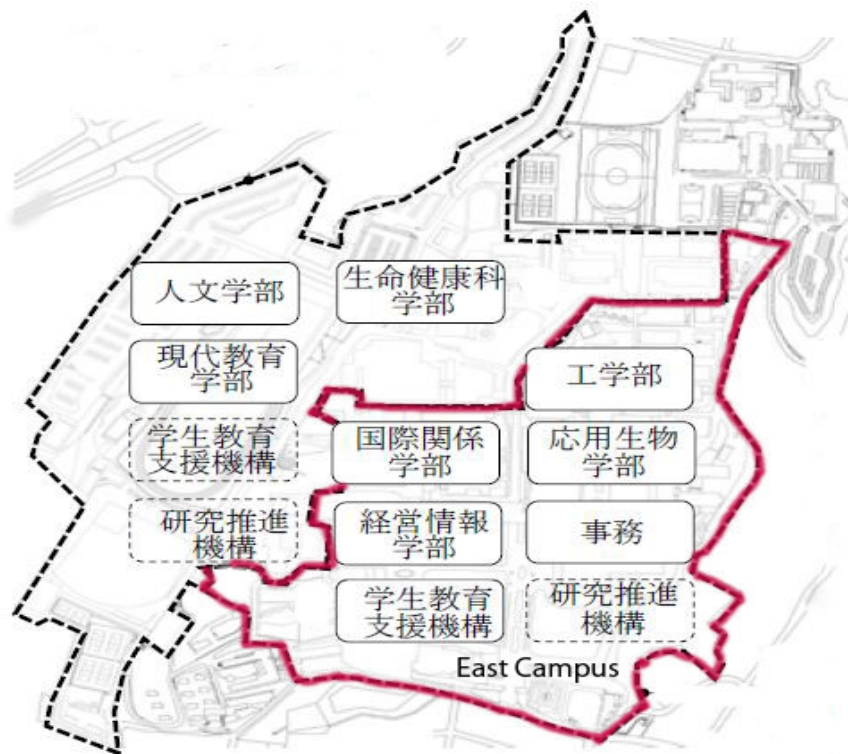
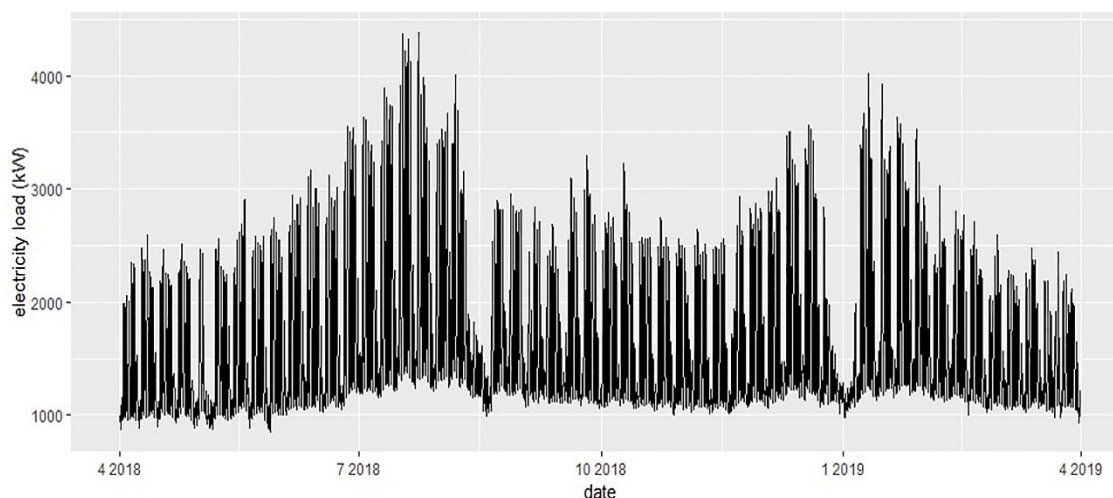


Figure 4.3 Map of Chubu University with East-campus in red boundaries



Figure 4.3 represents the map of Chubu university with the red boundaries indicating the boundaries of East-campus, the target of this research.



**Figure 4.4** Line graph of whole university electricity load of academic year 2018

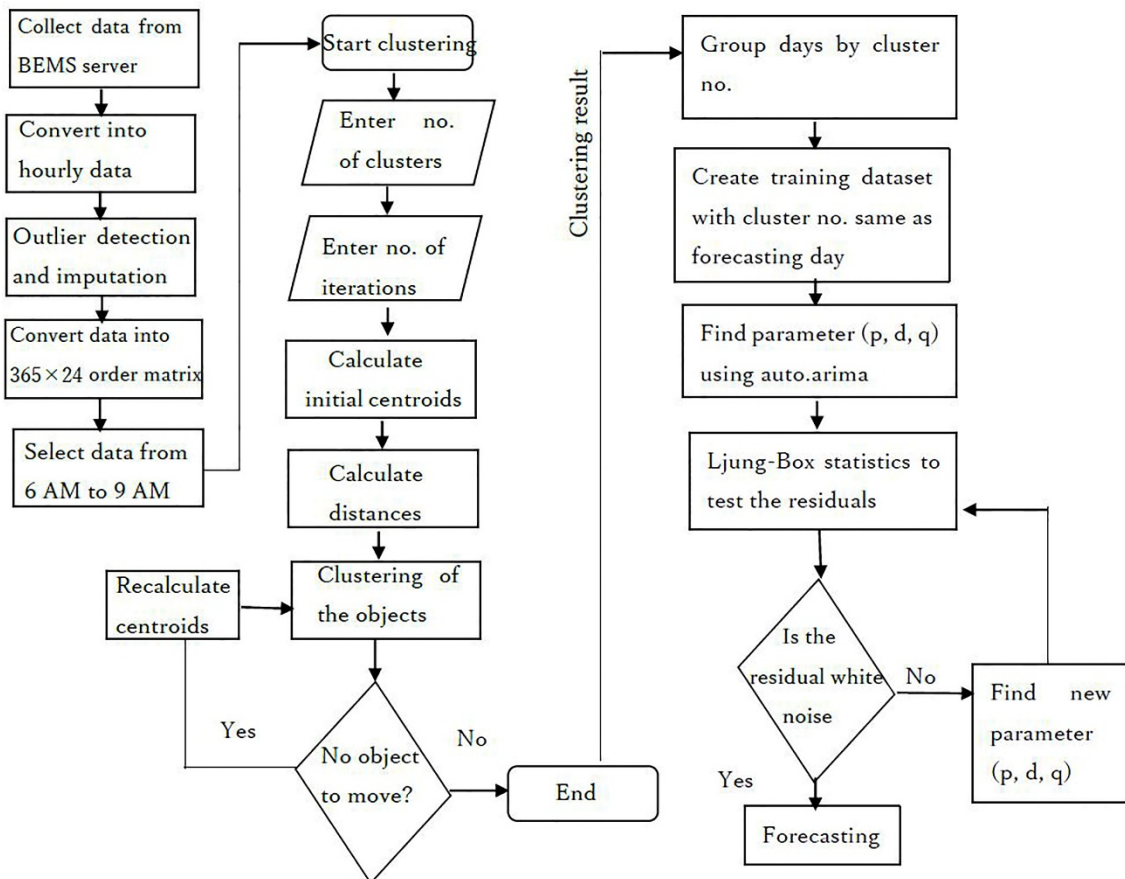
As seen in figure 4.4, the electric load of Chubu university changes over the year, with the highest electricity load in air-conditioning usage period in the summer and winter season. During vacations, electricity is found to fall low, near to the base electricity load of the university. Throughout the week too, there is a rise and fall in the electricity load due to the effect of lectures, partial lectures (Saturdays), holidays (Sundays and public holidays), occurrences of events, etc. Thus there is a presence of both seasonal and weekly trend.

#### **4.5 Methodology**

Data used in this research are the hourly electricity load data of the East Campus of Chubu university of the academic year 2017 and 2018. Chubu university has a Building Energy Management System (BEMS) where electricity load data are measured at the transformer of each building, and the data is collected in the BEMS server every minute. The minute data are summed to produce hourly data, and thus, one-year data consists of 8760 data for each building. Due to technical problems, data collected in the BEMS server contain outliers and missing values. The presence of outliers in the training data can lead to errors in the forecasting output. Thus, in this research the outliers present in the raw data are converted into not available (NA) values as in [51] and were imputed with appropriate values using linear interpolation using the zoo package of R programming language.

Data free from missing values and outliers are then analyzed using K-means clustering method. As in [51], the number of cluster = 6 is found to be appropriate and one whole year electricity load data are classified into 6 clusters. The main objective of this analysis is to

forecast the peak energy consumption of university buildings and develop an automatic process that receives data from the BEMS sever and forecast the peak load so that some strategy can be implemented for peak load reduction. The electricity load in university buildings depends upon different factors such as presence or absence of holidays, presence or absence of lectures, use or no use of air-conditioning, the occurrence of events, and use or no use of experimental facilities. Thus the energy load is changing day to day. Forecasting of electricity load without considering above factors can cause a high error in the forecasting result. Thus, to make an automatic system for data processing, we created a hybrid model with clustering and ARIMA model, that automatically clusters data so that the cluster number of the forecasting day is known. In the case of ARIMA model, the selection is made automatically using the auto.arima function of R programming language. The hybrid model combining clustering and ARIMA not only make the process automatic but also improves the accuracy of the electricity load in university buildings.



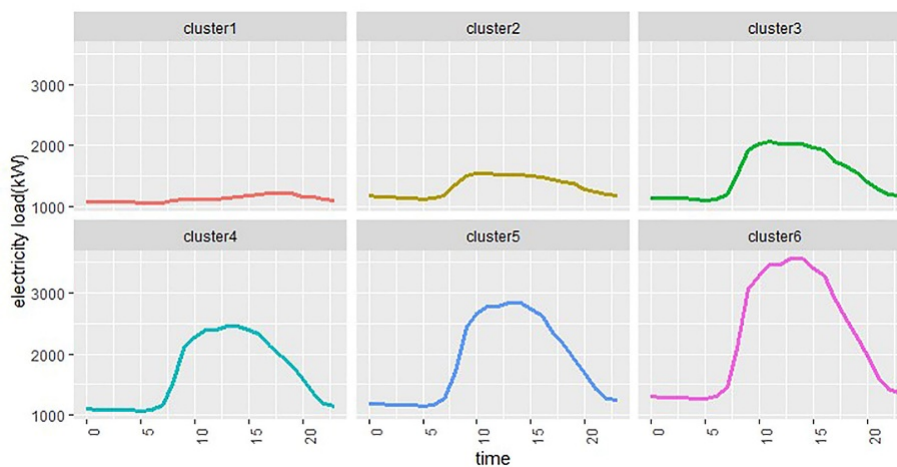
**Figure 4.5** Flow chart of Proposed method

The process from data collection to electricity load forecasting is shown in the flow chart

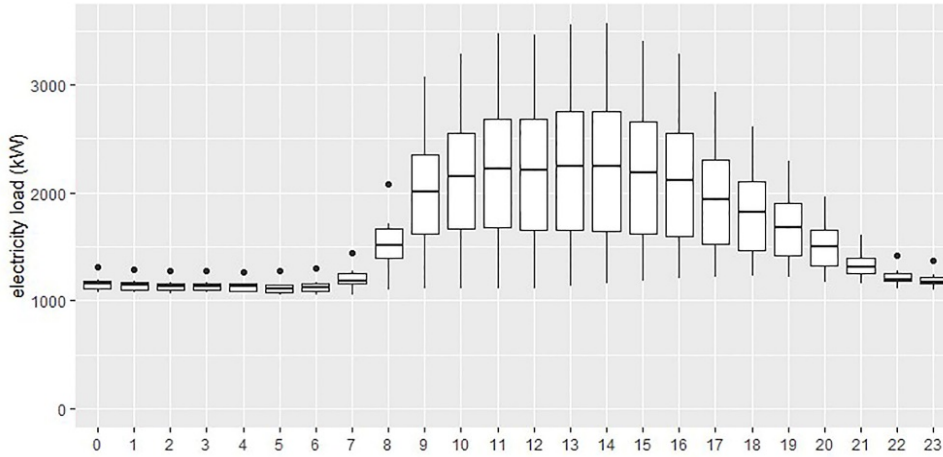
in figure 4.5 The data used in this paper are the building energy data of the East campus of Chubu University. The data is collected from the BEMS server are 30-minute interval data which is converted into hourly data. After outlier detection and imputation of missing values, one-year data (365 days) including the forecasting day until 9 AM are created. These data are converted into  $365 \times 24$  order matrix, out of which electricity load data from 6 AM to 9 AM from all 365 days including the forecasting day until 9 AM are extracted, on which clustering is performed. The data from 6 AM to 9 AM of the forecasting day is used only for clustering, and not for electricity load forecasting. In this study, six clusters were found to be suitable and the initial centroids are calculated using the percentile method. The clustering result classifies all 365 days into 6 clusters with days that have the similar electricity load characteristics. Then, a dataset with days belonging to the same cluster as the forecasting day is created, which is used as the training data for the forecasting model. The parameters  $(p, d, q)$  of the best model for the training dataset is calculated using the auto.arima function of R programming language. The next step is the testing of the residuals of the model using Ljung-Box statistics; if the residuals is a white noise, then the model is ready for forecasting, otherwise selection of the new model is required.

The electricity load data of each day from 6 AM to 9 AM is used for clustering the day of which the electricity load forecasting is expected. Here, the data from 6 AM to 9 AM is used only for clustering the day on which the forecasting is done, not for electricity load forecasting. The combination of clustering and ARIMA model have been used before as discussed in the literature review section but most of them are related to clustering the group of customers and finding the aggregate forecasting of individual clusters. This concept of clustering the forecasting day has never been discussed in any other journals before.

Using K-means clustering we classified the electricity load of the university into 6 clusters based on the hourly distribution of one-year electricity load.



**Figure 4.6** Electricity load of cluster center 1 to 6



**Figure 4.7** Electricity load distribution of cluster centers using box plot

Figure 4.6 is the electricity load plot of each cluster and figure 4.7 is the box plot showing the electricity load distribution of Chubu university per cluster and hours of the day respectively. A large difference in the energy load from cluster 1 to cluster 6 is found, even the base energy is found to increase with a rise in the cluster number. The electric load from 10 PM till 6 AM is almost similar. From 8 AM, the electric load is found to rise sharply until it reaches its peak around 1 or 2 PM. As seen in figure 4.6, the electricity load peak for each of the cluster is different, and thus by clustering the forecasting day and using its cluster members for electricity load prediction can improve the peak load forecasting accuracy of any forecasting algorithm.

Many comparative studies have been conducted with the aim of identifying the most accurate methods for time series forecasting [52]. However, research findings indicate that the performance of forecasting methods varies according to the accuracy measure being used. A good accuracy measure should provide information on the clear summary of the error distribution. Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE) and Mean Absolute Error (MAE) are the very early and most popular accuracy measures. In this paper, we use these accuracy measures for the forecasting results.

RMSE and MAE are scale-dependent measures since their values depend on the scale of the data. They are useful in comparing forecasting methods on the same data. For 24 hours ahead forecasting, if “ $e_t$ ” represents the error in forecasting for each hour and “ $Y_t$ ” represents the actual values, RMSE and MAE can be defined as

$$MAE = \frac{1}{n} \sum_{t=1}^n |e_t| \quad (10)$$

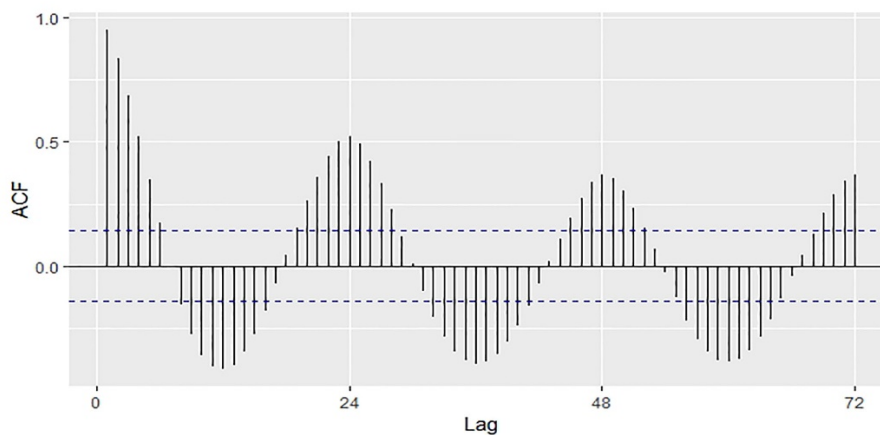
$$MSE = \frac{1}{n} \sum_{t=1}^n e_t^2 \text{ and } RMSE = \sqrt{MSE} \quad (11)$$

MAPE is based on percentage error of the observation and is scale-independent.

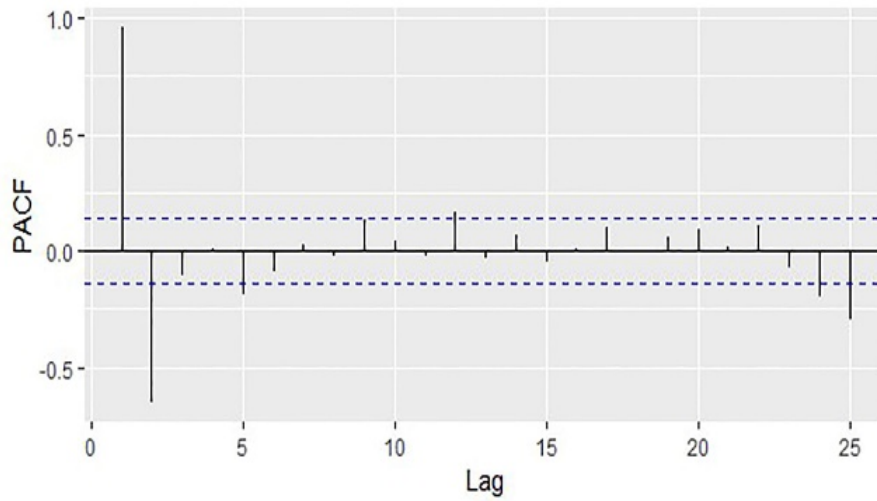
$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^n \frac{|e_t|}{|Y_t|} \quad (12)$$

#### 4.6 Fitting of ARIMA Model

The Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) are used to estimate the value of parameter  $p$  and  $q$  in the ARIMA model. Figures 4.8 and 4.9 shows the ACF and PACF of electricity load data of the training dataset for electricity load forecasting on 2<sup>nd</sup> April 2018. Based on the first two lags, it might be possible that AR (2) works based on the first two spikes in the PACF. To identify stationary/non-stationary processes of the time-series, we use Augmented Dickey-Fuller test (ADF). The null-hypothesis of an ADF test is that the data is non-stationary, and small p-values suggest stationary. On using the ADF test using `adf.test` function of R programming language, the series was found stationary. Thus the parameter  $d$  in the ARIMA model can be chosen to be zero. The pattern in ACF indicates the presence of seasonal behavior. To check the seasonal behavior, we check the situation on lags around 24 and 48. When data are seasonal, the autocorrelation will be larger for the seasonal lags (at multiple of the seasonal frequencies) than the other lags. In this paper hourly electricity load data is used. In the case of University data, electricity load at a certain time is likely to be similar to the electricity load at the same time in the previous days. Thus the presence of larger spikes at 24 and 48 in ACF indicates the data has seasonality. In ACF and PACF there is a presence of spikes at 24 indicating the data has seasonality.



**Figure 4.8** ACF of the training data



**Figure 4.9** PACF of the training data

To find the best model for this data, we fit different models and select the model with the minimum Akaike Information Criteria (AIC) value. AIC is an estimator of the relative quality of statistical models for a given data set. The selection of the model is important, as under-fitting a model may not capture the true nature of the variability in the outcome variable, while an over-fitted model loses generality. AIC is then a way to select the models that best balances these drawbacks and a smaller value of AIC represents a better model [53].

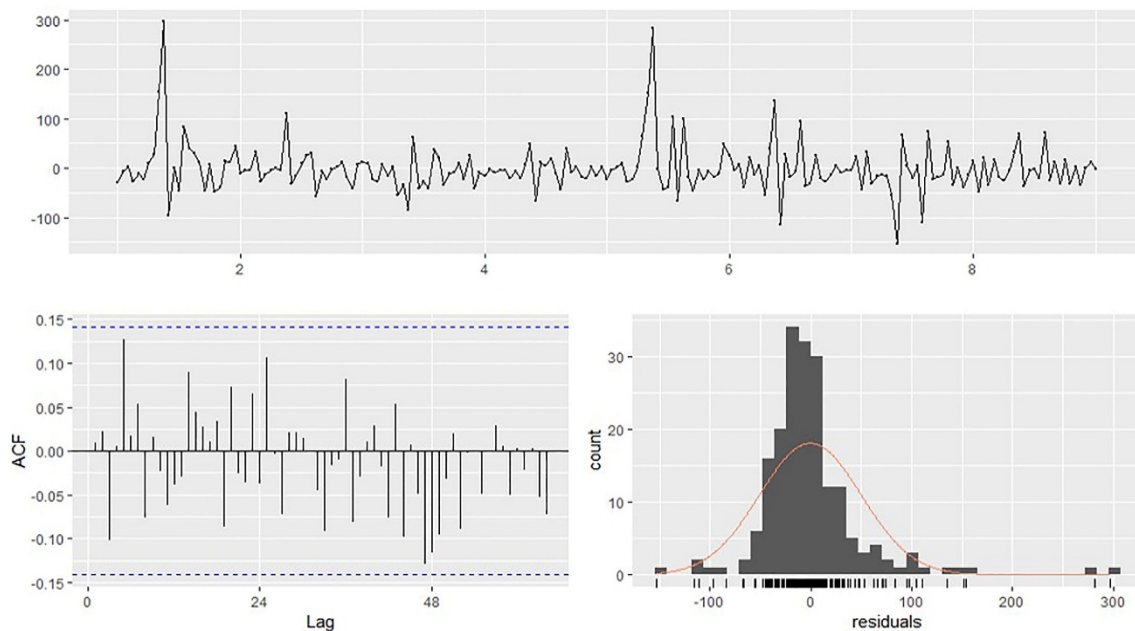
**Table 4.1** AIC values of suggested ARIMA models

ARIMA $(p, d, q)(P, D, Q)_s$	AIC
ARIMA(2,0,0)(0,0,0) <sub>24</sub>	2105.4
ARIMA(2,0,1)(2,0,0) <sub>24</sub>	2104.5
ARIMA(2,0,1)(2,0,1) <sub>24</sub>	2096.4
ARIMA(2,0,2)(2,0,1) <sub>24</sub>	2098.6
ARIMA(2,1,1)(2,0,1) <sub>24</sub>	2097.6
ARIMA(3,0,1)(2,0,0) <sub>24</sub>	2105.6
ARIMA(3,0,1)(2,0,1) <sub>24</sub>	2097.8
ARIMA(3,0,1)(3,0,1) <sub>24</sub>	2107.9

In table 4.4 the smallest value of AIC is obtained for seasonal ARIMA with the non-seasonal part  $(p, d, q) = (2, 0, 1)$  and the seasonal part of the model  $(P, D, Q) = (2, 0, 1)$ . Then, we



calculated the parameters of the best model for the training data using the `auto.arima` function of R programming language. The best model produced by `auto.arima` was “ARIMA(2,0,1) (2,0,1)<sub>24</sub>” with AIC value 2096.4, same as the least value of AIC in the table.



**Figure 4.10** Residual analysis for the ARIMA (2, 0, 1) (2, 0, 1)<sub>24</sub> fit to the data

Figure 4.10 shows the plot of residuals, its ACF values, and the histogram. As the ACF of residuals are within the significance level, the model selected using `auto.arima` is best fit for the training dataset. The objective of this research is not only to increase forecasting accuracy but also to make the forecasting process automatic. As the days pass on new data are added into the system and these data will be used in the forecasting in the future days. Thus, the training data used for forecasting goes on changing with time. Appropriate selection of these parameters ( $p, d, q$ ) can only provide accurate forecasting results. Thus, we decided to select the best model using `auto.arima` functions for other training datasets. However, we use the Ljung-Box test for confirming whether the residuals are white noise or not.

Some other reasons for making the forecasting automatic includes

- 1) Fitting a model is not a simple task and most users are not expert at fitting time series models
- 2) Automatic algorithms can produce better models than many experts.
- 3) Many businesses need hundreds of forecasts per week/month and they need it fast.
- 4) Some multivariate forecasting methods depends on many univariate forecasts.

#### 4.7 Results

To carry out the proposed algorithm, it is necessary to determine the cluster number of the forecasting day. For this purpose, it is essential to select the dataset according to the forecasting day. In this research, data for two years from the 1<sup>st</sup> of April 2017 to the 31<sup>st</sup> of March 2019 is used. If the forecasting of electricity load on the 1<sup>st</sup> of March 2018 is expected, then a one-year data starting from 2<sup>nd</sup> of March 2018 to 1<sup>st</sup> of March 2019 is created. The electricity load data from 6 AM to 9 AM of whole one year including the forecasting day is extracted. Then, K-means clustering is done on the extracted data that is converted into a  $365 \times 4$  order matrix by extracting 6 to 9 AM data from 365 days. The initial centroids required for K-means clustering is determined using the percentile method. The entire one-year data are classified into six clusters; and each cluster representing the group of days having similar electricity load characteristics.

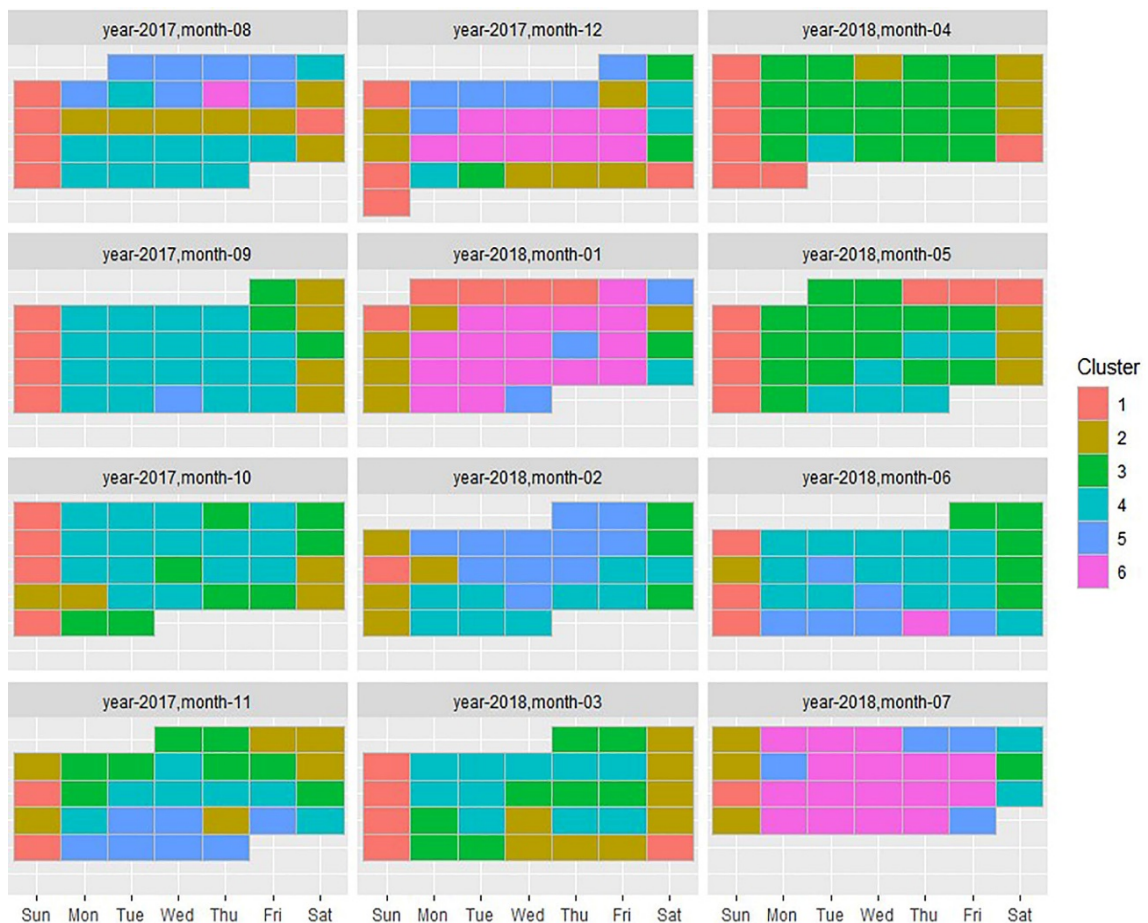
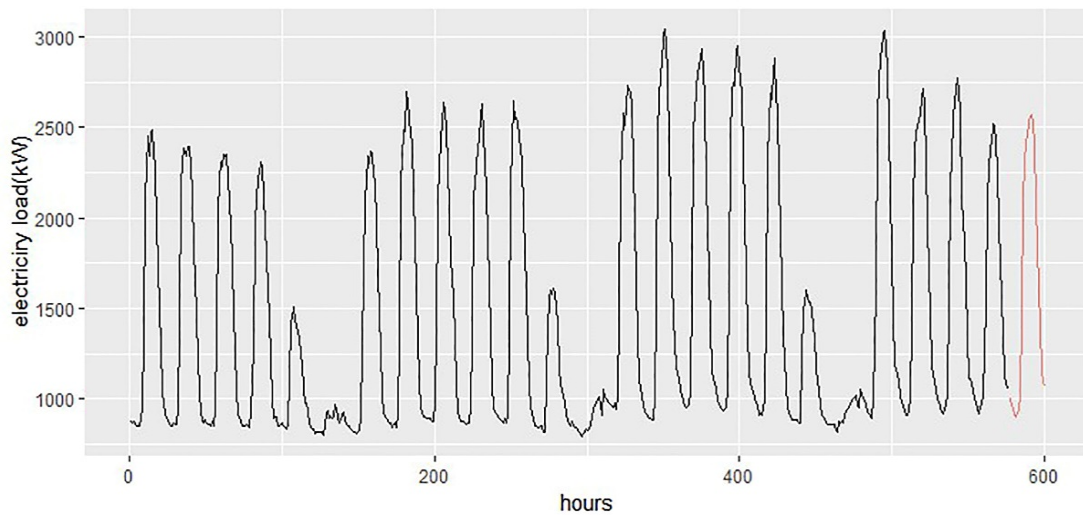


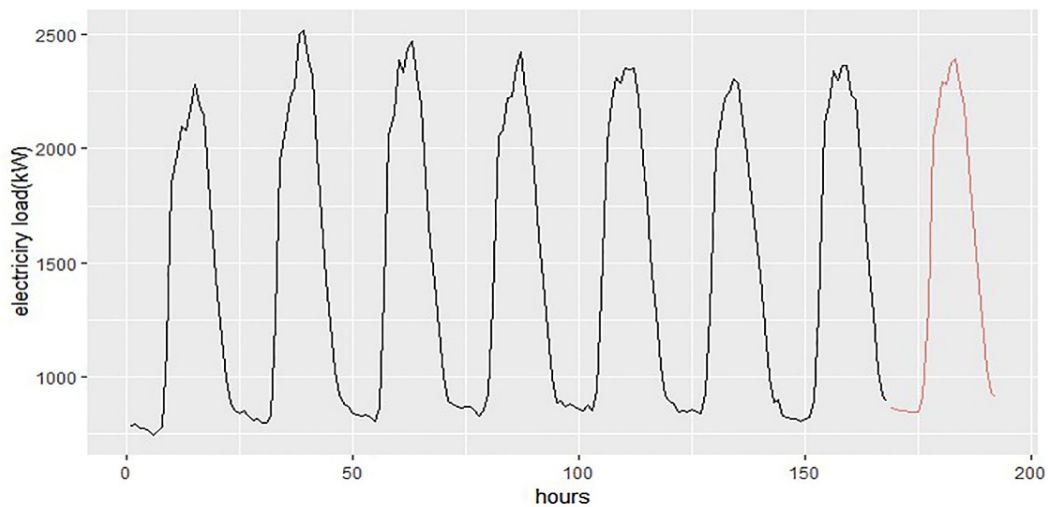
Figure 4.11 Calendar plot of clustering result with number of clusters  $k = 6$



Calendar plot of figure 4.11 is the clustering result of the one-year dataset including the forecasting day. The forecasting day, 27 July 2018 is assigned to cluster 5. The purpose of this research is to forecast the electricity load with the purpose of reducing the peak energy load in buildings. In the case of ARIMA forecasting, the training data is chosen a few days before the forecasting days. Thus, a variety of days with different electricity load pattern are mixed in the training data, which are used to forecast the electricity load of the next day, which may produce forecasting with high errors. In the case of the proposed method, the training data are selected using the clustering method. This makes the member of the training data have similar electricity load characteristics as the forecasting day and hence increasing the accuracy of the forecasting result.



**Figure 4.12.** Forecasting result ARIMA model

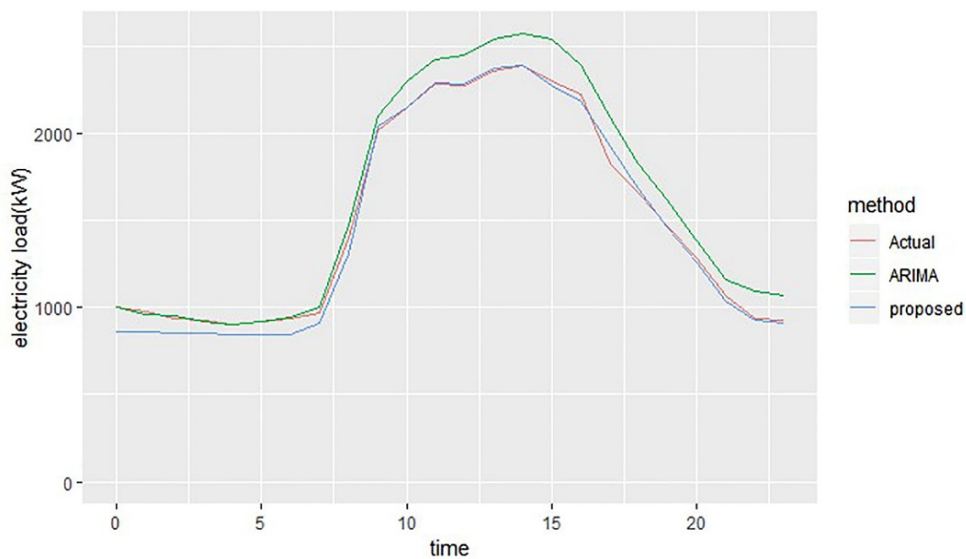


**Figure 4.13.** Forecasting result proposed method

Figure 4.12 and 4.13 are the forecasting results of the ARIMA model and proposed method respectively on 27 July 2018.

Black lines represent the training datasets whereas the forecasting result is indicated by red color. 24 preceding days are used as the training dataset in the ARIMA model whereas a group of days belonging to the same cluster as the forecasting day is used in the proposed method. The training dataset for ARIMA model consists of 24 days (576 hours), whereas the training data for the proposed method consists of 7 days (168 hours). The best model for both ARIMA and proposed method is calculated using the `auto.arima` function in R programming language. The computational time of the ARIMA method and the proposed method was measured using `system.time` function of the R programming language using computer with a processor, Intel(R) Core(TM)i7-6700 CPU @3.4GHz and 16 GB RAM. The elapsed time for the ARIMA model was 9.30 second whereas the elapsed time for the proposed method was significantly lower at 0.45 second.

The comparison between forecasting result of ARIMA model and proposed method using East campus data is shown in figure 4.14.



**Figure 4.14** Forecasting result of ARIMA and proposed method in comparison to actual data on 27 July 2018.

**Table 4.2 Forecasting result of suggested ARIMA model and proposed method**

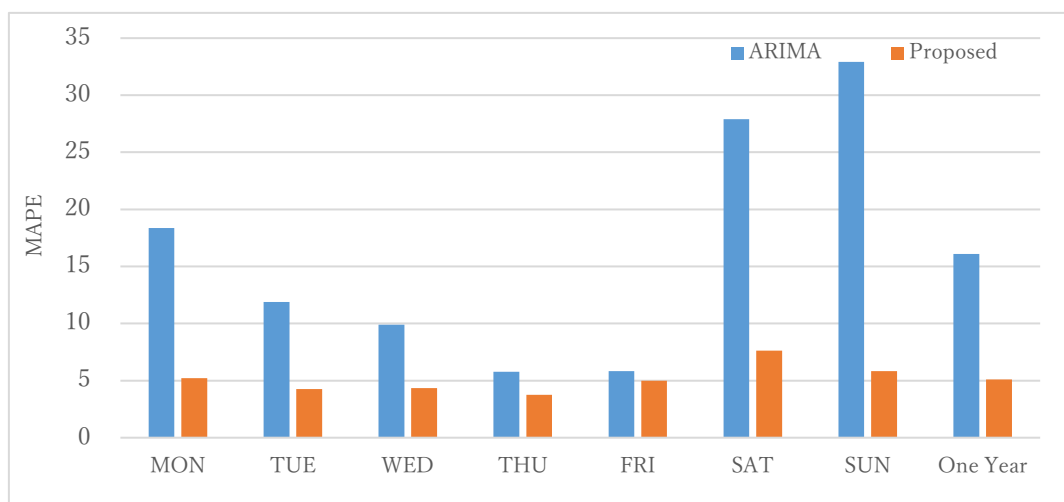
time	Actual data	Forecast using ARIMA model	Forecast using Proposed method	Error ARIMA model	Error Proposed method
t	$A_t$	$F_A$	$F_p$	$E_A$	$E_p$
0:00	1001	1003.2	866.8	-2.2	134.2
1:00	979	962.8	859.6	16.2	119.4
2:00	933	953.0	855.6	-20.0	77.4
3:00	924	915.8	851.2	8.2	72.8
4:00	905	901.6	848.7	3.4	56.3
5:00	917	916.7	846.4	0.3	70.6
6:00	936	944.2	845.6	-8.2	90.4
7:00	967	1006.5	913.9	-39.5	53.1
8:00	1404	1473.2	1310.2	-69.2	93.8
9:00	2016	2102.6	2044.0	-86.6	-28.0
10:00	2148	2302.0	2155.2	-154.0	-7.2
11:00	2285	2426.8	2293.5	-141.8	-8.5
12:00	2275	2446.9	2283.6	-171.9	-8.6
13:00	2357	2541.3	2376.6	-184.3	-19.6
14:00	2390	2574.1	2394.8	-184.1	-4.8
15:00	2300	2541.2	2274.9	-241.2	25.1
16:00	2224	2394.8	2184.6	-170.8	39.4
17:00	1827	2100.8	1926.1	-273.8	-99.1
18:00	1658	1830.4	1688.6	-172.4	-30.6
19:00	1471	1622.4	1461.7	-151.4	9.3
20:00	1289	1391.0	1258.2	-102.0	30.8
21:00	1071	1159.3	1039.9	-88.3	31.1
22:00	938	1098.3	927.1	-160.3	10.9
23:00	930	1073.5	910.1	-143.5	19.9
<b>MAE</b>		<b>108</b>	<b>47.5</b>		
<b>RMSE</b>		<b>134.6</b>	<b>60.7</b>		
<b>MAPE</b>		<b>6.6</b>	<b>4.2</b>		

From table 4.2, the values of MAE, RMSE, and MAPE of the proposed method is found to be small in comparison to that of the ARIMA model. Results from figure 4.14. and the values of MAE, RMSE, and MAPE of the proposed method shows that the proposed method produces better results than the ARIMA model.

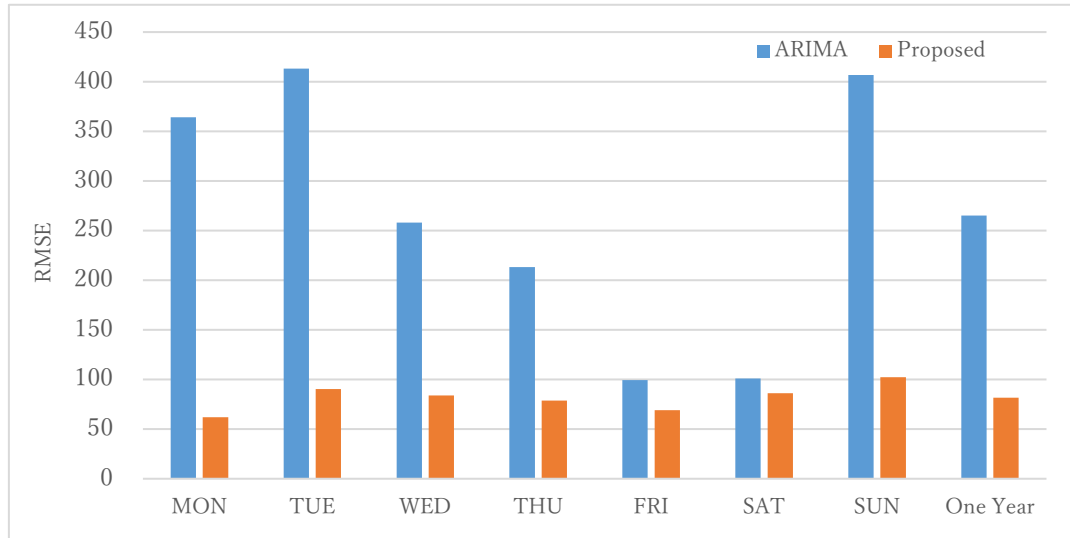
The electricity load forecasting of each day using the ARIMA model and the proposed method from 1<sup>st</sup> of April 2018 till 31<sup>st</sup> of March 2019 was performed. The MAPE, RMSE and MAE of each day between the actual data and the forecasted data using the ARIMA model and proposed model was calculated. Table 4.3 shows the one-year average value of MAPE, RMSE, and MAE between the actual and the forecasted value using the ARIMA model and the proposed method by day of the week.

**Table 4.3 Error values of forecasting result by ARIMA model and proposed method**

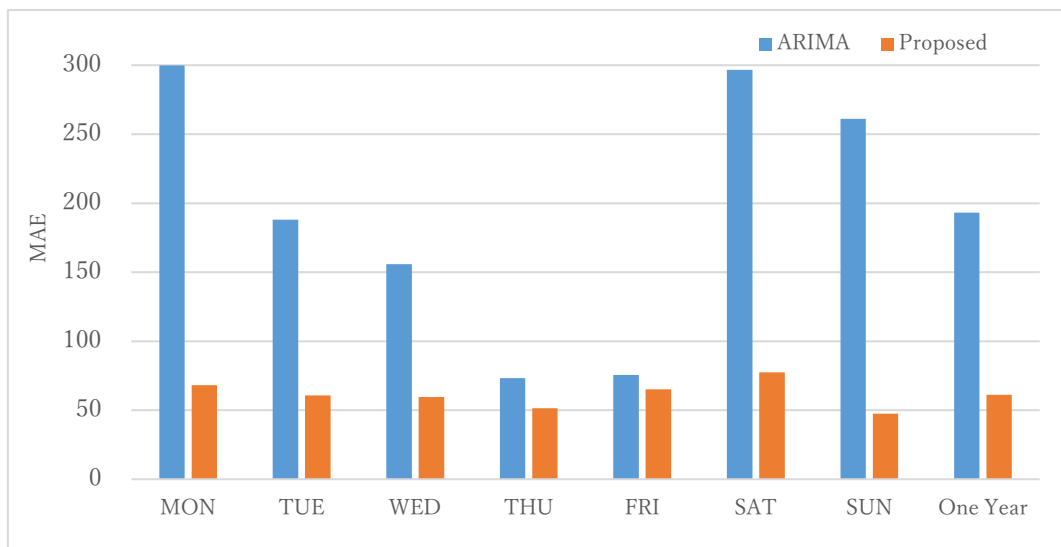
	ARIMA model			Proposed method		
	MAPE	RMSE	MAE	MAPE	RMSE	MAE
SUN	32.9	364	261.2	5.8	61.9	47.6
MON	18.4	413.1	299.9	5.2	90.4	68.1
TUE	11.9	258	188	4.3	83.8	60.7
WED	9.9	213.4	155.8	4.3	78.7	59.5
THU	5.8	99.4	73.2	3.8	69.0	51.4
FRI	5.8	100.9	75.6	5.0	86.2	65.2
SAT	27.9	406.8	296.7	7.6	102.4	77.4
One Year	16.1	265.3	193.1	5.1	81.7	61.3



**Figure 4.15 MAPE value for ARIMA and Proposed method**

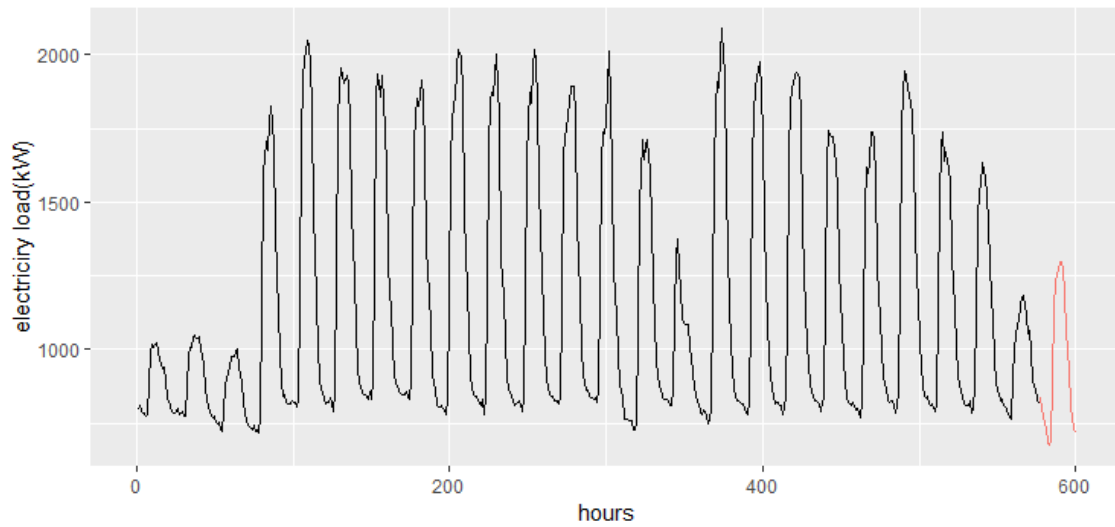


**Figure 4.16.** RMSE value for ARIMA and Proposed method

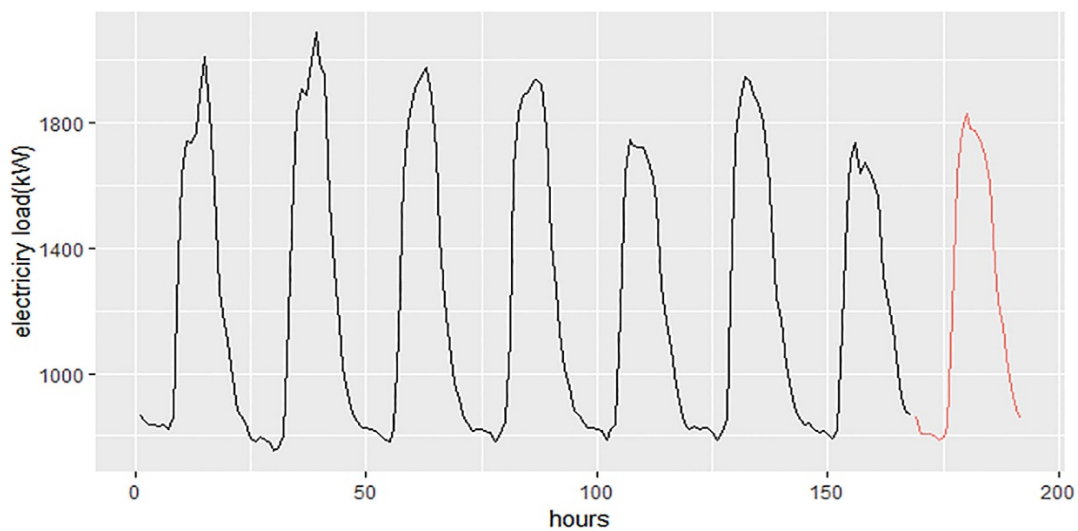


**Figure 4.17.** MAE value for ARIMA and Proposed method

Figures 4.15, 4.16 and 4.17 represent the bar graph showing the values of MAPE, RMSE, and MAE for the ARIMA and proposed method, respectively. The value of MAPE, RMSE, and MAE for ARIMA model is quite high at Saturday, Sunday, Monday and Tuesday in comparison to the proposed method. As ARIMA model forecasts based on the electricity load of the previous days, sudden rise and fall of electricity load before the forecasting days which can occur due to the holiday (Sunday), partial holiday (Saturday) and the occurrence of events like campus festivals, open campus, sports events etc. can affect on the forecasting accuracy.



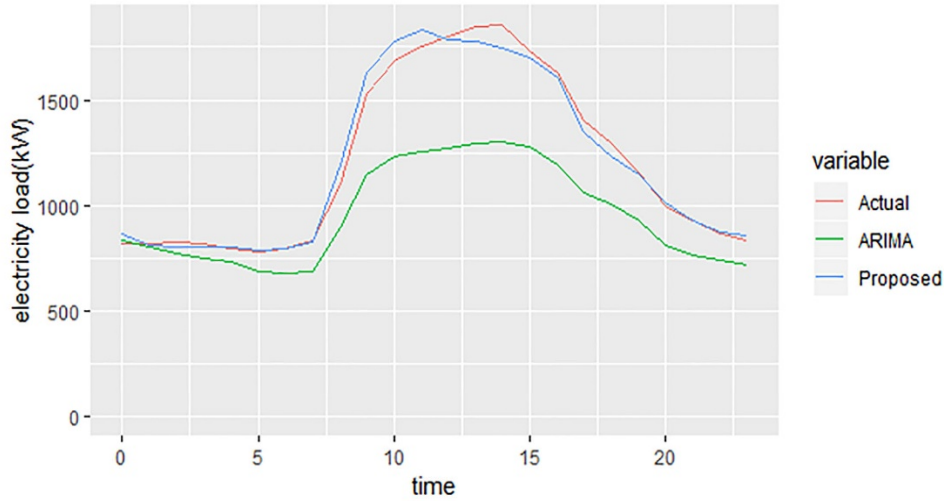
**Figure 4.18.** Forecasting using ARIMA model



**Figure 4.19.** Forecasting using proposed model

Figures 4.18 and 4.19 are the one day ahead forecasting result after removing Saturday and Sunday from the dataset. Figures 4.18 and 4.19 shows the forecasting results of ARIMA model and proposed method respectively on 18<sup>th</sup> September 2018. To remove the effect of the holiday (Sunday) and partial holiday (Saturday), a dataset was created removing all Sunday and Saturday data from the academic year data 2018. Although there is no presence of Saturday and Sundays on this data, fluctuations in the electricity load data can be found. The presence of reduced electricity load on 17<sup>th</sup> September (a day before forecasting) led to a considerable error in the forecasting result on 18<sup>th</sup> September as shown in figure 4.18. In case of proposed method, the presence of increased or reduced electricity load a day or few days

before the forecasting day does not affect the forecasting result because the training data of the proposed method are the days belonging to the same cluster group with similar electricity load characteristics as the forecasting day.



**Figure 4.20** Forecasting result of ARIMA and proposed method after removing Saturday and Sunday on the training data.

The MAPE, RMSE, and MAE values between the actual and forecasted values when Saturday and Sunday are removed using the ARIMA model are 18.5, 311.5, 253.7 respectively. Whereas, the MAPE, RMSE, and MAE values between the actual and forecasted values using the proposed method is significantly lower at 2.7, 50.5 and 36.4 respectively.

Grouping by the day of the week can be a slightly better option than using the only random days for forecasting. For example, to forecast the energy load on 18<sup>th</sup> January 2018 (Thursday), the training data for ARIMA model can be constructed by collecting previous days belonging to the same day of the week (Friday), as energy load is likely to be similar on the same day of the week, rather than another day of the week. In figure 4.11, the forecasting day 18<sup>th</sup> January 2018 belongs to cluster 5. The days belonging to the same day of the week (Thursday) on the preceding weeks are found to belong to clusters 6, 2, and 1. This will cause deviation in the forecasting values in the forecasting day.

Exclusion of holidays and grouping by the days of the week can improve the accuracy of forecasting when done manually; however the accuracy is less in comparison to the proposed model because the exclusion of holidays and grouping by the day of the week cannot ensure that the training data have similar characteristics as in the case of clustering. Thus, clustering is the best method to group the days on the basis of their similar electricity load characteristics.

#### **4.8 Discussion**

In this paper, we proposed a method to increase the performance of the ARIMA model using clustering technique and forecasting the electricity load automatically. Using proposed method, the analyst doesn't need to think about the energy load pattern in the previous days (e.g. Holidays, occurrences of events, etc.) as the training data in the proposed method is created using the group members of belonging to the same cluster as the forecasting day.

The proposed method can be useful for the conservation of energy in buildings and demand response. The electricity power company usually apply higher rates to their customers in the day time than in the night. Higher consumption of electricity in the electricity peak hours in the day can lead to higher electricity bills of the customers. Forecasting of electricity load before reaching the peak electricity load can provide the management authorities, faculty staff, and students with sufficient time to make a strategy for reduction of peak electricity load. For example, the authorities can decide the time of discharging the battery, time for using cogeneration, etc. before the peak hours and faculty staff and students can take actions for reducing electricity consumption if they are informed about the forecasting. Moreover, this method can be useful in the demand response for reducing the electricity bills by avoiding electricity usage during the high electricity rate hours. Thus forecasting of electricity load can be helpful in energy conservation in buildings. The measures for peak load reduction differs from buildings to buildings. Some general measures that can be adopted for peak load reduction in university buildings are discussed below.

#### **4.9 Measures for electricity peak reduction**

- In summer raising the air-conditioning setting temperature can help a lot for peak electricity load reduction. In Chubu University the air-conditioning setting temperature is usually done at 26 degrees. Hence, by raising the setting temperature from 26 to 28 degree can save at least 10% of the electricity load.
- Turning off the unnecessary lighting. Opening the blinds of the window and letting the natural light to enter the room.
- Using the electricity stored in the battery charged during off-peak hours and the electricity generated by PV during the electricity load peak hours.
- Turning off the lights on the lecture room strictly before and after the lectures.
- Turning off the computers that are not being used.
- Requesting the researchers to conduct the experiments that consume lots of electricity beyond the peak load hours (i.e. beyond 10 AM – 3 PM).



#### **4.10 Measures for base energy reduction**

Base energy includes that equipment that operates 24 hours and all the year-round. Some examples of refrigerators, servers, emergency lamps, ventilation fans, vending machines, sensors, etc. Some measures for base energy reduction is listed below

- Raising the setting temperature of refrigerators without affecting the quality of contents within it.
- Requesting to turn off the experimental equipment during the peak load hours if possible.
- Replacing emergency light and another lighting with energy-efficient LED lamps.
- Sleep mode of computers, printers, and other equipment when not in use should be turned off.
- Lights of the vending machine should be turn off at the peak load hours during the day.
- Motivating the common use of electric hot water pots, refrigerators, etc. rather than individual use or use per rooms.

# Chapter 5

## Data analysis and visualization using Shiny

### 5.1 R programming language

R is a language and environment for statistical computing and graphics. R provides a wide variety of statistical (linear and non-linear modelling, classical statistical tests, time series analysis, classification, clustering etc.) and graphical techniques, and is highly extensible. One of R's strength is the ease with which well-designed publication-quality plots can be produced, including mathematical symbols and formulae where needed. Great care has been taken over the default for the minor design choices in graphics, but the user retains full control. R is available as free software under the terms of the Free Software Foundation's GNU General Public License in source code form. It compiles and runs on a wide variety of UNIX platforms and similar systems, Windows and MacOS.

Although used mainly by statisticians and other practitioners requiring an environment for statistical and software development, R can also operate as a general matrix calculation toolbox with performances benchmarks comparable to GNU Octave or MATHLAB.

The capabilities of R are extended through user-created packages, which allow specialized techniques, graphical devices, import/export capabilities, reporting tools, etc. These packages are developed primarily in R, and sometimes in Java, C, C++ and Fortran. The R packaging system are used by researchers to create compendia to organize research data, code and report files in a systematic way for sharing and public archiving. The core set of packages is included with the installation of R, with more than 15,000 additional packages and other repositories.

### 5.2 Shiny

Shiny is an R package that makes it easy to build interactive web apps straight from R. We can host standalone apps on a webpage or embed them in R Markdown documents or build dashboards. We can also extend your Shiny apps with CSS themes, htmlwidgets, and JavaScript actions.



**Figure 5.1** Example of graphical visualization of data in the form of histogram using Shiny

Figure 5.1 shows an example plot of a histogram of a dataset called “faithful” with a configurable number of bins. Users can change the number of bins with a slider bar, and the app will respond to their input.

### 5.2.1 Structure of the Shiny App

Shiny apps are contained in a single script called `app.R`. The script `app.R` lives in a directory (for example, `newdir/`) and the app can be run with `runApp(“newdir”)`. `app.R` has three components:

- a user interface object
- a server function
- a call to the `shinyApp` function

The user interface (`ui`) object controls the layout and appearance of the app. The server function contains the instructions that our computer needs to build our app. Finally, the `shinyApp` function creates Shiny app objects from an explicit UI/server pair. In the newer version of Shiny, the `ui` object and server are supported in a single file which makes it possible to copy and paste the entire app into the R console, which makes it easy to quickly share the code to others to experiment with.

To prepare ui for the Shiny as shown in the figure 5.1, we require following codes

```
# Define UI for the app that draws a histogram
```

```
ui <- fluidpage (
```

```
# Add App title
```

```
titlePanel ("Hello Shiny!"),
```

```
# Sidebar layout with input and output definations
```

```
sidebarLayout(
```

```
#Sidebar panel for inputs
```

```
  SidebarPanel (
```

```
# Input: Slider for the number of bins
```

```
sliderInput (inputId = "bins",
```

```
             label = "Number of bins:",
```

```
             min = 1,
```

```
             max = 50,
```

```
             value = 30)
```

```
),
```

```
# Main panel for displaying output
```

```
mainPanel (
```

```
#Output: Histogram
```

```
plotOutput (outputId = "disPlot")
```

```
)
```

```
)
```

```
)
```

### **server**

The server function can be written as

```
server <- function (input, output) {
```

```
  outputdistPlot <- renderPlot ({
```

```
    x <- faithful$waiting
```

```
    bins <- seq(min(x), max(x), length.out = input$bins + 1)
```

```
    hist(x, breaks = bins, col = "*75AADB", border = "white",
```

```

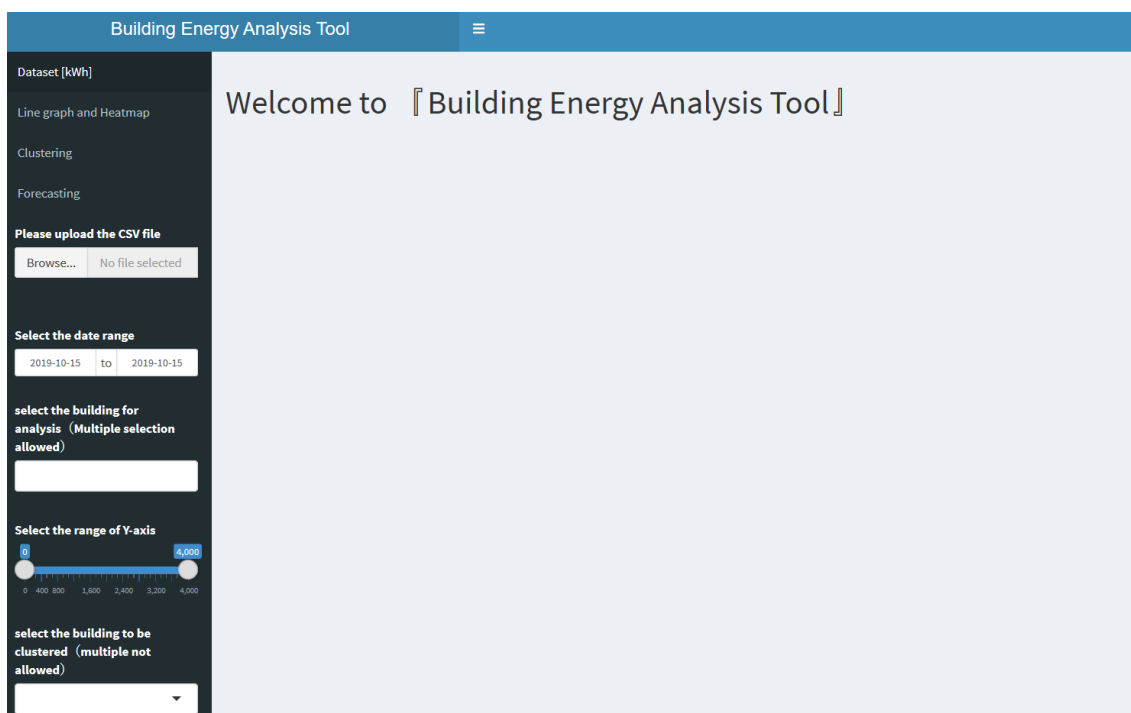
    xlab = "waiting time to next eruption (in mins)",
    main = " Histogram of waiting times")
  })
}

```

The script does some calculations and then plots a histogram with the requested number of bins. We can see that most of the script is wrapped in a call to renderPlot.

### 5.3 Shiny as a visualization platform

In this research we use Shiny for the visualization of the analyzed result. The first part of the data visualization and analysis process using Shiny is the uploading of the data that is to be analyzed. This can be done by clicking the Browser option in the in the “Please upload the CSV file” menu. On opening the Browser button a dialog box appears through which we can select the file we want to upload. Figure 5.2 is an image of Shiny Building Analysis tool.



**Figure 5.2** Image of Shiny Building Energy Analysis tool

As shown in the above figure 5.2, the main components of visualization and data analysis includes “Line graph and Heatmap”, “Clustering” and “Forecasting”. All these menus are in the slider bar on the left side of the screen as shown the figure 5.2. Selection of the range of Y-axis and the selection of building to be clustered are also available.

From the select the data range option as shown in figure 5.3, we can select the desired

range of date, which we want for the analysis and visualization.

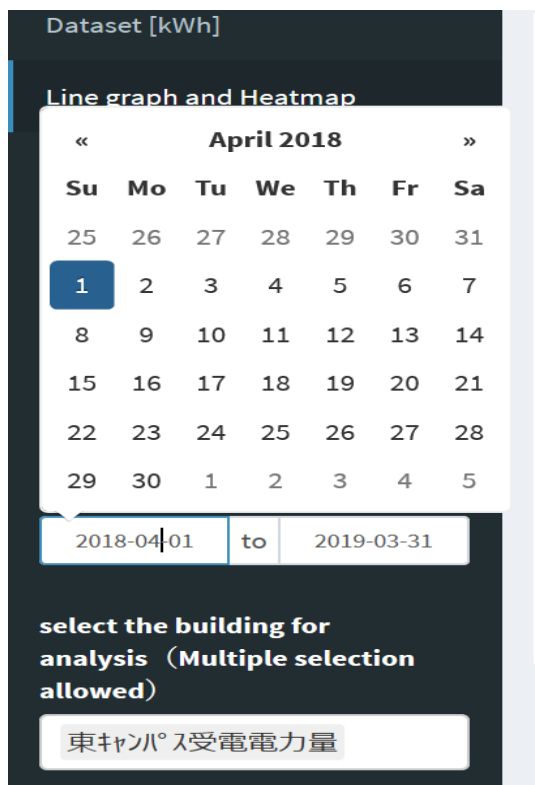


Figure 5.3 The date selection feature of Shiny App

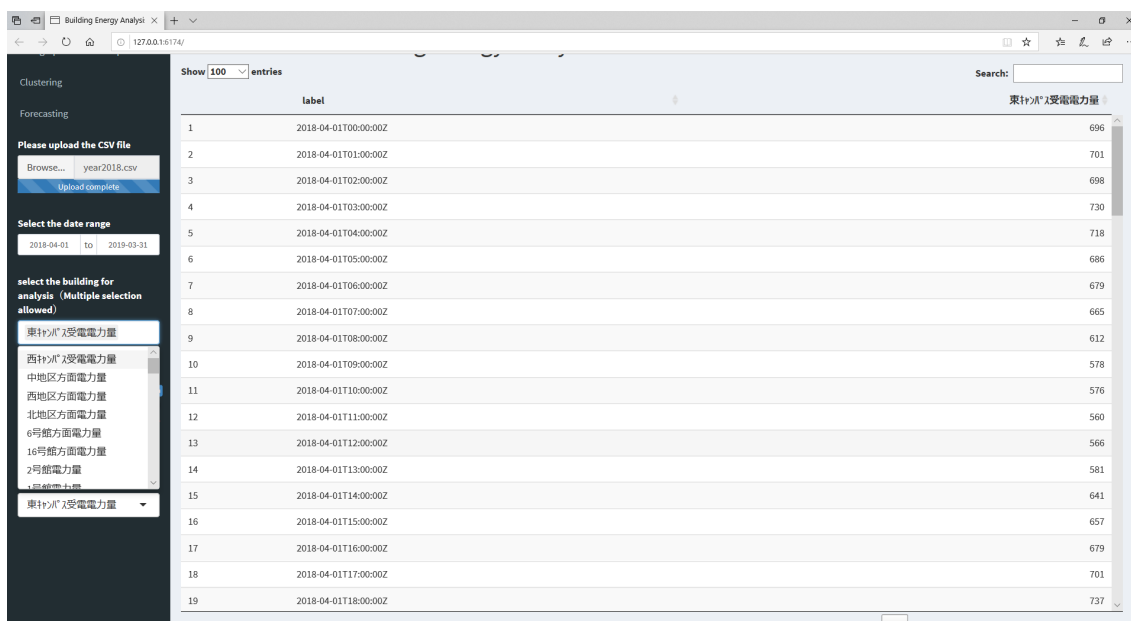


Figure 5.4 Building selection feature of Shiny App

Next option is the selection of buildings for the clustering. On clicking on the “select the building for analysis” option as shown in figure 5.4, a list of all the names of the buildings appears. By using the scroll down menu, desired buildings can be selected. Multiple selection of buildings is also possible. Figure 5.5 shows the visualization of data in the form of trend graph and heat map using Shiny



Figure 5.5 Producing line graph and heatmap using Shiny App

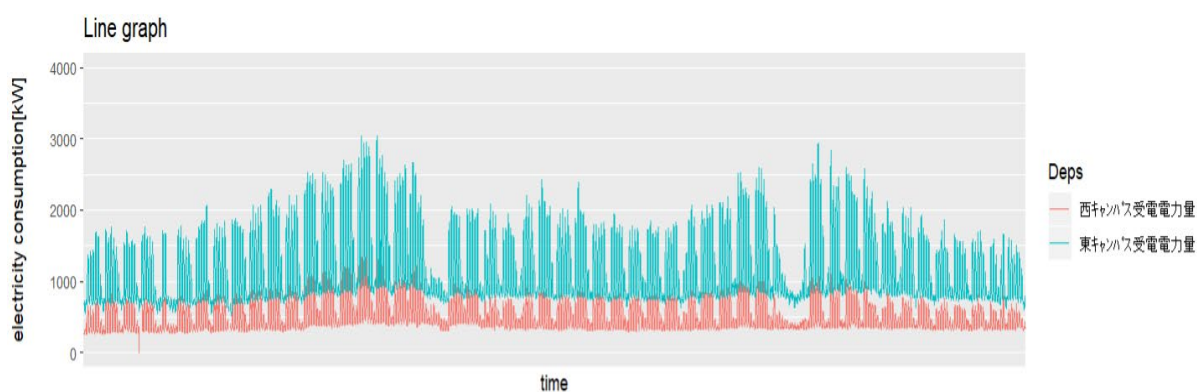
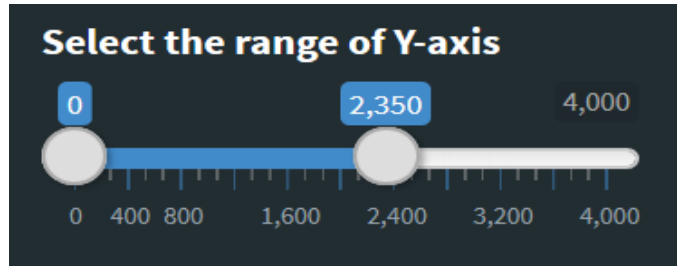


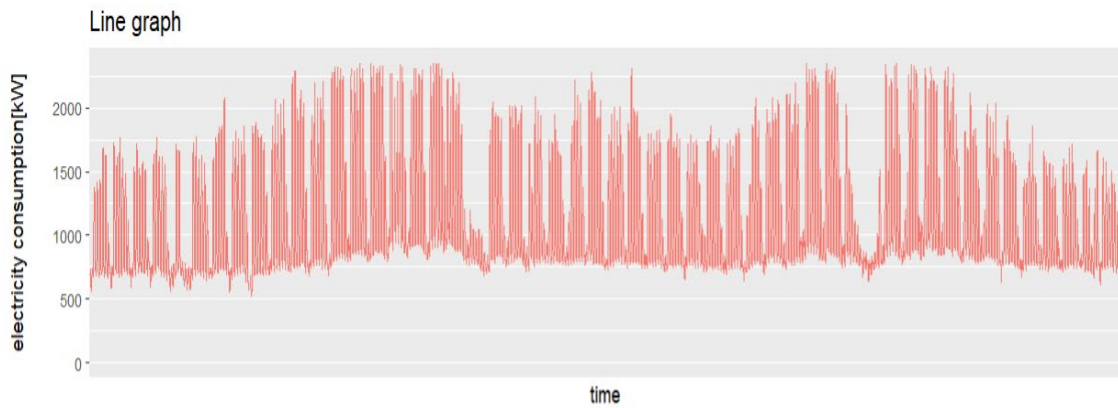
Figure 5.6 Producing multiple building line graph using Shiny App

Figure 5.6 shows the plotting of East campus and West campus of Chubu University by the selection of the both the campuses using the multiple selection features.



**Figure 5.7** Selection of the range of Y-axis feature of Shiny App

By using the “select the range of Y-axis” menu as shown in figure 5.7, the range of Y-axis can be altered. This option comes to be handy when plotting the line graph of two different building with large difference in electricity consumption.

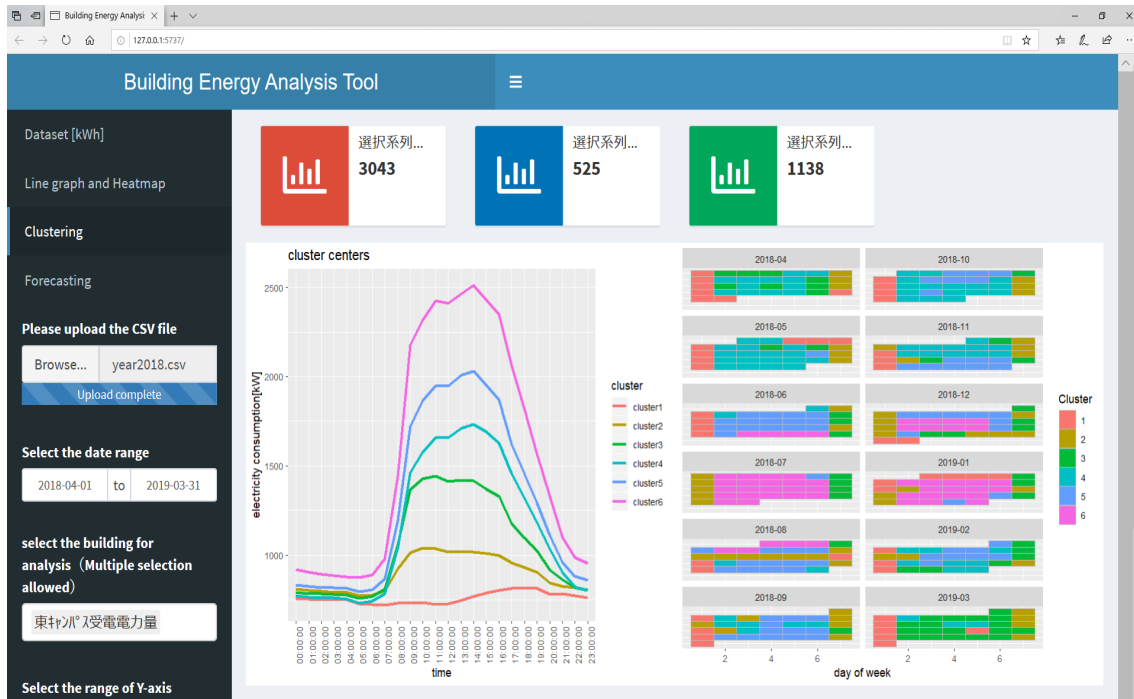


**Figure 5.8** Line graph after the range of Y-axis changed

Line graph when the range of the y-axis is changed from 0-4000 kWh to 0-2500 kWh is shown in figure 5.8.

The clustering button allow us to perform clustering of the imported data and print the clustering result in the form of cluster centres and calendar plot as shown in figure 5.9.

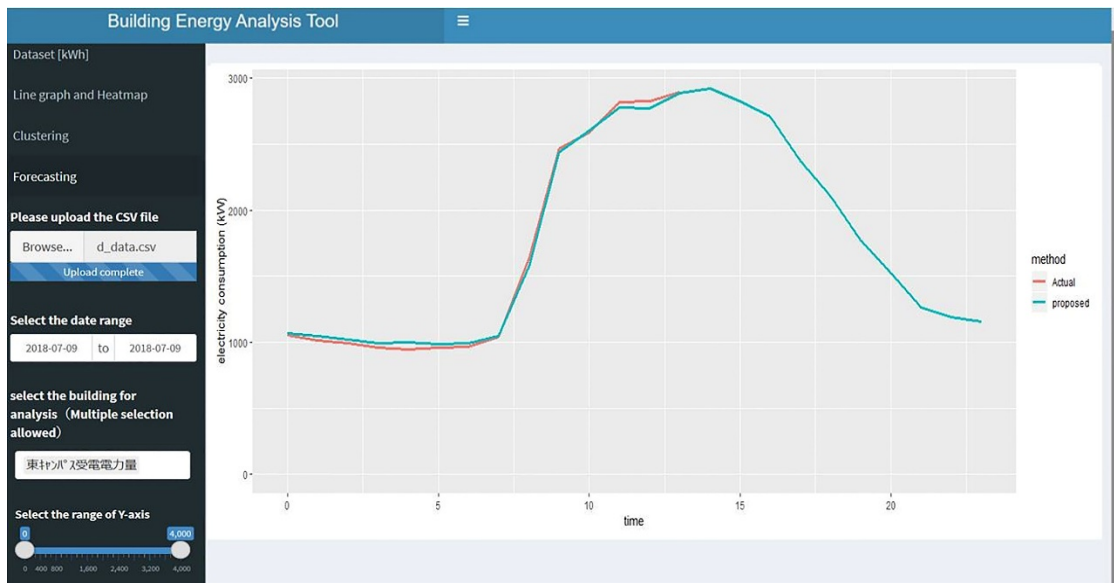




**Figure 5.9** Production of cluster centers and calendar plot using Shiny App

By analysing the cluster center graph and calendar plot graph, we can analyze in which day of the year, how much electricity is being used. The base energy and the peak energy consumption can be visualized.

The final option we plan to develop is the plotting of the electricity load forecasting result. Using a hybrid model of clustering and ARIMA model we can forecast the energy consumption of the forecasting day with noticeably good accuracy. A 24 hour forecasting of the forecasting day is shown in figure 5.10. The “Forecasting” option in the side menu bar allow us to forecast the electricity consumption of the day. This is helpful for the planning of the peak load reduction in buildings.



**Figure 5.10** Forecasting result using Shiny App

Once the Shiny App is deployed on a web-based server such as Amazon Web Service (AWS) anybody with the server address can access the app and interact with it. The real-time data can be uploaded into the server every hour and the real-time comparison between the forecasted and the actual values can be performed. Thus, the efforts taken for peak load reduction can also be seen in real time.

The proposed method using Shiny App and AWS includes

- Making a group e-mail and SNS of energy conservation members of Chubu University which includes professors, master`s course students, and faculty staff.
- Sending a mail with the link to the AWS server to the group members at 9:10 AM on weekdays.
- Requesting the group members to act towards peak load reduction based on the forecasted result.

Deploying the Shiny App to the web-based server has made it possible to view the energy status not only on the computer but also on the smartphone, and this makes the Shiny App more accessible regardless of the location of group members. Moreover, the selection of buildings makes it possible to view the energy pattern of the buildings in which they are interested. Hence, through the proper usage of the Shiny App and cooperation of the group members, electricity peak load reduction can be contributed.

## Chapter 6

### Conclusion

In this paper, we analyzed the electricity consumption pattern of Chubu University and proposed different methods for the energy management of university buildings. For the conservation of energy in university buildings, it is necessary to know when and how much energy is being consumed. For analyzing the energy consumption, we used K-means clustering method. To improve the performance of K-means clustering, we need to choose the initial centroids. In this paper we used percentile method based on the empirical cumulative distribution for the selection of initial centroids. Using this method, the accuracy was improved in comparison to the random initialization method and the empty clusters were also removed. The method was also found to produce better clusters in the case of building energy time series data, which was supported by the fact that the clusters produced by the proposed method has small within-cluster sum of square distance and large between-cluster sum of squared distance. The uncertainty of the K-means algorithm was also removed because the result produced by the proposed method were same irrespective of the number of trials. It is also necessary to determine the number of clusters before running the K-means algorithm and to get better result it is essential to select proper number of clusters. To determine the proper number of clusters for Chubu University, we analyzed the electricity consumption into six clusters. The calendar plot for three to five clusters produced using K-means clustering did not match the university schedule. For six clusters, the clustering result was similar to the university schedule. This was also verified using `clValid` package of R programming language. So, we found that six clusters were appropriate for Chubu University.

We used the K-means clustering to classify the energy consumption of Chubu University into energy consumption by factor which includes “base energy consumption”, “human activity factor energy consumption” and “air-conditioning factor energy consumption”. This classification was done for the departments as well as the buildings of Chubu University and their energy consumption characteristics were analyzed. It was found that College of Engineering has large base energy consumption and the difference between the base energy of cluster 1 and remaining clusters was found high. Characteristics similar to the College of Engineering was also found in College of Bioscience and Biotechnology and College of Life and Health Sciences. The office buildings and non-science department of Chubu University was found to have low base energy consumption and the difference in the base energy of cluster1 and remaining cluster was found low. Thus the energy consumption pattern in case of office buildings of Chubu University was found to have similar pattern to

the buildings of non-science departments. Moreover, the peak of peak days in case of buildings of Science department was found to increase with the increase in the base energy consumption whereas, no such relation was found in the case of buildings of non-science department. In case of buildings of Science department, the possibility of peak energy consumption was found higher, whereas in the case of buildings of non-science departments and office buildings the possibility of peak energy reduction was found higher.

For the peak load management in university buildings forecasting of the electricity load is essential because forecasting earlier to the peak load hours provides the energy conservation team with sufficient time and information for making strategy and act for electricity load reduction. In this paper a hybrid model of clustering and ARIMA model is used for improving the accuracy of forecasting result and making the forecasting of electricity load automatic. For clustering the electricity consumption data, one-year data including the forecasting day from 6 AM to 9 AM is used. Then the one-year data is classified into 6 clusters using the K-means clustering algorithm. The members of the day in which the forecasting belong are used as the training data for forecasting the electricity load. The probability that the cluster members of the forecasting day have similar electricity load pattern like the forecasting is quite high in comparison to other clusters. This novel concept of collecting the days with similar electricity load pattern as the forecasting day has proved to produce better forecasting result than the ARIMA model. The accuracy of the forecasted result of the proposed method and ARIMA model are compared by calculating the MAPE, RMSE and MAE values. One-year forecasting result from 1<sup>st</sup> April 2018 to 31<sup>st</sup> of March 2019 shows that the value of MAPE, RMSE and MAE of the ARIMA model is 16.1, 265.3 and 193.1 whereas the proposed method is significantly lower at 5.1, 81.7 and 61.3 respectively. The proposed method can also be used for the demand response of energy in buildings. As an application of the proposed we developed a data analysis and visualization platform using Shiny App. Shiny App allows the user to interactively view the forecasted result of the desired building. Once the app is deployed into the web-based server such as Amazon Web Service (AWS) anybody having the server address the App and interact with it. The real time data can be uploaded into the server every hour and real time comparison between the forecasted and actual valued can be done. Creating an email or SNS group of members related to the energy conservation in university buildings and sending them messages everyday on the weekdays stating the status of forecasted electricity load of the day and requesting to act for the peak load reduction is expected to help for the energy management of energy in buildings.

## References

- [1] BP Statistical Review of World Energy, 68<sup>th</sup> edition, 2019, <https://www.bp.com/content/dam/bp/businesssites/en/global/corporate/pdfs/energy-economics/statistical-review/bp-stats-review-2019-full-report.pdf>
- [2] Renewables 2019, Global Status report, [http://ren21.net/gsr-2019/?gclid=Cj0KCQjw5MLrBRCIARIsAPG0WGx7\\_HKzgnZBx4nOnW3epq2Ya6ilzk7TaunLgjieO2mMeTtXEbL534QaAltmEALw\\_wcB](http://ren21.net/gsr-2019/?gclid=Cj0KCQjw5MLrBRCIARIsAPG0WGx7_HKzgnZBx4nOnW3epq2Ya6ilzk7TaunLgjieO2mMeTtXEbL534QaAltmEALw_wcB)
- [3] Long-Term Energy Outlook, Trends, Challenges, and Opportunities in a Rapidly Evolving System, 2017, [https://www.enecho.meti.go.jp/committee/studygroup/ene\\_situation/002/pdf/002\\_006.pdf](https://www.enecho.meti.go.jp/committee/studygroup/ene_situation/002/pdf/002_006.pdf)
- [4] World Nuclear Association, Nuclear Power in Japan, Updated August 2019, <https://www.world-nuclear.org/information-library/country-profiles/countries-g-n/japan-nuclear-power.aspx>
- [5] EIA, Country Analysis Brief: Japan, Updated February 2017, [https://www.eia.gov/beta/international/analysis\\_includes/countries\\_long/Japan/japan.pdf](https://www.eia.gov/beta/international/analysis_includes/countries_long/Japan/japan.pdf)
- [6] Energy Technology Perspectives. 2017. Buildings. Available online <https://www.iea.org/topics/energyefficiency/buildings/>
- [7] Kawamura, M. Practical study on save energy and peak shaving for the existing buildings. PhD Thesis, 2017.
- [8] Hayasimoto, Y., Yamaha, M. A Study on operational data analysis method using clustering, *SHASE*, Chubu branch conference, 14, pp. 2012.
- [9] Han, J., Pei, J., Kamber, M. Data Mining Concept and Techniques. Elsevier, 2011.
- [10] Kotsiantis, S., Pintelas, P. Recent advances in clustering: A brief survey. *WSEAS Transactions on Information Science and Applications*, 1(1), pp. 73-81, 2004.

- [11] Khan, SS, Ahmad, A. Cluster center initialization algorithm for K-means clustering. *Pattern recognition letters*, 25(11), pp. 1293-1302, 2004.
- [12] Ilango, MR., Mohan, V., A survey of grid based clustering algorithms. *International Journal of Engineering Science and Technology*, 2(8), pp. 3441-3446, 2010.
- [13] Chitra, K., Maheswari, D. A Comparative Study of Various Clustering Algorithm in Data Mining. *International Journal of Computer Science and Mobile Computing*, 6(8), pp. 109-105, 2017.
- [14] Shah, M., Nair, S. A survey of data mining clustering algorithms. *International Journal of Computer Applications*, 128(1), pp. 1-5, 2015.
- [15] Popat, SK., Emmanuel, M. Review and comparative study of clustering techniques. *International journal of computer science and information technologies*, 5(1), pp. 805-812, 2014.
- [16] Saini, S., Rani, P. A survey on STING and CLIQUE grid based clustering methods. *International Journal of Advanced Research in Computer Science*, 8(5), 2017.
- [17] Ahmara, AS., Napitupulu, D., Rahim, R., Hidayat, R., Sonatha, Y., Azim, M. Using K-means Clustering to Cluster Provinces in Indonesia. *In Journal of Physics*. IOP publishing, 1028(1), pp. 012006, 2018.
- [18] Amri, Y., Fadhilah, AH, Setani, N., Rani, S. Analysis Clustering of Electricity Usage Profile Using K-means Algorithm. *In IOP Conference Series: Materials Science and Engineering*, 105, pp. 012020.
- [19] Damayanti, R., Abdullah, AG., Purnama, W., Nandiyanto, AB. Electricity Load Profile Analysis Using Clustering Techniques. *In IOP Conference Series: Materials Science and Engineering*. IOP Publishing, 180, pp. 012081, 2017.
- [20] Santamouris, M., Mihalakakou, G., Patargias, P., Gaitani, N., Sfakianaki, K., Papaglastra, M., Pavlou, C., Doukas, P., Primiki, E., Geros, V. Using Intelligent Clustering Technique to Classify the Energy Performance of School Buildings. *Energy and Buildings*, 39, pp. 45-51, 2007.

- [21] Arai, K., Barakbah, AR., Hierarchical K-means: An Algorithm for Centroids Initialization for K-means. *Report of the Faculty of Science and Engineering*, 31, pp. 25-31, 2007.
- [22] Yedla, M., Pathakota, SR., Srinivasa, TM. Enhancing K-means Clustering Algorithm with Improved Initial Center. *International Journal of Computer Science and Information Technologies*, 1(2), pp. 121-125, 2010.
- [23] Shakti, M., Antony, ST. An Effective Determination of Initial Centroids in K-means Clustering Using Kernel PCA. *International Journal of Computer Science and Information Technology*, 2(3), pp. 955-959, 2011.
- [24] Huang, JZ., Ng, MK., Rong, H., Li, Z. Automated Variable Weighing in K-means Type Clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1(5), pp. 657-668, 2005.
- [25] Fahim, AM., Salem, AM., Torkey, FA., Ramadan, MA. An Efficient Enhanced K-means Clustering Algorithm. *Journal of Zhejiang University-Sciences*, 7(10), 1626-1633, 2006.
- [26] Prahastono, I., King, DJ, Ozveren, CS. A Review of Electricity Load Profile Classification methods. In Proceedings of the 42<sup>nd</sup> Universities Power Engineering Conference. *IEEE*, pp. 1187-1191, 2007.
- [27] Molina-Solana, M. Ros, M. Ruiz, MD., Gómez-Romero, J., Martín-Bautista, MJ. Data Science for Building Energy Management: A review. *Renewable and Sustainable Energy Reviews*, 70, pp. 598-609, 2017.
- [28] Kim, SS. Variable selection and outlier detection for automated K-means clustering. *Communications for Statistical Applications and Methods*, 22(1), pp. 55-67, 2015.
- [29] Yu, ZJ., Haghighat, F., Fung, BC., Morofsky, E., Yoshino, H. A methodology for identifying and improving occupant behavior in residential buildings. *Energies*, 36, pp. 6596-6608, 2011.
- [30] Bessa, RJ., Trindade, A., Mirinda, V. Spatial-Temporal Solar Power Forecasting for

- Smart Grids. *IEEE Transactions on Industrial Informatics*, 11(1), pp. 232-241, 2014.
- [31] Ceci, M., Corizzo, R., Malerba, D., Rashkovska, A. Spatial autocorrelation and entropy for renewable energy forecasting. *Data Mining and Knowledge discovery*, 15(33), pp. 698-729, 2019.
- [32] Iglesias, F., Kastner, W. Analysis of similarity measure in times series clustering for the discovery of building energy patterns. *Energies*, 6(2), pp. 579-597, 2013.
- [33] UCI Machine Learning Repository. Available online: Available online: <https://archive.ics.uci.edu/ml/dataset.php> (accessed on 6 April 2019).
- [34] Ansari, Z., Azeem, MF., Ahmed, W. Qualitative Evaluation of Performances and Validity Indices for Clustering the Web Navigational Sessions, *World of Computer Science and Information Technology Journal*, 11(5), pp. 217-226, 2011.
- [35] Brock, G., Pihur, V., Datta, S. clValid, an R package for cluster validation, *Journal of Statistical Software*, 25(4), 2008.
- [36] Bon, AT., Jun, OS. Demand Forecasting for Strategic Resource Planning. In *Proceedings of the International Conference on Industrial Engineering and Operations Management*, Detroit, Michigan, USA, pp. 23-25, 2016.
- [37] Wijaya, TK., Vasirani, M., Humeau, S., Aberer, K. Cluster –based Aggregate Forecasting for Residential Demand Using Smart Meter Data. In *2015 IEEE International Conference on Big Data*. *IEEE*, pp. 879-887, 2015.
- [38] Hernández,L., Baladrón, C., Aguiar, J., Carro, B., Sánchez-Esguevillas, A. Classification and Clustering of Electricity Demand Patterns in Industrial Parks. *Energies*, 5(12), pp. 5215-5228, 2012.
- [39] Wang, Y., Chen, Q., Kang, C., Xia, Q. Clustering of Electricity Consumption Behavior Dynamics Towards Big Data Applications. *IEEE transactions on smart grid*. 7(5), pp. 2437-2347, 2016.
- [40] Maksood, FZ., Achuthan, G. Sustainability in Oman: Energy Consumption Forecasting



- Using R. *Indian Journal of Science and Technology*,10(10), pp. 1-14, 2017.
- [41] Sibi, S., Pushpalatha, S. Demand Forecasting of Electricity Consumption Using Dynamic Clustering of Time Series Data. *International Journal of Pure and Applied Mathematics*, 119(15), pp. 525-533, 2018.
- [42] Gajowniczek, K., Ząbkowski, T. Simulation study on clustering approaches for short-term electricity forecasting. *Complexity*, 2018.
- [43] Kore, S., Khandekar, S. Residential Electricity Demand Forecasting Using Data Mining. *International Journal of Engineering Trends and Technology*, 49(1), 2017
- [44] Auder, B., Cugliari, J., Goude, Y., Poggi, JM. Scalable Clustering of Individual Electrical Curves for Profiling and Bottom-Up Forecasting. *Energies*, 11(7), 2018.
- [45] Zhang, GP. Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model. *Neurocomputing*, 50, pp. 159-175, 2003.
- [46] Makridakis, S., Wheelwright, S. and Hyndman, RJ. Forecasting: Methods and Applications, *John Wiley & Sons*, New York, 1998.
- [47] Jakaša, T., Andročec, I., Sprčić, P. Electricity Price Forecasting—ARIMA Model Approach. In *2011 8th International Conference on the European Energy Market (EEM)*, IEEE, pp. 222-225, 2011.
- [48] Hipel, KW. and Mcleod, AI. *Time Series Modelling of Water Resources and Environmental Systems*. Amsterdam. Elsevier, 1994.
- [49] Wateo, S., Churakham, K. and Intarasit, A. Forecasting Time Series Movement Direction with Hybrid Methodology. *Journal of Probability and Statistics*, 2017.
- [50] Box, GE., Jenkins, G.M., Reinsel, GC. and Ljung, GM. *Time Series Analysis: Forecasting and Control*. John Wiley & Sons, Hoboken, New Jersey, 2015.
- [51] Nepal, B., Yamaha, M., Sahashi, H. and Yokoe, A. Analysis of Building Electricity Use Pattern Using K-Means Clustering Algorithm by Determination of Better Initial

Centroids and Number of Clusters. *Energies*, 12(12), 2019.

- [52] Chen, C., Twycross, J. and Garibaldi, J.M. A new accuracy measure based on bounded relative error for the time series forecasting. *PloS one*, 12(3): e0174202, 2017.
- [53] Snipes, M. and Taylor, DC. Model Selection and Akaike Information Criteria: An Example from Wine Ratings and Prices. *Wine Economics and Policy*, 3(1), pp. 3-9, 2014.

本文（日本語）

## 第1章 はじめに

### 1.1 世界エネルギー消費

現在、世界には2つの主要なエネルギー関連の課題がある。大きな課題の1つは、世界の高まるエネルギー需要を満たすことであり、次は化石燃料からの炭素排出量の削減である。エネルギー需要の重心は変化しており、アジアの拡大する中流階級が世界のGDPの成長の大部分を占めている。一方、エネルギーの消費方法はリアルタイムで変化しており、世界が電化し、エネルギーはますます競争が激しく効果的なデジタル市場で売買されるより広範なサービスの一部になっている。

2018年のレポート<sup>1)</sup>によると、2018年の世界のエネルギー需要と炭素排出量の使用は、2010/11年以来最も速いペースで増加し、パリの気候目標で想定された加速的な移行からさらに遠ざかりました。報告書は、エネルギー成長の増加の大部分は、天候に関連した影響による冷暖房の需要の増加によるものであると述べた。中国、米国、およびインドは合わせて、需要増加全体のほぼ70%を占めた。化石燃料消費の増加により、世界のエネルギー関連の二酸化炭素排出量は、この年に推定1.7%増加した<sup>2)</sup>。

世界エネルギーの良い面の1つは、再生可能エネルギーへの投資が増えていることである。図1.1は、近年の投資の約3分の2が再生可能セクターへの投資であることを示している。太陽光と風力は現在、電力部門の主流の選択肢であり、ますます多くの国が太陽光と風力で電力の20%以上を発電している。しかし、現在の傾向は、エネルギーシステムを持続可能なものにするために、エネルギー最終用途のすべてのセクターで広範囲にわたる政策決定が必要であることを示している<sup>2)</sup>。図1.2は、再生可能エネルギーの増加と化石燃料への依存度の減少にもかかわらず、2040年まで化石燃料がエネルギー市場の約50%を占めることを示している。

### 1.2 日本のエネルギーシナリオ

2011年まで、日本は原子力で約30%を発電しており、これは2017年までに少なくとも40%に増加すると予想された。現在、計画は2030年までに少なくとも20%である。最近の日本では、エネルギー需要量の約90%を輸入する必要がある。現在、37基の原子炉が稼働な状態である。最初の2つは2015年8月と10月に再起動し、さらに7つが再起動した。現在、17の原子炉が再起動承認のプロセスにある。図1.3(a)は2000年から2015年までの日本の総エネルギー消費量を示し、1.3(b)は2015年のエネルギー消費量を示す円グラフである。

### 1.3 建物のエネルギー消費

建築部門のエネルギー需要は、総エネルギー使用量の30%以上、世界の電力需要の55%以

上を占めている。建築部門のエネルギー需要は、発展途上国におけるエネルギーへのアクセスの改善、エネルギー消費機器の所有と使用の拡大、および建物の床面積の急速な成長により、急速に増加している。持続可能な建築物を開発する取り組みは進んでいるが、建築部門の成長とエネルギー需要の増加に追いついていない。建物や建物の建設からのエネルギー需要は、発展途上国のエネルギーへのアクセスの改善、エネルギー消費機器の所有と使用の拡大、および世界の建物の床面積の急速な成長により、年間約3%増加している<sup>6)</sup>。したがって、化石燃料の枯渇による将来のエネルギー危機の問題を克服し、CO<sub>2</sub>の過剰排出による地球温暖化から環境を保護するためには、化石燃料から再生可能エネルギーへの移行と省エネルギーが同様に重要である。

本論文では、クラスタリング手法を用いたエネルギー消費パターン分析を使用して、建物の省エネルギーに関連する方法について議論する。中部大学が省エネルギーのために取ったさまざまなアプローチは以下で説明する。

#### 1.4 中部大学の省エネルギーへの取り組み

中部大学は愛知県にある日本の私立大学である。大学は7つの異なる学部で構成されている。東日本大震災の後、ほとんどの発電所が閉鎖され、日本は突然エネルギー危機の問題に直面した。それ以来、日本政府や一般市民、民間企業や大学での意識が高まった。2015年12月、パリ気候会議(COP21)が開催され、195か国が世界で初めて法的かつ法的拘束力のある世界気候協定を採択した。日本は、2013年と比較して2030年までに国内の排出削減量を26.0%削減することを約束した。

2010年には、中部大学の電力消費量が電力需要を上回り、2011年に政府から省エネルギーの要請があり、中部大学は省エネ活動を強化した。2010年、中部大学のCO<sub>2</sub>排出率は70.7 kgCO<sub>2</sub> / m<sup>2</sup> / 年であり、これは日本の理科系大学のエネルギー消費量に近い。

エネルギー消費量を削減するために、多くの省エネルギー対策が実施されている。中部大学では、BEMSというシステムを導入した。これは、学内のエネルギー計量システムおよび、イントラに接続され、各グリッドのエネルギー消費状況の一元管理と、節電メールを配信する。図1.9の制御システムにおいて、学部やキャンパスのエネルギーの予測を天気などを基に行う。図中の中央にあるBEMSは、電力管理値とこの予測値の差から節電目標量を算出し、図の下部の設備機器の自動制御やマイクログリッドの放電制御を行う。合わせて図中の右の大学のイントラを介して、施設利用者に予報や注意報の節電メールを配信し、自動制御に依らない実験研究機器や研究室の設備機器の節電を要請する。図中左下の計量システムからリアルタイムの計量値から、節電目標値を超過すると予測された場合は、節電警報を該当の学部グリッドやキャンパスグリッドに配信して節電活動を促す。学部とキャンパスグリッドの概要を図1.8に示す。ここで学部グリッドがここで述べている一層目でその上位にキャンパスグリッドが二層目に位置付けられる。各学部グリッドとキャンパスグリッドの電力消費予測は、気象予報と電力消費実績の重み

付けで算出される。学部グリッドでは、各々の電力消費が、節電目標値の90%に達した時間に注意報(L1)と95%の警報(L2)の2段階で、節電メールの配信および、空調と照明の自動制御による節電を行う。

東西のキャンパスグリッドは、それぞれの消費予測と各学部グリッドの電力消費合算値を管理する。目標値を超過すると予測された場合は、受電系統内の全学部グリッドに、前述の節電メールの配信と自動制御が実行され、節電を行う。発電側には、マイクログリッドが西受電系統の生命健康科学部に続いて、東受電系統の学生教育推進機構にも設置された。注意報(L1)の配信時に放電して、学部グリッドおよび、各受電系統の電力ピークの平準化を行う。2012年には、前年より24.4%少ない115kWの電力消費が削減され、また、6.3%低い炭素排出量が達成されました。

各施設グリッドとキャンパスグリッドの消費電力予測は、天気予報と消費電力パフォーマンスを比較することで計算される。部門レベルでは、各電力消費量が節電目標値の90%に達すると、節電メールは2段階で配信される。警告メッセージと最終警報は節電目標の95%レベルである。東および西のキャンパスグリッドは、独自の消費予測と、各部門のグリッドの合計電力消費を管理する。目標値を超えると予測される場合、上記の節電メールが送信され、節電のためにすべての部門グリッドに自動制御が実行される。発電側では、蓄えられた電気の放電は最初のアラームの配信時に行われる。図 1.10 に示すように、2016年のCO<sub>2</sub>排出量は、70.7 kgCO<sub>2</sub> / m<sup>2</sup> 年から50.8 kgCO<sub>2</sub> / m<sup>2</sup> 年に削減された。CO<sub>2</sub>排出量は、中部大学を日本の文系大学のレベルまでなった。

## 1.5 消費電力分析にクラスタリング手法を使用する理由

中部大学のエネルギー消費削減のためのさまざまな手段が<sup>7)</sup>に記載されている。図 1.11 に示すように、スマート BEMS を使用して中部大学の電力消費量を分類するために、大学全体と学部の電力消費量を季節と年末年始(すなわち、夏、冬、春、秋、新年年末年始)代表週から、ベースエネルギー消費とピークエネルギー消費が決定された。しかし、これは時間をかけた分析であり、各季節の同じ週が各建物の代表的な日を表さない場合がある。

<sup>8)</sup>では、大学のカレンダーが大学のクラスタリング結果と一致していることがわかった。建物の毎日の電力消費データを分析することで建物のエネルギー消費パターンを分類する方法が不可欠であり、この場合、クラスタリング手法が有効になる。したがって、本研究では、クラスタリング手法を使用して、大学のエネルギー消費パターンを分類する。

## 第2章

### クラスタリングと提案手法

#### 2.1 クラスタリング

クラスタリングは、データオブジェクトをグループまたはクラスタに分割するプロセスであるため、クラスタ内のオブジェクトは互いに類似しており、別のクラスタ上のオブジェクトとは異なる<sup>9)</sup>。クラスタ分析は、インテリジェントなデータ分析プロセスの基礎として機能するこれでよい方法である。クラスタリングには、視覚化、最適化、財務、製造、および医療組織のさまざまな分野で応用されている<sup>11)</sup>。センサー、データロガー、検出、メモリ技術の改善、およびインターネット検索、デジタルイメージング、ビデオ監視の著しい進歩により、膨大な量のデータが毎日生成されている。データの量と種類の両方が急速に増加しているため、データを自動的に理解、処理、および要約できる方法の開発が必要になる。この問題に対処するには、クラスタリング手法が役立つ。クラスタリングアルゴリズムは、パーティションアルゴリズム、階層アルゴリズム、密度アルゴリズム、グリッドアルゴリズムに分類できる<sup>12-14)</sup>。

さまざまなクラスタリングアルゴリズムの中で、MacQueen が開発した K 平均法が最も広く使用されている。この論文では、有名な分割アルゴリズムである K 平均法のアルゴリズムを使用する。K 平均法は、大量のデータを計算時間で比較的高速かつ効率的にグループ化できるため、よく使用される<sup>17)</sup>。

#### 2.2 クラスタリングの文献レビュー

クラスタセンターの初期化の問題を解決しようとする試みがいくつか行われた。Amri はクラスタリングの K 平均法を使用して、2011 年から 2014 年までに収集された 370 クライアントの電力使用パターンを分析した。次元削減手法を使用して、データ全体を 4 つの属性に削減した。各属性は、春、夏、秋、冬の 4 つの季節の値の合計を表している。次に、K 平均クラスタリングを使用して、データを 5 つのクラスタに分類した。その結果、夏と春の最高と最低のエネルギー使用量が決定された<sup>18)</sup>。Damayanti は K ハーモニッククラスタリング手法を使用して、1 年間の電力消費プロファイルをグループ化した。1 年間の電気使用データは 2 つのクラスタに分割された。両方のクラスタに対して電力消費プロファイルが生成された。最初のクラスタパターンは休日に支配される不規則な負荷パターンを示したが、2 番目の負荷プロファイルはオフィスや産業の負荷需要が高いために平日の負荷を示した<sup>19)</sup>。Santamouris はインテリジェントなファジークラスタリング手法を使用して、学校の建物のエネルギー性能を分類した。総エネルギー使用量と暖房用エネルギー使用量の両方が 5 つのクラスタに分類された。クラスタリング手法は、周波数分析手法と比較された。ファジークラスタリング手法は、より堅牢なクラスを生成し、既存の類似性に従って建物を分類することが分かった<sup>20)</sup>。Arai と Barakbah は、K 平均アルゴリズムの初期重心を最適化する階層的方法を提案した。このアルゴリズムは、K 平均アルゴリズムのクラスタリング結果を使

用し、階層アルゴリズムと組み合わせて K 平均の初期重心を決定することにより、クラスタリング結果のすべての重心を変換した。この方法は、大きなデータセットと多くの属性を持つ複雑なクラスタリングに適している。ただし、この方法は、速度と精度について K 平均アルゴリズムを利用する<sup>21)</sup>。

Yedla は、より良い初期重心を見つけるための新しい方法を提案し、時間の複雑さを軽減して適切なクラスタにデータポイントを割り当てる効率的なプロセスを提供した。この方法では、原点から各データポイントの距離を計算して初期重心を選択し、データポイントを並べ替えて k 個の等しい部分に分割した。次に、各データポイントの中間点が初期重心として取得した。この方法は、より良い初期重心を生成することがわかり、適切なクラスタにデータポイントを割り当てる効率的な方法を提供した<sup>22)</sup>。Shakti と Thanamani は、時間の複雑さを軽減し、K 平均アルゴリズムの精度を向上させる方法として、KPCA 方法を提案した。次元削減手法では、主成分分析(PCA)が使用され、KPCA は PCA の拡張形式である。クラスタリングに大きなデータセットが使用される場合、K 平均アルゴリズムの時間の複雑度は高くなる。したがって、K 平均アルゴリズムを適用する前に、データの次元は KPCA を使用して削減される。この方法は、高次元データの K 平均アルゴリズムの問題を克服することができる<sup>23)</sup>。

Huang は可変重みを自動的に推定できる K 平均アルゴリズムの自動可変計量方法を提案した。このアプローチによって作成される変数の重みは、クラスタリングにおける変数の重要性を推定し、大きく複雑な実データがよく使用されるさまざまなデータマイニングアプリケーションの変数選択に展開できる。重み値による重要でない変数の識別と削除は、クラスタリング結果を改善することがわかった<sup>24)</sup>。

Fahim は K 平均アルゴリズムの性能を向上させる方法を提案した。元の K 平均アルゴリズムでは、各点とクラスタセンター間の距離が各反復で計算されるため、計算時間が長くなる。この方法では、2 回目の反復から、距離が増加し、計算時間が減少したポイントのみについて距離が測定される。ただし、初期重心はランダムに選択されるため、この方法は初期重心の影響を受けやすい<sup>25)</sup>。プラハストノ他顧客の分類と電力消費プロファイルの生成のためのさまざまなクラスタリング手法とその特性について説明した。各クラスタリングアルゴリズムには独自の特性があるため、データに応じてクラスタリングアルゴリズムを選択することが不可欠である<sup>26)</sup>。

Moliana-Solana は報技術の開発と建物のエネルギー管理への影響をレビューした。スマートメータリング、IoT 等エネルギー問題を解決するために適用された、または適用できるクラウドコンピューティングエネルギー管理と技術を構築するためのデータサイエンスのさまざまな側面が議論した<sup>27)</sup>。

上記で要約された文献は、より良い初期重心の選択のための新しい方法論を提案している。ただし、時系列のエネルギー使用データの初期値の選択に関しては、ほとんど研究が行われていない。電力使用データを構築するためのより良い初期重心を決定するためにこ



の研究で使用された方法論には、2つの主な利点がある：(1)提案された方法によって生成された結果は、試行回数で変わらない(2)提案された方法の精度は、K平均ランダム初期化方法の精度より優れている。

### 2.3 外れ値の検出と欠損値の補完

外れ値とは、他の観測とは異なる観測であり、異なるメカニズムによって生成されたというおそれがある。また、データの想定モデルに適合しないノイズの多い観測値である場合もある<sup>28)</sup>。異常値は、一般的に建物の電気使用量の測定で発生する。クラスタリングを行うデータに異常値が存在すると、不適切な結果になる可能性がある。クラスタリングデータに欠損値がある場合、クラスタリングは不可能になる。ただし、外れ値として固定数の値を単に削除すると、重要な観測値が誤って失われる可能性がある。したがって、この研究では、外れ値を検出するために、最小-最大正規化手法を使用してデータを[0,1]に正規化した。この研究で使用されるデータに存在する異常値にはいくつかの極端な値が含まれているため、異常値の検出に四分位数を使用する Tukey の方法を使用した。

Tukey の方法<sup>29)</sup>は、極端な値に耐性のある四分位数を使用するため、極端な値にあまり敏感ではない。四分位範囲は、下(Q1)と上(Q3)の四分位間の距離である。内側のフェンス =  $[Q1 - 1.5 IQR, Q3 + 1.5 IQR]$ 、外側のフェンス =  $[Q1 - 3 IQR, Q3 + 3 IQR]$ 。内側と外側のフェンスを超えた値は、可能性のある外れ値と見なされたが、外側のフェンスを超えた値は、可能性の高い外れ値と見なされる。この研究で使用したデータには、残りのデータと比較していくつかの極端な値が含まれていた。その結果、すべてが内側のフェンスで検出されたわけではないため、外側のフェンスが異常値の検出に使用された。外側のフェンスを超える外れ値は、使用不可(NA)値に変換された。次に、Rプログラミング言語の zoo パッケージを使用した線形補間手法を使用して、NA 値の代わりに適切な値を代入した。図 2.1 は、外れ値のある箱ひげ図である。極値が存在するため、箱ひげ図は線の形に見える。図 2.2 は、Tukey の方法を使用して外れ値を削除した後の箱ひげ図である。内側のフェンスの上にはいくつかの値が見られたが、これは外れ値ではなく、夏と冬の空調ピーク日の電気使用データで発生する実際の値である。

### 2.4 K平均法クラスタリングアルゴリズムの概要

式2のようにK平均はEuclidean距離式を使用して、2つのオブジェクト間の相関を見つけます。初期重心は、K平均クラスタリングを使用してランダムに決定される。実行される手順は次のとおりである。

- i) kの目的の値を決定する。kの値は目的のクラスタの数である。
- ii) 初期重心を決定する。初期重心は既存のデータからランダムに割り当てられ、クラスタの数は初期重心の数に等しくなる。
- iii) Euclidean距離式を使用して各重心までの距離を計算することにより、各データポイント

の最も近い重心を見つける。

iv) データを最小距離でグループ化する。データポイントは、クラスタの中心から最も近い距離にある場合、クラスタの一部になる。

v) 各クラスタのデータの平均に基づいて新しい重心を見つける。

vi) ステップ(iii)に戻る。

vii) クラスタの割り当てにデータの変更がない場合は停止する。

この研究で使用したデータは、建物の電力消費データの時系列であり、入力ベクトルの形状には、時間ごとに配置された特徴があった。時系列データは一般的に予測に使用される。

Ricardo は<sup>30)</sup>および Michelangelo は<sup>31)</sup>太陽光発電の予測と再生可能エネルギーの予測にそれぞれ時系列データを使用した。

## 2.5 提案手法

K 平均クラスタリングは初期重心をランダムに選択するため、クラスタリングアルゴリズムの結果は毎回変化する。クラスタリングの実行回数に関係なく、同じ結果を生成するクラスタリングアルゴリズムを持つことが望ましい。より良い初期重心を選択することにより、K 平均アルゴリズムのパフォーマンスを向上させる方法を提案する。このメソッドは、同じクラスタリング結果を生成し、クラスタリングの精度は K 平均ランダム初期化メソッドよりも優れている。

中部大学の電力使用パターンの変化を示すために、**図 2.4** に示すように、電力使用期間の異なる 6 セットのグラフを示す。日曜日と祝日を除いて、中部大学の電力使用パターンはほぼ同様であった。電力使用量は午前 8 時ごろに急激に増加し、午後 12 時から 1 時ごろにピークに達する。夕方から早朝にかけての電力使用量は少なく、大学のベース電力消費量を表している。主に違いは 1 時間ごとの電気使用量の観点から見られた。したがって、この研究では、1 年間の電力使用量の時間ごとの分布に基づいて初期重心を選択することにした。

K 平均アルゴリズムの初期クラスタセンターは、累積頻度分布関数に基づくパーセンタイル法を使用して決定された。k 個のクラスタの場合、累積密度は (k + 2) 等分されたパーセンタイルに分割された。次に、累積頻度分布のパーセンタイルに対応する電力消費データの値が選択された。最大値と最小値、および可能性のある外れ値の影響を減らし、空のクラスタを防ぐために、最小と最大の初期値は無視された。したがって、k + 2 値のうち最後の k 値が K 平均の初期重心として採用された。

**図 2.5** は、パーセンタイル法を使用した累積頻度分布における初期重心の決定を示している。**図 2.5** では、23 : 00 のデータの分布が示されている。残りのタイムスタンプのその他の初期重心も同様の方法で計算された。k 個のクラスタの場合、(k + 2) 等間隔のパーセンタイル値が選択され、(k + 2) 値から k 値が取得された。

## 2.6 実験と結果

### 2.6.1 精度検証

#### 2.6.1.1 依存データセットを使用した精度検証

提案された方法の精度を検証するために、4つの異なる依存データセット、すなわち、アイリスデータ、ワインデータ、Ruspini データ、および New-Thyroid データで検証を行った。Iris データ、Wine データ、および New-Thyroid データセットは、UCI Machine Learning Repository<sup>33)</sup>からダウンロードされた。Ruspini データは R パッケージ「クラスタ」に含まれている。

K 平均および提案された方法を使用して得られた 100 回の試行の最大、最小、および平均精度を表 2.1 に示す。提案された方法の精度は、ランダムな初期化を伴う K 平均アルゴリズムを使用した 100 回の試行の平均精度よりも著しく高いことがわかる。表 2.2 に示すように、提案された方法の平方距離のクラスタ内合計は小さいが、提案された方法の平方距離のクラスタ間合計は、4 つすべてのデータセットのランダム初期化法を使用した K 平均よりも大きかった。これは、提案されたアルゴリズムが K 平均ランダム初期化法よりも優れたクラスタリングを生成することを示す。

#### 2.6.2 大学データを使用したクラスタ品質の比較

提案された方法を使用して、建物の電力消費データを分析した。この研究で使ったデータは、中部大学の電力消費データである。大学の各建物の電量消費データは BEMS を使用して測定され、データは BEMS サーバーによって 1 分毎に収集される。これらのデータを合計して、1 時間ごとの電力消費データを作成した。したがって、各建物からの 1 年間のデータは 8,760 個のデータで構成される。

データは、2015 年 4 月 1 日から 2016 年 3 月 31 日までの日順に並べられている。毎日、0:00 から 23:00 までの 24 時間のデータで構成されている。この形式のデータはクラスタリングに適していないため、**図 2.6** に示すように、データを  $366 \times 24$  次の行列に変換する必要がある。**図 2.7** は、この研究で使った  $366 \times 24$  オーダー行列のデータフレームの一部を表している。2016 年はうるう年なので 366 日分のデータがある。

クラスタの品質を決定するために、K 平均アルゴリズムと提案手法のクラスタ内およびクラスタ間の平方和の合計を決定した。クラスタ化を向上させるには、クラスタ内の距離を小さくし、クラスタ間距離の大きくすることが求められている。表 2.3 に示すように、ランダム初期化を使用して K 平均から取得した値は、10 回の K 平均クラスタリングから取得した値の平均である。提案された方法の試行回数は、毎回同じ結果が得られるため、実行は 1 回のみである。表 2.3 に示すように、クラスタ 4 からクラスタ 7 の各ケースでは、提案された方法のクラスタ内の平方距離の合計は、ランダム初期化の K 平均よりも小さく、クラスタ間の平方距離では提案された方法の重心は、ランダム初期化方法よりも大きかった。したがって、提案されたアルゴリズムは、ランダム初期化法を用いた K 平均よりも優れている。

た。

## 2.7 クラスタ数の決定

大学での電力消費に影響する多くの要因がある。日曜日と祝日には、講義や照明、空調、OA機器は使用されない。したがって、電力消費は主に大学のベースエネルギーによるのである。ベースエネルギー消費は、サーバー、冷蔵庫、非常口照明、研究機器など、常時稼働している機器の電力消費である。中部大学では、土曜日は休業日ではないが、出席する学生数、講義と活動も平日に比べて非常に低い。したがって、土曜日の電力使用パターンは、日曜日、休日、平日とは異なる。中部大学の空調の使用は期間で管理されている。気象条件の突然の変化などの例外的な条件を除き、中間期間の空調の使用は許可されない。日本の大学の場合、通常は電力会社と契約が結ばれ、大学の電力需要の契約値は固定されている。空調のピークがある夏と冬に、電力使用量が契約電力需要を超える場合、大学は電気代に高い金額を支払う必要がある。したがって、大学当局は契約値を超えて電力消費が発生しないように努力する。したがって、大学の電力使用パターンを分析する一方で、空調のピーク期間も考慮する必要がある。表 2.4 に示すように、中部大学の電力使用量は、講義の有無、空調の有無、休日の有無によって 6 種類に分類でる。

建物の電力使用パターンを分析するには、使用されている電力量と日の関係を知る必要がある。そのため、電気の使用パターンを適切に説明する適切な数のクラスタ(k)を選択することが不可欠である。この目的のために、3~6 のクラスタ番号を選択して K 平均クラスタリング分析を実行し、大学のスケジュールを使用して作成された実際のカレンダープロットと精度を比較した。図 2.8 および 2.9 は、中部大学のクラスタ 3~6 のクラスタリング結果を示している。

2015 年の中部大学の場合、空調の冷房期間は 6 月 15 日から 9 月 15、暖房期間は 2015 年 11 月 15 日から 2016 年 4 月 15 である。表 2.4 に記載されているクラスタ番号が変更される場合もある。たとえば、日曜日はクラスタ 1 に該当するが多いが、オープンキャンパス、大学祭などのイベントが発生すると、特定の日のクラスタ番号がクラスタ 1 から 2、3、またはそれ以上のクラスタに変更される。したがって、表 2.4 に加えて、各日のクラスタ数を決定するには、各日の電力量消費を分析する必要がある。したがって、表 2.4 を考慮し、毎日の電力使用量を分析することにより、中部大学のカレンダープロットが作成される。このカレンダープロットを実際のカレンダープロットと見なし、これを使用してクラスタリング結果の精度を比較する。

中部大学の実際のカレンダープロットと 3 つのクラスタのクラスタリング結果のカレンダープロットを比較すると、日曜日と土曜日は分離されていない。また、空調期間と非空調期間は分離されない。4 つのクラスタでは、空調期間と中間の非空調期間は分離されていない。5 つのクラスタの場合、空調のピークは、大学のスケジュールと比較して、実際の期間より前に開始された。したがって、クラスタ 3~5 は、中部大学の実際の電力使用パターン

を表すことはできない。6つのクラスタの場合、空調と非空調の中間期間が分離された。また、ピーク時の電力使用期間も大学のスケジュールと一致していた。6つのクラスタのクラスタリング結果の精度は、クラスタリングによって生成されたカレンダープロットを大学のスケジュールと比較することによって決定された。図 2.11 は、各クラスタの曜日と日数を表すクラスタリング結果を示している。クラスタ 1 は、主に日曜日、土曜日、および講義のない日に集中していることがわかった。非講義日の電力消費量を表すクラスタ 2 は、主に照明と OA 機器の電力消費によるもので、主に土曜日に集中していた。空調の日は、ほぼ常に平日に集中していた。

## 第3章

### 提案された K 平均クラスタリングを使用したエネルギーパターン分析

#### 3.1 要因別エネルギー消費の推定

Kawamura ら<sup>7)</sup>は、春夏秋冬と年末年始の代表週を実測から求め、平均値を重ね合わせ、年度ごとの電力消費の変化を確認している。代表週は5月、7月、1月の第三週と年末年始を含む週としている。ベース電力やピーク電力を抽出し、対策の検討の要否や要因の検討を行っている。図 3.1 は夏期の例を示すが、不定期なピーク電力の出現があり、課題を抽出できる。ここでは、代表週を決めて分析しているが、運用の変化などにより特徴的な消費となる期間が変わることも考えられる。図 3.2 に後述するクラスタリング手法である k 平均法で分類された結果を示す。この方法では、分類は電力消費の大小で行われ、図-3.2 は夏期のピーク期間に相当する。期間が一週間に限定されていないため、図-3.1 より多くの日が選ばれているが、ピーク期間の抽出という点では図-3.1 と同様に分析に使用できると考える。

##### 3.1.1 分析に使用されるデータ

データは中部大学スマート BEMS で得られるデータで、各建物のキュービクル毎の計測ポイントで1分間隔に収集される。データはサーバーに集積され、日毎のファイルがダウンロードできる。解析は2015年度のデータを用いた。表-3.1 に部局と建物の対応を示す。

#### 3.2 クラスタリング結果

全学での電力消費によるクラスタリングを行った。クラスタリングの結果、各クラスタの重心となるデータが求められた。これを図-3.3 に示す。電力消費量の大小で6つのクラスタとなっている。

図 3.4 から 3.11 に各部門でのクラスタリング結果を表示する。大学全体と同様に、電力消費の低いクラスタ1から電力消費の高いクラスタ6の順番になっている。午後10時から午前8時までのエネルギー消費は、サーバー、冷蔵庫などのエネルギー消費によるものである。図 3.4 では、工学部のベースエネルギー消費が異なっている。クラスタ1とクラスタ6のベースエネルギー消費量の差が大きいことがわかる。この効果は、夜間の空調使用や研究施設の影響と考えられる。

#### 3.3 エネルギー管理の適用

##### 3.3.1 要因別エネルギー消費の推定

大学では冷暖房を期間で管理しており、2015年は、冷房期間は2015年6月15日～9月15日、暖房期間は2015年11月16日～2016年4月15日であった。ただし、研究室では申請により、講義室では気象状況により、この期間を越えて運用されることもある。これらを考慮して、それぞれのクラスタと授業日の関係は表 3.2 のようになる。大学スケジュールとの関係から、クラスタ1は主に休日のため、人間活動や空調に関わる電力はわずかで、ベース電力消費と考えられる。クラスタ2, 3は冷暖房期間から外れ就業している期間なので

電力消費は人間の活動によるものと考えられ、クラスタ 1 との差が人間活動分、クラスタ 4～6 とクラスタ 3 の差は空調の使用が主となるので空調要因分と呼ぶこととする。

表 3.3 に学部毎のクラスタとの要因別エネルギー消費の推定方法を示す。空調は大部分でビル用マルチパッケージ型が採用されており、電気式だけでなく、ガスでも行われており、学部毎に構成が異なる。第 5 列にはガス空調の区別を示す。大学スケジュールとの対応により、冷暖房期間から空調の有無を判断しているが、クラスタリングしたデータが電力消費なので、ガス空調についてはガス消費量から換算し、空調用エネルギー消費に計上する。表 3.2 に示すように、クラスタとエネルギーの利用用途の分類から要因別エネルギー消費を推定する。

図 3.13 に推定した一次エネルギー消費比率を示す。ベースの比率は理系学部である工学部で最も大きくなっているのは、研究機器のエネルギー消費の影響と考えられる。実験施設等が少ない事務系建物、人文学部、経営情報学部、現代教育学部等はベースの比率が小さい。大学全体である全学でベースの比率が大きくなるのは冷蔵庫、サーバー、実験施設などの影響があると考えられる。

### 3.4 クラスタリング分析結果の精度測定

以上までに、k 平均法によるクラスタリングからエネルギー消費を要因別に分類できることを示した。この内容の確認について、個別に照明、空調とコンセントの電力消費が計測されている 52 号館で要因別エネルギー消費の確認を行った。電力が系統別に計測されている唯一の建物が 52 号館である。ここでは、各階の分電盤内で照明、コンセント、動力が個別に計測されている。照明には照明器具、換気扇および空調室内機の電力消費が含まれている。動力には排水ポンプ、冷蔵庫の電力消費が含まれ、昇降機は含まれていない。

計測データは 1 時間間隔のデータであり、2015 年 4 月から 2016 年 3 月まで系統別に計測されたデータである。52 号館で個別に計測した照明、コンセントと動力の電力消費を合計して 52 号館全体の電力消費にした。その全体のデータにクラスタリングを行って、要因別エネルギー消費を推定した。図 3.14 は各クラスタの重心となるデータを示す。電力消費の小さいものから 1 から 6 の順番となっている。図 3.15 にクラスタリングされた日をカレンダーに落とし込んだものを示す。この建物では電力消費のもっとも大きいクラスタ 6 は冬期に現れている。

室外機動力盤は別建物に所属されており 52 号館の電力消費には入っていない。一方、室内機は全建物で 44 台であり、その消費電力は合計して 1.2kW 程度で、接続されている照明系統の 8% 程度で稼働時間を考慮すると電力に占める空調分の比率はわずかと考えられ、空調要因分は分類せず人間活動分に含める。52 号館の空調はガス式空調であり、空調に使用するガスを個別に計測している。空調の一次エネルギー消費はガス消費を一次エネルギー消費に換算したものとする。従ってクラスタリングの結果と実測の一次エネルギー消費量は同じである。照明とコンセントは別々に測定されるが、ベースと人間活動両方のエネルギー

ー消費が含まれている。したがって、ベースと人間活動の両方のエネルギー消費を分離する必要がある。

図 3.16 にコンセント系統を合計し年平均したものを示す。この図の 0 時から 5 時の平均値をベース分とし残りを人間活動分とする。図 3.17 には照明系統の年平均を示すが、同様に 0 時から 5 時の平均値をベース分とし残りを人間活動分とする。コンセント系統のベース分と照明系統のベース分、さらにこれに動力系統を合計して 52 号館のベース分とする。図 3.18 に要因別エネルギー消費比率の実測とクラスタリングの比較を示す。したがって、クラスタリングを使用して、大学の建物の一次エネルギー消費データをベースエネルギー消費、人間の活動分と空調要因分に分類できる。

### 3.5 エネルギー管理の活用

本節までに k 平均法を用いた要因別エネルギー消費分析内容の確認を行った。さらに、全学における低炭素化の目標について検討する。前報では 2016 年度に 2011 年度比で 14%の低炭素化を達成したことを報告した<sup>34)</sup>。2030 年度までの削減目標について 2016 年度の実績に k 平均法による要因別分析から考察する。表-4 には 2016 年度の全学の実績から求めた要因別の一次エネルギー消費量を示す。人間活動分については 2030 年度までに LED 化と OA 機器の効率を良くすることで 40%の電力消費量削減を見込む。空調は 2030 年までの機器更新でエネルギー消費量が 15%削減されると考える。エネルギー白書の目標では、電力の排出係数は  $0.37\text{kgCO}_2/\text{kWh}$  であり現在の  $0.489\text{kgCO}_2/\text{kWh}$  より 24%少ない。政府が閣議決定した温暖化対策における民生部門の削減率 40%を達成するには電力で 21%の削減を目標として算出すると、ベースは 17.8%の削減が必要となる。

### 3.6 用途毎の建物のエネルギー消費パターン

#### 3.6.1 事務所施設の特徴

事務所施設は、図 3.19 の日電力模式図に 24 時間消費のベース電力に昼間の空調、照明の設備機器とコンセント電力消費が乗った形となる。

ベース電力と昼間電力の 20%程度となる、ベース電力の消費は昼間電力に比べて小さい。昼間電力は朝の出勤時の前から立ち上がる、昼間のピーク時に昼休みの消灯が行われればその分の電力消費が削減される。昼間のエネルギー消費の増加は、空調、照明、プリンター、コンピューター、ファックスなどのオフィス機器のエネルギー消費とコンセントのエネルギー消費によるものである。

#### 3.6.2 病院施設の特徴

図 3.20 は病院施設の電力消費の模式図である。事務所用途に比べてベース電力の割合が 50%以上と大きいことである。この事例でベース電力と昼間電力の割合が 1 対 1 程度となっており、ベース電力は平日と休日に関わらず年間を通じて消費されるのでエネルギー消



費型の用途となる。ベース電力の要因は、主に医療機器の電力消費と考えられ、昼間は事務等の管理部門の電量消費と考えられる。

ベース電力は医療機器の特徴電力が大きな要因と想定されるが、具体的な節電対策は、医療行為との関連があるので関係者との協議が必要となる。

### 3.6.3 物販施設の特徴

図 3.21 の物販施設の電力消費の模式図では、事務用途と同じくベース電力と昼間電力の差が大きく、この事例では、1 対 10 程度となっている。また、朝の立ち上がりと夕方の立下りが、開店と閉店に伴って電量消費は急上昇と急低下する特徴があり、事務所用途の様に残業の影響が無い。週末も開店しているんで休日が少ないことも特徴であり、年間のエネルギー消費の多い用途となる。

### 3.6.4 宿泊施設の特徴

図 3.22 に宿泊施設の電力消費の模式図を示す。熱源は地域冷暖房を受け入れているので季節間の格差は比較的少なくここでは夏期のものを示している。電量消費の特徴は、朝の外出時と夜の帰宅時にピーク電力を生じているが、夜間の就寝時間以外の電力消費はほぼ一様である。このベース電力消費は昼間の約 60%と大きく、休日が無いのでエネルギー多消費型の用途となる。各室等の多数室の構成となる用途なので、昼間の各室の不在時の空調や照明の停止が大きな省エネルギー項目となる。

## 3.7 クラスタリング手法を用いた大学建物のエネルギー消費パターンの分析

中部大学の異なる建物のエネルギーパターンを分析するために、提案された K 平均法を適用して、中部大学の 3 つの建物の電力消費を調べた。工学部の 3-7-8 号館、経営情報学部の 21 号館と事務系建物 1 号館の電力消費パターンを調べた。工学部には、継続的に実行されるさまざまな研究機器とサーバーが含まれている。これらのエネルギー消費パターンを分析すると、建物のエネルギー消費に対する実験および研究機器の影響を明らかになる。

工学部の建物の場合、図 3.24 のクラスタ 1 で表されるベースエネルギー消費およびクラスタ 1 とクラスタ 6 の午前 0 時から午前 5 時までの間隔も大きいことがわかる。工学部の建物のこの特徴は、文系や事務系建物とは異なる。図 3.25 のクラスタ 1 から 3 で表されているように、非空調日におけるエネルギー消費が大きくなるのは研究機器の影響を示している。

図 3.26 は中部大学の事務系建物である 1 号館のカレンダープロットを示す。事務系建物は、工学部建物のような 24 時間稼働する研究機器はない。したがって、図 3.27 のクラスタ 1 で表されるベースエネルギー消費量は小さいことがわかる。図 3.28 に示すように、クラスタ 1 から 3 のエネルギー消費量が少ないことは、事務系建物に研究設備がないことを示している。クラスタ 4 から 6 で表す空調使用日が多く電力消費消費を占めていること

が分かる。従業員が仕事を始める午前8時ごろに、エネルギー消費は急激に増加する。同様に、午後12:00から午後1:00まで昼食休憩のために、クラスタ2から6まで電力消費は減少する。クラスタ1から6は、午前0時から午前5時に互いに一致する。これは、事務系建物に研究機器がないこともサポートする。

図3.29は経営情報学部の建物のカレンダープロットを示す。図3.30のクラスタ1で表されるベースエネルギー消費量は小さく、午前1時から午前5時までのクラスタ1とクラスタ6の差は工学部に比べて小さくなっている。図3.31に示すように、クラスタ1から3のエネルギー消費量が少ないことは、研究機器が少ないことを示す。事務系建物のようにクラスタ4から6で表す空調使用日が多い電力消費消費を占めていることが分かる。これは、文系のエネルギー消費が大学の事務系建物のエネルギー消費と似ていることを示している。2015年の中部大学の場合、空調冷房期間は6月15日から9月15日、暖房期間は2015年11月15日から2016年4月15日である。中部大学では、中間期間の空調の使用は行われていないが、極端な気象条件への突然の変更およびその他の例外的なケースは許可される。クラスタ1は、通常、日曜日、および祝日や空調が使用されていないときの祝日に当たる。この電力消費は、センサー、サーバー、冷蔵庫のような一年中24時間稼働する研究機器の電力消費である。これは一日に発生する一最小の電力消費を示す。したがって、クラスタ1の1年分の合計を求める場合、以下の式(6)にて算出できる、これは大学のベース電力消費量を示す。

ベースエネルギー消費は、建物のエネルギー消費削減のために重要な課題である。昼間に数時間発生するピークエネルギーとは異なり、ベースエネルギーは24時間全体で発生するため、建物の1時間あたりのベースエネルギー消費を1kWh削減すると、年間ベースエネルギー消費が8,760 kWh削減される。

人間の活動によるエネルギー消費は、照明、コンピューター、OA機器、エレベーターなどの使用によって消費されるエネルギーとして、空調のエネルギー消費は除く。クラスタ2と3はほとんど空調を使用しない日に該当するため、クラスタ2と3からクラスタ1のエネルギー消費を差し引いて合計することで、人間の活動によるエネルギー消費を取得できる。クラスタ4、5、および6は、空調の使用日に該当するため、空調によるエネルギー消費量は、クラスタ3のエネルギー消費量をクラスタ4、5、および6からそれぞれ減算して合計したものである。中部大学では、電気とガスの両方が空調に使用されるため、ガスの一次エネルギー消費も考慮される。

図3.32は、中部大学の3つの異なる建物のエネルギー使用量を用途別に分類したものである。他の2つの建物の場合、空調エネルギー消費量が多いのに対し、工学部のベースエネルギー消費量は多くなっている。式(6)を使用して、工学部、事務系建物、および経営情報学部の建物のベースエネルギー消費量を計算した。図3.33は、工学部の床面積あたりのベースエネルギー消費量が事務系建物の5.9倍、経営情報学部の建物の4.2倍であることを示す。工学部のエネルギー消費量が多いのは、他の2つの建物にはない研究機器と機械が1日

24 時間稼働しているためである。

上記の建物のピークとベースのエネルギー消費量の違いを見つけるために、クラスタ 6 とクラスタ 1 のエネルギー消費量の比率を午前 10 時から午後 17 時まで計算した。図 3.34 に示す箱ひげ図は、これら 3 つの建物のピークエネルギー消費量の違いを示している。工学部、事務系建物、および経営情報学部の建物でクラスタ 6 とクラスタ 1 の間の平均比は 2.5、5.62、4.70 であった。工学部ではベースエネルギー消費が大きいですが、ベースエネルギー消費とピークエネルギーの比率が小さいことは、ベースエネルギー消費の削減が主な焦点であることを示している。文系と事務系建物ではベースエネルギー消費よりもピークエネルギー消費の削減の可能性が高いことを示している。

中部大学の理系と文系の建物のベースとピークエネルギーの関係を見つけるために、追加の建物が含まれ、図 3.35 に示すような散布図を作成した。青い点は文系の建物を表し、赤い三角形は理系の建物を表す。点と三角形の近くの数字は、中部大学のそれぞれの号館の番号である。この図の平均ベースエネルギーは午前 0 時から午前 5 時までのエネルギー消費の平均であり、平均ピークエネルギー消費は午前 10 時から午後 5 時までのエネルギー消費の平均である。文系の建物の  $R^2$  値が小さいため、ベースとピークのエネルギー消費量の関係は見つからなかったが、理系の建物の場合は  $R^2$  値が 0.904 であったため、ピーク日のピークエネルギー消費量がベースエネルギー消費の増加に伴い増加することが分かった。

### 3.8 中部大学のエネルギー消費パターンの年次変化

クラスタリング手法は、年間のエネルギー消費量の分析に使用する。そのため、2015 年から 2018 年まで年度毎のエネルギー消費を計算する。2018 年の空調エネルギー消費量は 2017 年と比較して多くなっている。これは、図 3.37 に示すように、ガス消費量が前年と比較して高いことが原因と考えられる。2018 年の外気温度が前年に比べて高かったため、空調エネルギー消費量が増加したと考えられる。

### 3.9 ガス空調利用ビルと電気空調利用ビルの比較

#### 3.9.1 エネルギー消費パターンの年間変化の分析

各建物のエネルギー特性を調査するために、クラスタリング手法を適用する。ガス空調を使用する 52 号館と電気空調を使用する 20 号館を対象とする。52 号館は生命健康科学部であり、4 階建ての建物で延床面積は 2,137m<sup>2</sup> である。20 号館は国際関係学部であり、その主要な建物で 12 階建てで延床面積は 5,576m<sup>2</sup> であり関連する建物を含め 9,566m<sup>2</sup> である。電力計測はこれらの建物群で行なわれている。

図 3.38、3.39 は、クラスタリング結果の積み上げ棒グラフを使用して推定された原単位あたりのエネルギー消費量を表す。20 号館の 2015 年の延床面積あたりのエネルギー消費量は 2014 年と比較して削減されていることがわかる。エネルギー消費の削減は、ベース、人間活動分、および空調エネルギー消費の削減によるものである。空調エネルギー消費量削

減の理由の1つは、20号館の空調の更新と考えられる。20号館は、実験装置を使用しない文系の建物である。したがって、この建物には研究および実験装置の影響はない。202台の機器のうち190台が更新され、エネルギー消費の削減に貢献した。

空調のエネルギー消費削減に対する電動ヒートポンプ(EHP)更新の効果を調べるため、2014年から2018年にかけて、20号館の空調ピーク日のエネルギー消費量を表すクラスタ6を抽出した。したがって、クラスタ6はクラスタリング結果から抽出され、照明、コンピューターなどによるエネルギー消費を表すクラスタ3から削減された。クラスタ6とクラスタ3の差は、ピーク時の空調の電力消費と考えられる。ピーク空調期間の空調によるエネルギー消費量を図3.40に示す。EHP更新の効果により、空調エネルギーのピーク消費量が20.2%減少したことがわかる。

図3.39では52号館のエネルギー消費は2017年を除いた2014年から2018年にかけて増加したことがわかる。エネルギー消費の増加は、主にベースおよび空調のエネルギー消費に見られる。図3.41および3.42は、20号館のクラスタリング結果のクラスタセンターを示す。それぞれ2014年と2015年はクラスタ6で表されるピーク電力消費量の減少がはっきりと観察できる。図3.43および3.44は、20号館のクラスタリング結果のカレンダープロットを示す。2014年には、クラスタ6で表されるエネルギー消費量が最も高い日は、主に冬季の12月と1月に見られる。2015年には、夏と冬の両方で最もエネルギー消費の多い日が見られる。

図3.45および3.46は、52号館の52号館の電力消費原単位のクラスタリング結果を示す。52号館の2015年と2017年のエネルギー消費量にはそれほど差がないことがわかり。しかし、クラスタ6で表されるピーク日のエネルギー消費のパターンに違いが見られる。52号館の空調はガスのみが使用するため、クラスタリングの結果には空調の効果は現れない。図3.47および3.48に示すように、両方のカレンダープロットで、エネルギー消費量の多い日は主に冬季に表れている。

## 第4章 電力消費予測

### 4.はじめに

電力消費予測とは、需要と供給の均衡を満たすために、将来の負荷挙動を予測することである。電力消費予測は、発電会社や配電会社、建物の管理者などの効率を高めることができる重要なプロセスである。この予測は、電力会社が顧客への供給を運用および管理するに役立つ。負荷予測のいくつかの利点を以下述べる。

- 電力会社とビル管理者は、将来の負荷需要の消費を理解しているため、計画を立てることができる。
- 将来の長期的な負荷を理解することは、企業が将来の発電および送電投資に関して経済的に実行可能な決定を計画し、行うのに役立つため、公益企業のリスクを最小限に抑える。
- 発電所の最大利用率。予測により、発電不足または発電過剰が回避される。
- 電力システムのメンテナンスの決定と計画を支援する。予測により、発電不足または発電過剰が回避される。

東日本大震災後、ピーク負荷の低減と省エネルギーが課題になった。多くの日本の大学や商業ビルは、省エネルギーとCO<sub>2</sub>排出量の削減に取り組んだ。建物のエネルギー管理と需要対応には、正確な電力消費予測が不可欠である。本研究の目的は、既存の予測手法と比較して予測精度を向上させる予測モデルの開発により、大学建物の需要応答、ピーク負荷削減などのエネルギー管理をサポートすることである。

この論文では、大学の建物のより正確なピーク負荷予測を生成するために、クラスタリングとARIMAモデルの組み合わせたハイブリッドモデルを提案した。auto.arima関数を使用することによってARIMAモデルの選択が自動的に行われるため、時系列データのみがある場合、建物のスケジュール、休日の有無、イベントの有無などを考慮することなく、正確な電力消費の予測を行うことができる。

### 4.1 電力消費予測の動機

電気料金は、消費電力量とピーク需要契約値に応じて請求される。電力消費がピーク需要契約値を超える場合、顧客は契約料金が増えるため、この契約値を超えないように協力する。

図4.2は、中部大学西キャンパスにおける2018年7月17日、19日、23日および8月6日の電力消費を示す。これらの日中部大学西キャンパスの電力消費は、図4.2に示す赤い点線で表されるピーク需要契約値を超えた。これにより、過去のデータを使用して適切な精度で電力消費を予測する自動技術が必要となる。中部大学は、東キャンパスと西キャンパス、2つの電力契約地に分かれている。この分析では、東キャンパスの電力消費データのみを使用した。

## 4.2 文献レビュー

Wijaya ら<sup>37)</sup>は、住宅の電力予測の1時間および24時間先までの住宅の電力消費の短期および総合予測のためのさまざまな機械学習アルゴリズムが評価された。予測精度は、スケールに依存せず、ゼロに近い値に対して堅牢な評価メトリックスを使用して評価された：正規化二乗平均誤差(NRMSE)および正規化平均絶対誤差(NMAE)を使用した。異なる世帯の電力消費量を事前定義されたクラスタ数に分類し、各クラスタの総電力消費量の予測値を合計すると、各世帯の電力消費量を個別に予測して合計するよりも良い結果が得られる。自己組織化マップとK平均クラスタリングアルゴリズムのカスケードアプリケーションに基づいて、工業団地のエネルギー消費パターンを分析するデータ処理システムを<sup>38)</sup>示した。このシステムは、スペインの工業団地の実負荷データで検証された。検証結果は、システムが意味のあるさまざまな行動パターンを適切に見つけ、監督なしに、またデータに関する予備知識なしにそうすることができることを示している。密度ベースのクラスタリング手法CHSFDPを実行して、電力消費の典型的なダイナミクスを発見し、顧客をさまざまなグループにセグメント化することを提案した<sup>39)</sup>。動的クラスタリングの結果に対して時間領域分析とエントロピーが実行され、各グループの顧客の需要応答の可能性が特定される。オマーンの住宅および産業部門の電力消費を予測するためのR言語のK平均法適用した<sup>40)</sup>。各クラスタの予測結果はTSI+RWD、TBATS、ARIMAなどを使用して分析され、予測精度は平均絶対誤差と二乗平均平方根誤差を使用して計算した。TBATSモデルを使用すると、より正確な結果が得られる。ただし、この調査は5か月間のデータのみで行われたため、予測を改善するには少なくとも1年のデータが望ましい。<sup>41)</sup>では、季節、電力消費、電力が消費されるフィールドなど、さまざまな基準に基づくDynamicクラスタリングアルゴリズムが使用された。クラスタリング後、負荷需要予測に基づいてDeep Belief Network(DBN)が各クラスタに実装され、その後、長期需要を予測するためにマルチ予測DBNが実装された。2005年の月間エネルギー消費予測は、2004年のデータを使用して実行された。住宅消費電力プロファイルのクラスタリング方法と、住宅顧客のより正確な電力予測の作成に焦点を当てたさらなるアプリケーションの包括的な研究<sup>42)</sup>が示される。さまざまな距離測定を使用した適切なクラスタ数の決定と、母集団全体の集約モデルの予測結果について説明した。Kores ら<sup>43)</sup>は、線形回帰(LR)とサポートベクトル回帰(SVR)を使用して電力需要を予測する方法が提案した。SVRとLRは、データを分類およびクラスタ化するwekaツールを使用して実装する。誤差計算により、提案された方法により精度が向上し、パフォーマンスが向上することがわかる。ボトムアップ短期負荷予測のクラスタリングを<sup>44)</sup>で使用した。分解戦略では、ノンパラメトリックモデルを使用して予測とウェーブレットを処理し、負荷曲線間の類似性のさまざまな表記を定義する。フランスの個々の消費者に適用して、予測精度が16%向上した。

### 4.3 ARIMA モデル

Box と Jenkins は、時系列データをデータに適合させ、予測に適合モデル(つまり、ARIMA モデル)を使用することにより、時系列を予測する数学モデルを開発した<sup>45-46)</sup>。一般に、ARIMA プロセスは  $ARIMA(p, d, q)$  という表記で記述される。ここで、 $p$  はモデル内の自己回帰次数を示す。自己回帰の順序は シリーズの以前の値は現在の値を予測するために使用される。差( $d$ )は、モデルを推定する前に系列に適用される差の順序を指定する。トレンドが存在する場合(トレンドのあるシリーズは通常、非定常であり、静止を想定した ARIMA モデリング)、それらの効果を除去するために、差分が必要となる。移動平均( $q$ )は、前の値の系列平均からの偏差を使用して現在の値を予測する方法を指定する<sup>47)</sup>。パラメーター ( $p, d, q$ ) は、平均と分散に関して静止している静止時系列を使用して常に推定される。

ARIMA モデルの季節性に対処するために、Box および Jenkins<sup>50)</sup>は、適切な次数の季節差分を使用して系列から非定常を除去する一般化された ARIMA モデルである。一次の季節差は、観測と前年の対応する観測との差である月次時系列  $s = 12$ 、および四半期時系列  $s = 4$  の場合、モデルは一般に  $SARIMA(p, d, q) \times (P, D, Q)^s$  モデルと呼ばれる。

### 4.4 研究の対象

図 4.3 は中部大学の地図を示し、赤い境界線はこの研究の対象である東キャンパスの境界線を示している。図 4.4 に見られるように、中部大学の電力消費は年間を通じて変化し、夏季と冬季の空調使用期間の電力消費が最も高くなる。休暇中、電力は大学のベース電力消費の近くで低下することがわかる。1 週間を通して、講義、部分講義(土曜日)、休日(日曜日と祝日)、イベントの発生などの影響により、電力消費の増減がある。

### 4.5 方法論

この研究で使用されるデータは、2017 と 2018 年度の中部大学東キャンパスの 1 時間ごとの電力消費データである。中部大学には、各建物の変圧器で電力消費データを測定する建物エネルギー管理システム、BEMS を使用している。1 分のデータが合計されて毎時データが生成されるため、各建物は 1 年間 8760 時間のデータで構成される。技術的な問題により、BEMS サーバーで収集されたデータには、外れ値と欠損値が含まれることがある。トレーニングデータに外れ値が存在すると、予測出力にエラーが生じる可能性がある。したがって、この研究では、データに存在する外れ値は NA 値に変換され<sup>51)</sup>、R プログラミング言語の zoo パッケージを使用した線形補間を使用して適切な値を代入した。データ収集から電力消費予測までのプロセスを図 4.5 のフローチャートに示す。この論文に使用するデータは、中部大学東キャンパスの建物エネルギーデータである。外れ値の検出と欠損値の代入の後、午前 9 時までの予測日を含む 1 年のデータ(365 日)が作成される。これらのデータは  $365 \times 24$  行列に変換され、その中から、予測日を含む 365 日すべての午前 6 時から午前 9 時までの電力消費データが抽出され、クラスタリングが実行する。予測日の午前 6 時から午前 9 時

までのデータは、クラスタリングにのみ使用され、電力消費の予測には使用されない。この研究では、6つのクラスタが適切であることがわかり、初期重心はパーセントイル法を使用して計算される。クラスタリング結果は、365日間すべてを、類似の電力負荷特性を持つ日を持つ6つのクラスタに分類する。次に、予測日と同じクラスタに属する日を含むデータセットが作成され、予測モデルのトレーニングデータとして使用される。トレーニングデータセットに最適なモデルのパラメーター( $p, d, q$ )は、Rプログラミング言語の `auto.arima` 関数を使用して計算される。次のステップは、Ljung-Box 統計を使用したモデルの残差のテストである。残差がホワイトノイズの場合、モデルは適切である。そうでない場合は、新しいモデルの選択が必要となる。

電力消費予測が予想される日を含めた午前6時から午前9時までの各日の電気消費データをクラスタリングに使用する。ここでは、午前6時から午前9時までのデータは、予測が行われる日のクラスタリングにのみ使用され、電力消費の予測には使用されない。文献レビューセクションで説明したように、クラスタリングとARIMAモデルの組み合わせの既報の研究では、それらのほとんどは顧客グループのクラスタリングと個々のクラスタの集計予測の検索に関連している。予測日のクラスタリングのこの概念は、議論されていない。

K平均クラスタリングを使用して、1年間の電力消費の時間ごとの分布に基づいて、大学の電力消費を6つのクラスタに分類した。図4.6は各クラスタの電力消費のグラフを示す。図4.7は中部大学のクラスタごとの電力消費分布と時間帯をそれぞれ示す箱ひげ図である。クラスタ1からクラスタ6まで電力消費に大きな違いが見られ、クラスタ数の増加に伴ってベースエネルギーも増加することがわかる。午後10時から午前6時までの電力消費はほぼ同じである。午前8時から、午後1時または午後2時頃に電力消費がピークに達するまで急激に上昇することがわかる。図4.6に示すように、各クラスタの電力消費のピークは異なるため、予測日をクラスタ化し、クラスタメンバーを電力消費の予測に使用することで、予測アルゴリズムのピーク負荷予測の精度を向上させることができる。

時系列予測の最も正確な方法を特定することを目的として、多くの比較研究が実施されている<sup>52)</sup>。ただし、調査結果によると、予測手法のパフォーマンスは、使用されている精度測定値によって異なることが示される。適切な精度の尺度は、エラー分布の明確な概要に関する情報を提供する必要がある。Root Mean Square Error(RMSE)、Mean Absolute Percentage Error(MAPE)、Mean Absolute Error(MAE)は、非常に初期の最も一般的な精度測定である。この論文では、これらの精度測定値を予測結果に使用する。

#### 4.6 ARIMA モデルの適合

自己相関関数(ACF)と部分自己相関関数(PACF)は、ARIMAモデルのパラメーター $p$ と $q$ の値を推定するために使用される。図4.8および4.9は、2018年4月2日の電力消費予測のトレーニングデータセットの電力消費データのACFおよびPACF結果を示している。最初の2つのラグに基づいて、最初の2つのスパイクに基づいてAR(2)が可能性である。時系



列の定常/非定常プロセスを識別するには、Augmented Dickey-Fuller test(ADF)を使用する。ADF 検定の帰無仮説は、データが非定常であり、小さな  $p$  値が定常であることを示している。R 言語の `adf.test` 関数を使用した ADF テストを使用すると、データは静止していることがわかった。したがって、ARIMA モデルのパラメーター  $d$  はゼロになるように選択できる。ACF のパターンは、シーズナルな行動の存在を示している。季節的な動作を確認するには、24 と 48 前後のラグの状況を確認する。データが周期性のある場合、周期のラグの自己相関は、他のラグよりもシーズンの周波数の倍数で大きくなる。この論文では、1 時間ごとの電力消費データが使用される。大学のデータの場合、特定の時間における電力消費は、前日の同時の電力消費に類似している可能性がある。したがって、ACF の 24 および 48 に大きなピークが存在することは、データに周期性があることを示している。このデータに最適なモデルを見つけるために、さまざまなモデルを近似し、Akaike 情報量基準(AIC)値が最小のモデルを選択する。AIC は、特定のデータセットの統計モデルの相対的な品質の推定量である。モデルの適合が不十分な場合、結果変数の変動性の真の性質を把握できない可能性があるが、適合しすぎるモデルは一般性を失う。AIC は、これらの欠点を最もよくバランスさせるモデルを選択する方法であり、AIC の値が小さいほど、より良いモデルを表すことができる<sup>53)</sup>。

表 4.4 では、ノンシーズナル部分  $(p, d, q)=(2, 0, 1)$  およびモデルのノンシーズナル部分  $(P, D, Q)=(2, 0, 1)$  で最小 AIC 値があった。次に、R プログラミング言語の `auto.arima` 関数を使用して、トレーニングデータに最適なモデルのパラメーターを計算した。`auto.arima` によって作成された最適なモデルは、「ARIMA(2,0,1)(2,0,1)<sup>24</sup>」で、AIC 値は 2096.4 で、表の AIC の最小値と同じだった。

図 4.10 は、残差 ACF 図、およびヒストグラムのプロットを示している。残差の ACF は有意水準内にあるため、`auto.arima` を使用して選択されたモデルは、トレーニングデータセットに最適である。この研究の目的は、予測の精度を高めるだけでなく、予測プロセスを自動化することでもある。日が経過すると新しいデータがシステムに追加され、これらのデータは将来の予測で使用される。したがって、予測に使用されるトレーニングデータは時間とともに変化する。正確な予測結果得られるには適切パラメーター  $(p, d, q)$  をに選択する必要がある。したがって、他のトレーニングデータセットに `auto.arima` 関数を使用して最適なモデルを選択することにした。ただし、残差がホワイトノイズかどうかを確認するには、Ljung-Box テストを使用する。

予測を自動化するその他の理由には、次のものがある。

- 1)モデルのフィッティングは単純なタスクではなく、ユーザーは時系列モデルのフィッティングに精通していないことが多い。
- 2)自動アルゴリズムは、多くの専門家よりも優れたモデルを作成できることがある。
- 3)多くのビジネスでは、週/月あたり数百の予測が必要であり、迅速な予測が必要となる。
- 4)一部の多変量予測方法は、多くの単変量予測に依存している。

## 4.7 結果

提案されたアルゴリズムを実行するには、予測日のクラスタ番号を決定する必要がある。このためには、予測日に従ってデータセットを選択することが不可欠である。本研究では、2017年4月1日から2019年3月31日までの2年間のデータを使用した。2019年3月1日の電力消費の予測が予想される場合、2018年3月2日から2019年3月1日までの1年間のデータを作成する。予測日を含む1年間の午前6時から午前9時までの電力消費データを抽出する。次に、抽出されたデータに対してK平均クラスタリングが行われ、365日の午前6時から9時のデータを抽出することで $365 \times 4$ 次の行列に変換する。K平均クラスタリングに必要な初期重心は、パーセンタイル法を使用して決定する。1年間のデータ全体が6つのクラスタに分類される。また、各日を同様の電力消費特性を各クラスタに分類する。

図4.11のカレンダープロットは、予測日を含む1年間のデータセットのクラスタリング結果である。2018年7月27日の予測日はクラスタ5に割り当てられている。ARIMA予測の場合、トレーニングデータは予測日の数日前の日付のデータを選択する。このため、異なる電力消費パターンを持つさまざまな日がトレーニングデータに混在し、翌日の電力消費を予測するために使用され、予測には誤差が生成される可能性がある。提案手法の場合、クラスタリング手法を用いて学習データを選択する。これにより、トレーニングデータのメンバーは予測日と同様の電力消費特性を持つようになり、予測結果の精度が向上する。図4.12と4.13は、それぞれ2018年7月27日のARIMAモデルと提案された方法の予測結果である。

黒い線はトレーニングデータセットを表し、予測結果は赤色で示される。前の24日間はARIMAモデルのトレーニングデータセットとして使用されるが、提案された方法では予測日と同じクラスタに属する日のグループが使用する。ARIMAモデルのトレーニングデータセットは24日間(576時間)で構成するが、提案された方法のトレーニングデータは7日間(168時間)で構成する。ARIMAと提案された方法の両方に最適なモデルは、Rプログラミング言語の`auto.arima`関数を使用して計算する。ARIMA手法と提案手法の計算時間はIntel(R)Core(TM)i7-6700 CPU @ 3.4GHz、16 GB RAMを使用して、Rプログラミング言語の`system.time`関数を使用して測定した。ARIMAモデルの経過時間は9.30秒であったが、提案された方法の経過時間は0.45秒で大幅に短くなった。ARIMAモデルの予測結果と東キャンパスデータを使用した提案方法の比較を図4.14に示す。

表4.2から、提案された方法のMAE、RMSE、およびMAPEの値は、ARIMAモデルの値と比較して小さいことがわかる。図4.14の結果、提案された方法のMAE、RMSE、およびMAPEの値は、提案された方法がARIMAモデルよりも優れた結果を生成することを示している。

2018年4月1日から2019年3月31日まで、ARIMAモデルと提案された方法を使用した毎日の電力消費予測が実行された。ARIMAモデルと提案モデルを使用した実際のデータ

と予測データの間の各日の MAPE、RMSE および MAE が計算された。表 4.3 は、曜日ごとに ARIMA モデルと提案された方法を使用した実際の値と予測値の間の MAPE、RMSE、および MAE の 1 年間の平均値を示す。

図 4.15 から 4.17 は、ARIMA と提案方法の間の誤差値を示す棒グラフである。図 4.15、4.16、および 4.17 は、それぞれ ARIMA および提案された方法の MAPE、RMSE、および MAE の値を示す棒グラフである。ARIMA モデルの MAPE、RMSE、および MAE の値は、提案した方法と比較すると、土曜日、日曜日、月曜日、火曜日に非常に高くなることがわかる。ARIMA モデルは、前日の電力消費、休日(日曜日)土曜日、および学園祭などのイベントの発生により発生する可能性がある予測日の前の電力消費の突然の増減に基づいて予測するため、オープンキャンパス、スポーツイベントなどは、予測精度に影響を与える可能性がある。

図 4.18 と 4.19 は、土曜日と日曜日をデータセットから削除した後の 1 日先の予測結果である。図 4.18 と 4.19 は、それぞれ ARIMA モデルと提案方法の 2018 年 9 月 18 日の予測結果を示している。9 月 17 日(予測の前日)に電力消費が減少したため、図 4.18 に示すように、9 月 18 日の予測結果にかなりの誤差が生じた。提案された方法の場合、電力消費特性を持つ同じクラスタグループに属する日であるため、予測日の数日前は予測結果に影響しない。

ARIMA モデルを使用して土曜日と日曜日を削除した場合の実際の値と予測値の間の MAPE、RMSE、および MAE 値は、それぞれ 18.5、311.5、253.7 である。一方、提案された方法を使用した実際の値と予測値の間の MAPE、RMSE、および MAE の値は、それぞれ 2.7、50.5、36.4 で大幅に低くなっている。曜日ごとにグループ化することは、予測にランダムな日のみを使用するよりも少し良い選択である。たとえば、2018 年 1 月 19 日の電力消費を予測するには、同じ曜日に属する前日を収集することにより、ARIMA モデルのトレーニングデータを構築できる。図 4.11 では、2018 年 1 月 18 日の予測日はクラスタ 5 に属している。前の週の同じ曜日に属する日はクラスタ 6、2、および 1 に属していることがわかる。休日を除外し、曜日でグループ化すると、手動で行った場合の予測の精度が向上する。ただし、休日を除外して曜日ごとにグループ化すると、トレーニングデータがクラスタリングの場合と同様の特性を持つことを保証できないため、提案されたモデルに比べて精度は低くなる。したがって、クラスタリングは、類似の電力消費特性に基づいて日をグループ化するための最良の方法である。

#### 4.8 考察

本論文では、クラスタリング手法を用いて ARIMA モデルの性能を向上させ、電力消費を自動的に予測する方法を提案した。提案された方法を使用すると、分析者は、トレーニングデータが所属するグループメンバーを使用するため、前日のエネルギー消費パターン(休日、イベントの発生など)を考慮する必要がない。提案された方法は、建物のエネルギー保存と

需要応答に役立つと考えられる。電力会社は通常、夜間よりも昼間に高い料金を顧客に適用する。ピーク時の電力消費量が多いと、顧客の電気代が高くなる可能性がある。ピーク電力消費に達する前の電力消費の予測は、管理部所、教職員、および学生にピーク電力消費を削減するための戦略を立てるのに十分な時間を提供できる。たとえば、当局はピーク時間前にバッテリーの放電時間、コージェネレーションの使用時間などを決定でき、教員と学生は予測について通知された場合、電力消費を削減するための措置を講じることができる。さらに、この方法は、電気料金の高い時間帯に電気の使用を回避することにより、電気料金を削減するための需要応答に役立つ。したがって、電力消費の予測は、建物の省エネに役立つ。ピークエネルギー消費削減の測定値は、建物ごとに異なる。大学建物のピーク電力消費削減のために採用できるいくつかの一般的な対策を以下で説明する。

#### 4.9 電力ピーク削減の対策

- \* 夏には、空調の設定温度を上げることで、ピーク時の電力消費を大幅に削減できる。中部大学では、空調設定温度は通常 26 度で行われる。したがって、設定温度を 26°C から 28°C に上げることにより、電力消費の少なくとも 10% を節約できる。
- \* 不要な照明をオフする。窓のブラインドを開き、自然光が部屋に入るようにする。
- \* オフピーク時に充電されたバッテリーに蓄えられた電気と、電力消費のピーク時に太陽光パネルによって生成された電気を使用する。
- \* 講義の前後に、講義室の照明を完全にオフにする。
- \* 使用していないコンピューターの電源を切る。
- \* ピーク電力消費時間にわたって(つまり、午前 10 時から午後 3 時を超えて)多くの電力を消費する実験の実施を研究者に依頼する。

#### 4.10 ベースエネルギー削減の対策

ベースエネルギーには、一年中 24 時間稼働する機器が含まれる。冷蔵庫、サーバー、非常灯、換気扇、自動販売機、センサーなどの。ベースエネルギー削減のためのいくつかの対策を以下に示す。

- \* 冷蔵庫の中身の品質に影響を与えることなく、冷蔵庫の設定温度を上げる。
- \* 可能であれば、ピーク電力消費時間中に実験装置の電源をオフにするよう要求する。
- \* 非常灯やその他の照明をエネルギー効率の高い LED ランプに交換する。
- \* 使用していないときは、コンピューター、プリンター、その他の機器のスリープモードをオフにする。
- \* 自動販売機の照明は、日中の最大電力消費時間に消灯する。
- \* 個別の使用や部屋ごとの使用ではなく、電気温水ポット、冷蔵庫などの使用を共通する。

## 第5章

### Shinyを使用したデータ分析と視覚化

#### 5.1 R プログラミング言語

Rは、統計計算とグラフィックスのための言語と環境である。Rは、さまざまな統計およびグラフィカルな手法を提供し、高度に拡張可能である。Rの強みの1つは、必要に応じて数学記号や数式など、適切に設計された出版品質のプロットを簡単に作成できることである。グラフィックスのマイナーなデザインの選択については、デフォルトよりも細心の注意が払われているが、ユーザーは完全なコントロールを保持している。Rは、Free Software FoundationのGNU General Public Licenseの条件に基づいてソースコード形式でフリーソフトウェアとして利用できる。さまざまなUNIXプラットフォームおよび同様のシステム、WindowsおよびMacOSでコンパイルおよび実行される。主に統計学やソフトウェア開発のための環境を必要とする統計学者やその他の実務家を使用するが、RはGNU OctaveまたはMATHLABに匹敵するパフォーマンスベンチマークを備えた一般的な行列計算ツールボックスとしても動作する。

Rの機能は、特別な技術、グラフィカルデバイス、インポート/エクスポート機能、レポートツールなどを可能にするユーザー作成パッケージによって拡張される。これらのパッケージは、主にRで開発され、Java、C、C++、Fortranで開発されることもある。Rパッケージングシステムは、研究者が研究データ、コード、およびレポートファイルを体系化して共有および公開アーカイブするための体系を作成するために、研究者によって使用される。パッケージのコアセットはRのインストールに含まれており、15,000以上の追加パッケージと他のリポジトリが含まれている。

#### 5.2 Shiny

Shinyは、Rから直接インタラクティブなWebアプリを簡単に構築できるRパッケージである。スタンドアロンのアプリをWebページでホストしたり、R Markdownドキュメントに埋め込んだり、ダッシュボードを構築することができる。また、CSSテーマ、htmlwidgets、およびJavaScriptアクションでShinyアプリを拡張することもできる。図5.1は、構成可能な数のピンを持つ「faithful」と呼ばれるデータセットのヒストグラムのプロット例を示している。ユーザーはスライダーバーでピンの数を変更でき、アプリは入力に応答する。

##### 5.2.1 Shiny アプリの構造

Shinyのアプリは、ui.Rとserver.Rという2つのスクリプトに含まれている。スクリプトは同じディレクトリにあり、runAppでアプリを実行できる。スクリプトには以下の3つの機能がある。

- ユーザーインターフェイスオブジェクト(ui)
- サーバー機能

### ● shinyApp関数の呼び出し

ユーザーインターフェイス(ui)オブジェクトは、アプリのレイアウトと外観を制御している。サーバー関数には、コンピューターがアプリをビルドするために必要な指示が含まれている。最後に、shinyApp関数は、明示的なUI/サーバーペアから Shiny アプリオブジェクトを作成する。Shiny の新しいバージョンでは、UI オブジェクトとサーバーが単一のファイルでサポートされているため、アプリ全体をコピーして R コンソールに貼り付けることができる。これにより、実験するコードを他のユーザーと簡単に共有することができる。

### 5.3 視覚化プラットフォームとしての Shiny

この研究では、分析結果の視覚化に Shiny を使用する。Shiny を使用したデータの視覚化および分析プロセスの最初の部分は、分析するデータのアップロードである。これは、[Please upload the CSV file]メニューの[Browse]オプションをクリックして実行できる。Browse ボタンを開くと、アップロードするファイルを選択するためのダイアログボックスが表示される。図 5.2 は、Shiny ツールの画像である

図 5.2 に示すように、視覚化とデータ分析の主要なコンポーネントには、「折れ線グラフとヒートマップ」、「クラスタリング」、「予測」が含まれている。これらのメニューはすべて、図 5.2 に示すように、画面の左側にあるスライダーバーにある。Y 軸の範囲の選択とクラスタ化する建物の選択も可能である。

図 5.3 に示すように、データ範囲の選択オプションから、分析と視覚化に必要な希望の日付範囲を選択できる。次のオプションは、クラスタリング用の建物の選択である。図 5.4 に示すように、「Select the building for analysis」オプションをクリックすると、建物のすべての名前がリストが表示される。スクロールダウンメニューを使用して、目的の建物を選択できる。また、複数の建物を選択することもできる。

図 5.6 は、複数選択機能を使用した両方のキャンパスの選択による、中部大学の東キャンパスと西キャンパスのプロットを示す。図 5.7 に示すように、「Select the range of Y-axis」メニューを使用して、Y 軸の範囲を変更できる。このオプションは、電力消費に大きな差がある 2 つの異なる建物の折れ線グラフをプロットするときに便利である。y 軸の範囲を 0~4000 kWh から 0~2500 kWh に変更したときの折れ線グラフを図 5.8 に示す。「Clustering」ボタンを使用すると、インポートしたデータのクラスタリングを実行し、図 5.9 に示すように、クラスタリング結果をクラスタセンターとカレンダープロットの形式でプロットできる。クラスタセンターグラフとカレンダープロットグラフを分析することにより、年間のどの日に、どのくらいの電力が使用されているかを分析できる。

最後の完成したいオプションは、電力消費予測結果のプロットである。クラスタリングと ARIMA モデルのハイブリッドモデルを使用すると、予測日のエネルギー消費量を非常に正確に予測することができる。予測日の 24 時間予測を図 5.10 に示す。サイドメニューバーの「Forecasting」オプションを使用すると、その日の消費電力を予測できると考えられる。

これは、建物のピーク電力消費低減の計画に役立つと考えられる。

Shiny App が Amazon Web Service(AWS)などの Web ベースのサーバーに実装されると、サーバーアドレスを持つユーザーはだれでもアプリにアクセスして操作できる。リアルタイムデータは 1 時間ごとにサーバーにアップロードでき、予測値と実際の値のリアルタイム比較を実行できる。したがって、ピーク電力消費の削減にかかった努力もリアルタイムで確認することができる。Shiny App と AWS を使用して提案された方法には、

- 教員、学生を含む中部大学の省エネメンバーのグループメールとSNSを作成する。
- 平日の午前9時10分に、AWSサーバーへのリンクを含むメールをグループメンバーに送信する。
- 予測された結果に基づいて、ピーク時の電力消費削減に向けて行動するようグループメンバーに要求する。

Shiny App をウェブベースのサーバーに展開すると、コンピューターだけでなくスマートフォンでもエネルギー消費状況を表示できるようになり、グループメンバーの場所に関係なく Shiny App にアクセスしやすくなる。さらに、建物を選択すると、関心のある建物のエネルギーパターンを表示できる。したがって、Shiny App の適切な使用とグループメンバーの協力により、電力消費の削減に貢献できると考えられる。

## 第6章

### 結論

本論文では、中部大学の電力消費パターンを分析し、大学の建物のエネルギー管理のためのさまざまな方法を提案した。大学の建物のエネルギーを節約するには、いつどのくらいのエネルギーが消費されているかを知る必要がある。エネルギー消費を分析するために、クラスタリングのK平均法を使用した。K平均クラスタリングのパフォーマンスを改善するには、初期重心を選択する必要がある。この論文では、初期重心の選択に経験的累積分布に基づくパーセントイル法を使用した。この方法を使用すると、ランダム初期化方法と比較して精度が向上し、空のクラスタも削除されました。また、この方法は、エネルギー時系列データを構築する場合により良いクラスタを生成できることがわかった。これは、提案された方法で生成されたクラスタのクラスタ内二乗距離が小さく、クラスタ間の二乗和が大きいためという事実によってサポートされた。提案された方法によって生成された結果は試行回数に関係なく同じであったため、K平均アルゴリズムの不確実性も除去された。また、K平均アルゴリズムを実行する前にクラスタの数を決定する必要があり、より良い結果を得るには、適切な数のクラスタを選択することが不可欠である。中部大学の適切なクラスタ数を決定するために、電力消費量を6つのクラスタに分類した。K平均クラスタリングを使用して作成された3~5個のクラスタのカレンダープロットは、大学のスケジュールと一致しなかった。6つのクラスタの場合、クラスタリング結果は大学のスケジュールに似ていた。これは、Rプログラミング言語のclValidパッケージを使用して検証された。したがって、6つのクラスタが中部大学に適していることがわかった。

K平均クラスタリングを使用して、中部大学のエネルギー消費量を、「ベースエネルギー消費量」、「人間活動要因分」、「空調要因分」などの要因別にエネルギー消費量に分類した。この分類は、中部大学の学部および建物に対して行われ、エネルギー消費特性が分析された。工学部のベースエネルギー消費が大きく、クラスタ1と残りのクラスタのベースエネルギーの差が大きいためことがわかった。工学部に似た特徴は、応用生物学部および生命健康科学部でも見られた。中部大学のオフィスビルおよび文系建物では、ベースエネルギー消費が低く、クラスタ1と残りのクラスタのベースエネルギーの差は小さいことがわかった。このように、中部大学のオフィスビルの場合のエネルギー消費パターンは、文系の建物と同様のパターンを持つことがわかった。さらに、理系の建物の場合のピーク日のピークは、ベースエネルギー消費の増加とともに増加することがわかったが、文系の建物の場合はそのような関係は見つからなかった。理系の建物の場合、ベースエネルギー消費の可能性が高いことがわかった。文系および事務系建物の場合、ピークエネルギー削減の可能性が高いことがわかった。

大学の建物のピーク電力消費管理には、電力消費の予測が不可欠である。ピーク電力消費時間より数時間前予測することで、省エネルギーチームは戦略を立て、電力消費を削減するための十分な時間と情報を得ることができる。この論文では、クラスタリングとARIMAモ



デルのハイブリッドモデルを使用して、予測結果の精度を向上させ、電力消費の予測を自動化した。電力消費データのクラスタリングには、午前 6 時から午前 9 時までの予測日を含む 1 年間のデータが使用した。その後、1 年データは、K 平均クラスタリングアルゴリズムを使用して 6 つのクラスタに分類した。予測が属する日のメンバーは、電力消費を予測するためのトレーニングデータとして使用した。予測日のクラスタメンバーが予測と同様の電力消費パターンを持っている可能性は、他のクラスタと比較して非常に高い。予測日と同様の電力消費パターンを持つ日を収集するというこの新しい概念は、ARIMA モデルよりも優れた予測結果を生成することが証明された。提案された方法と ARIMA モデルの予測結果の精度は、MAPE、RMSE、MAE の値を計算して比較した。2018 年 4 月 1 日から 2019 年 3 月 31 日までの 1 年間の予測結果は、ARIMA モデルの MAPE、RMSE および MAE の値が 16.1、265.3 および 193.1 であるのに対し、提案された方法はそれぞれ 5.1、81.7 および 61.3 で大幅に低いことを示している。

提案された方法は、建物内のエネルギーの需要応答にも使用できる。提案のアプリケーションとして、Shiny App を使用してデータ分析および視覚化プラットフォームを開発した。Shiny App を使用すると、ユーザーは目的の建物の予測結果をインタラクティブに表示できる。アマゾンウェブサービス(AWS)などのウェブベースのサーバーにアプリがデプロイされると、サーバーを所有している人は誰でもアプリをアドレス指定し、それと対話する。リアルタイムデータは 1 時間ごとにサーバーにアップロードでき、予測値と実際の値のリアルタイム比較を実行できる。大学の建物の省エネに関連するメンバーの電子メールまたは SNS グループを作成し、その日の予測される電力消費の状態を示すメッセージを平日毎日送信し、ピーク電力消費削減に対処するよう要求することは、エネルギーに役立つと考えられる。