2017

Sequential set-point control of thermostatic loads using extended Markov chain abstraction to improve future renewable energy integration

Ashraf Ghassab Radaideh
Iowa State University

Sequential set-point control of thermostatic loads using extended Markov chain abstraction to improve future renewable energy integration

by

Ashraf Radaideh

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Electrical Engineering (Electric Power and Energy Systems)

Program of Study Committee:
Venkataramana Ajjarapu, Major Professor
Manimaran Govindarasu
Umesh Vaidya
Zhaoyu Wang
Lizhi Wang

The student author, whose presentation of the scholarship herein was approved by the program of study committee, is solely responsible for the content of this dissertation. The Graduate College will ensure this dissertation is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University
Ames, Iowa
2017

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

ACKNOWLEDGEMENTS

I would like to to take this opportunity to express my sincere appreciation and gratitude to my advisor Prof. Venkataramana Ajjarapu for his valuable guidance and support throughout the course of my doctoral studies. His insights and words of encouragement have often inspired and motivated me to do quality research.

My deepest gratitude is due to my program of study committee members: Prof. Manimaran Govindarasu, Prof. Umesh Vaidya, Prof. Lizhi Wang, and Prof. Zhaoyu Wang, for their encouragement, advice, and comments to widen my research from various perspectives.

I would like also to express my special thanks to my family: my parents and my brothers and sisters, for their amazing support and sacrifices.

I would also like to express my thanks to all of my friends for supporting me and encouraging me throughout the whole journey of the Ph.D study, and the department faculty and staff for truly making this whole graduate school experience at Iowa State University a wonderful experience and a memorable one.

Last but not the least, I would like to thank my sponsors Yarmouk University for their precious and valuable support during my Ph.D study.

ABSTRACT

Additional flexible resources are required to achieve resilience and sustainable power systems. Challenges emerged due to the increasing amounts of renewable generation penetrations at both the bulk power system and the distribution sides. System operators are required to deal with higher levels of variable and uncertain power outputs for various time-scales. Moreover, replacing existing thermal units with other inertial-less technologies, make the system sensitive to even small contingencies. Demand-side control is becoming an ingredient part of our future power system operation. Effective utilization of demand-side resources can make the system more elastic to integrate the future renewable plans. To help in resolving these challenges, this work develops a demand-side control framework on the Thermostatically Controlled Loads (TCLs) to support the grid with minimal impacts on customers' comfort and devices' integrity.

The Markov chain abstraction method is used to aggregate the TCLs and describe their collective dynamics. Statistical learning techniques of hidden Markov chain analysis is used to identify the parameters of the resulting Markov chains at fixed temperature set-points. Various sensitivities are conducted to reveal the optimal Markov chain representation. To allow extracting or storing additional thermal energy, this thesis develops an Extended Markov Model(EMM) which describes devices' transition when a new set-point is instructed. The results have shown that the EMM is able to capture both devices' transient and steady-state behaviors under small and large set-point adjustments.

Parameters heterogeneity affects the accuracy of the EMM model. In contrast to what proposed in the literature, more comprehensive heterogeneous parameters are defined and considered. The K-mean clustering approach is proposed in our analysis to minimize the heterogeneity error. Devices are divided into multiple clusters based on the power ratings and cycling characteristics. The results

have shown that clustering highly improves the EMM performance and minimize the heterogeneity errors.

Under temperature set-point control the TCLs' aggregated power experience two main challenges before it converges to the new steady-state value, the abrupt load change, and the power oscillations. This is due to devices' synchronous operations once a new operating set-point is ordered. Such power profiles may cause serious stability issues. Therefore, Model Predictive Control (MPC) with direct ON/OFF switching capability is proposed to apply the set-point control sequentially and prevent any possible power oscillations. The MPC can determine the optimal devices' flow toward the new operating set-point. The results have shown that the proposed modeling and control approaches highly minimize the required switching actions. Control actions are required only during the transition between the set-points and finally converges to zero when all devices reach the new set-point setting. In contrast, the models proposed in the literature require very high switching rates which can cause damage or reducing devices' life expectancy.

The last part of this thesis proposes a dispatching framework to utilize the TCLs' flexibility. The developed modeling and control techniques are used to support the grid with three demand response ancillary services. Namely, spinning reserves, load reduction, and load shifting. The three ancillary services are designed as demand response programs and integrated into the Security Constrained Unit Commitment (SCUC) Problem. Three participation scenarios are considered to evaluate the benefits of aggregating the TCLs in the day-ahead markets.

CHAPTER 1. INTRODUCTION

1.1 General Introduction and Motivations

Current trends of expanding the capacities of renewable energy sources experience a rapid growth worldwide. Many factors have motivated this transition, including fuel-free, clean, and secure sources of power. Other factors like the feasibility of adopting large size capacities with attendant decreases in investment costs also play major roles. These factors have motivated many countries to establish their own renewable energy targets and to work diligently for achieving them [1],[2],[3].

Renewable energy sources are usually treated as non-dispatchable source of power, *i.e.,* sources will supply their maximum available power while avoiding curtailments; a practice required to improve their capacity factor and maintain competitive energy costs. This evolution has affected power system operation and control; higher renewable energy penetration is expected to displace more and more conventional thermal units. Current practices do not require renewable technologies to be equipped with frequency support facilities. This will expose the power systems to operate with lower inertia levels, higher power production variability, and uncertain environments. Therefore, in absence of sufficient grid support, the system can be more sensitive to contingencies and vulnerable to blackouts [4],[5].

System support can be deployed to protect the power systems in various short-term and long-term ancillary services, such as regulation, load following, spinning, and non-spinning reserves [6]. The regulation reserves is required to maintain the real-time balance between generation and loads in the seconds time scale. While the load following reserves are required to compensate for the short-term fluctuations produced by the load or smoothing out the renewable energy power output. Load following reserves are usually provided by the fast acting peaking units which are able to ramp very quickly in the minute time-scale. The Base-case generation such as the coal-fired or

nuclear power plants cannot change their output power very frequently and are not suitable for such applications [7], [8].

Procurement of such ancillary services can be provided solely from the peaking units. However, these units operate on fossil-fuel which has expensive operational costs. Thus, increasing the amount of power capacities provided from these units will incur the system operation additional costs that will be reflected in the electricity price. In addition, expansion of the generation fleet with more fossil-fuel based units is against the environment protection objectives of reducing the carbon emissions footprint [9].

These factors have motivated system operators to investigate additional flexible resources in order to support the existing grid infrastructure. The Federal Energy Regulatory Commission (FERC) in order 755 has motivated the utilities to search for additional resources to support the system flexibility and invest more in clean resources [10],[11]. Demand Response (DR) resources have been introduced recently for utilizing load-flexibility at the aggregate level and benefit the overall system needs. Electric loads once operating and consuming power are considered synchronized to the power systems and their contribution can provide potential support to the system in various time-scales. Reserves extracted from flexible loads are equivalent to contributions from other thermal units, and the aggregated response from DR programs can be more valuable and economically feasible than services provided by other peaking units. Therefore, DR resources are counted as a potential source for future grid flexibility [12].

In DR programs, customers are getting incentives to allow their loads being controlled as specified in their contracts or in the DR program specific rules. The contract also involves the control restrictions and limitation such as the number of load interruptions, time limitations, and the comfort limits. Load-acceptability is facilitated by connecting the end-users loads with fast monitoring and control infrastructure. The smart grids and the recent advancements in communication and control systems at the distribution side, such as the Advanced Metering Infrastructure (AMI) and the programmable thermostats, facilitate the deployment of the DR control and make it feasible even over a wide geographical areas [13].

The most prevailing loads at the distribution networks are the Thermostatically Controlled Loads (TCLs) such as air-conditioners, space heaters, and water heaters [15]. The large number of these devices makes the majority of energy consumption stems from the TCLs operation. For instance, Fig.1.1 presents a residential energy consumption survey made by the Energy Information Administration in 2009. The survey shows that the TCLs have the major energy consumption in the residential buildings of the United States. The dominant TCL category is different based on the climate characteristics of the region. For example, air-conditioners consume most of the energy in the hot and humid climate regions like Florida, whereas space heaters are the largest consumers in the cold region like New York. All of the cooling devices are operated by electricity. However, not all of the heating loads are operated by electricity, other sources of energy such as the natural gas also play a major role. Approximately, 40% of the heating loads in the United Stated operate based on electricity [15]. Accordingly, TCLs in residential buildings can constitute a substantial demand response resources, and their aggregation can provide potential support to power systems.

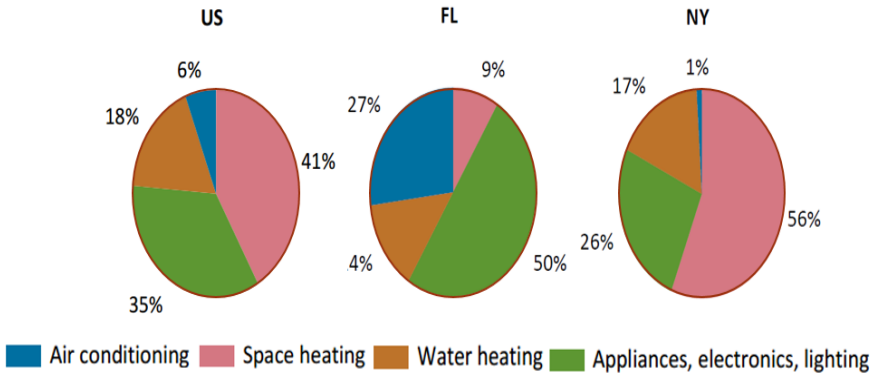

Figure 1.1 Residential energy consumption survey over the entier United States, Florida, and New York [15].

The biggest advantage in utilizing the TCLs is the inherent thermal storage capability. This feature makes them an optimal form of flexible loads, one eligible for extracting considerable amount of reserves by manipulating their temperature set-points. Implementing a demand response program on TCLs requires involvement of ultimately large number of devices, such that the effects of the

control actions will be minimal and will not cause customers discomfort (The indoor temperature will be maintained within acceptable ranges).

DR implementation is still in its infancy stage. Some utilities have started conducting studies and performing pilot projects to leverage the potentials behind utilizing the TCLs. Current implementations are restricted to a small number of customers getting incentives by bill credits or discount rates on a voluntary basis. For instance, the summer discount plan offered by the Southern California Edison Company (CA) uses a radio broadcasting signal to shut down the air-conditioners (A/C) for up to 6 hours a day during emergency events [16]. In this program, the utility allows the customers overriding six requesting signals over the year while receiving a $200 as a bill credits. The Commonwealth Edison Company (ComEd) offers the Peak Time Saving program by installing a wireless switch to control the A/C devices during the summer peaks [17]. The control actions are limited to completely turning the A/C devices OFF or reducing their ON cycling time. The Potomac Electric Power Company (MD) [18] under the Smart Grid Project install advanced programmable thermostats accessible through the internet to turn devices OFF or adjusting the operating ON and OFF cycling times. The City of Ames offers a DR program called "Efficient Air Conditioner Rebate" which allows the utility to control a group of air-conditioning units during the summer peaks. Participants are required to install a device called "Prime Time Power Switch". This device allows the air-conditioners receive a radio broadcasting signals to turn OFF the devices during the summer peaks [19]. There are many other electric utilities offering DR programs on the air-conditioning loads which have similar roles and objectives [20],[21].

The ultimate objective of all DR programs currently implemented is to manage the summer peaks by reducing the customers demand. This demand reduction will help the utilities in avoiding purchasing a high-priced power from the wholesale market and maximizing their own benefits. On the other hand, DR requesting events turn off the devices or reduce the ON cycling time. Accordingly, the indoor temperature will significantly increase, and customers comfort will be lost. DR calling events can be seen as load shedding schemes but only on the TCL devices. The scope of the ongoing research on TCLs is to maintain the power services without interruption while minimally

adjusting the operating set-point within a pre-specified limits. These limits are determined by customers' preference to regulate the temperature in a given comfort zones. Accordingly, the TCLs can be utilized and be more involved in the daily operational aspects while maintaining the power service and customers satisfaction.

1.2 Related Work

Work in literature rely on investigating the capability of the underlying TCLs by modeling the individual devices with a physical model. This model is called the Equivalent Thermal Parameter (ETP) which describes the heat storage and dissipation processes involved in devices operation. Heat storage comprises the heat injected from the heat pump and the heat loss caused by the interaction with the surrounding environment. The dynamics of an individual device constitutes a hybrid non-linear system. *i.e.*, the temperature of the controlled mass is modeled as a continuous variable, while the status of the device is modeled as a discrete variable [22],[23]. Therefore, dealing with such large scale non-linear systems for online applications such as prediction and control could lead to computationally intensive tasks. For these reasons, representing the TCLs requires developing an aggregate model able to describe the operation of all devices in a single computationally tractable model. Moreover, the aggregate model must accurately preserve devices critical information which will be used in the control applications such as the inside temperature dynamics and the total power consumption.

The early work on aggregating the TCLs are not mainly interested in DR control and adjusting the aggregated power. Instead, the aggregated models are developed to predict the the cold-load pickup events experienced by the feeders after prolonged service interruptions. Power outages cause depleting the TCLs internal thermal storage. Therefore, when the power service is restored, devices will be forced by their local thermostats to work simultaneously to bring the deteriorated temperature back to the set-point. This action causes the aggregated power to return to a much higher demand than the value before the outage. In addition, large number of devices stay synchronized

together and for relatively long time periods causing the aggregated power to have an oscillatory profile [24].

The first aggregated TCL model proposed for studying the cold-load pick-up events is presented in [25],[26]. In this work, the analogy of the stochastic diffusion process is used to describe the evolution of the probability density function (PDF) for both the ON and OFF state devices. The resulting system is described by two stochastic partial differential equations (Fokker-Planck equations) coupled at their boundaries. The model is designed and developed to capture the evolution of homogeneous TCL devices. Homogeneous TCL system refer to the case when all devices share identical set of parameters, in contrast to the heterogeneous system, where each device has its own distinct set of parameters. The homogeneous systems is considered in literature as a simplifying assumption since heterogeneity complicates deriving the aggregated model. The authors in [25],[26] have shown that the developed model captures the behavior of the homogeneous system. However, an analytical closed form solution is hard to obtain. In addition, the numerical solution of the resulting model requires discretizing the state and time. Therefore, the accuracy obtained by considering the continuous-state, continuous-time states will be lost while obtaining the numerical solution results.

The authors in [27],[28] have simplified the Fokker-Plank equations by considering the discrete-time, discrete-state, modeling approach. This simplification add a great advantage in transforming the evolution of the coupled probability density functions into a Markov chain. The authors have shown that the Markov chain system accurately captures the homogeneous TCLs behavior and the analytical solution of the model parameter can be easily obtained.

The potential of using the TCLs in DR control applications is investigated in [29],[30]. The System identification techniques of ARX and ARMAX are used to inform about a linear model that links the set-point changes with the aggregated power variations. In this work, 10,000 air-conditioning load are utilized to minimize the variability of wind power output by providing load-following reserves. The developed reduced order model does not provide sufficient information about the behavior of the underlying devices, *i.e.,* devices are seen as a black box. This short-

coming prevents determining the effects of the control actions on devices' distribution and obscure knowing their inside temperature state. Perturbations to the operating Set-point are obtained by implementing the minimum variance control law and the aggregated power in general shows a high level of flexibility.

The set-point control adopted in the previous work is restricted to small magnitudes, *i.e.,* less than the dead-band length, while the resulting optimal set-point adjustment shows high-level of variability. The variations in the set-point are used as a tool to turn ON or OFF certain number of devices as dictated by the desired power trajectories. Accordingly, the set-point perturbation is used to manipulate the percentage of the ON-state devices by switching them ON or OFF. Thus, the set-point control is not used to store or extract thermal energy. This control methodology leads to high switching control actions which could be beyond device's capabilities.

The principles of state queuing theory is implemented in [31] to develop an aggregated model for water heater loads. The model is developed to study the impact of dynamic pricing demand respons programs in modifying the load profile of water heater loads. The state queuing modeling approach holds similar characteristics as the Markov chain abstraction method. *i.e.,* both of the models divide the ON and OFF state devices in the discrete-state and discrete-time settings. The resulting system matrix is in the form of transition probabilities between the discrete temperature states. The state queuing model is developed for homogeneous water heaters while a comprehensive sensitivity analysis on parameter uncertainty is performed.

The authors in [32] develop a new aggregated model for Homogeneous TCLs. A linearized model of the aggregated response is derived using Laplace transformation, and a linear quadratic regulator is used to adjust the temperature set-point. The obtained aggregated model in frequency domain is rather complex and is hard to obtain the inverse Laplace transform. Therefore, the authors use computer software (MATHEMATICA) to expand the model expression in the frequency domain and obtain a closed form solution of the linearized system. The obtained model has similar disadvantages as the one developed in [29]. *i.e.,* both models regard the TCL devices as a black box and devices information are obscured.

The direct ON/OFF switching control is first introduced by the authors in [33],[34],[35], [36]. The authors have utilized the Markov chain abstraction method developed in [27] to implement the ON/OFF switching control and modify the TCLs' aggregated power. The modeling approach is extended for heterogeneous group of TCLs and an analytical derivation of the Markov transition probability matrix is provided. However, devices have limited heterogeneity level. *i.e.,* system is heterogeneous in thermal capacitance but homogeneous with respect to rated power and thermal resistance.

The work in [36] have shown that the homogeneous TCLs experience undamped oscillations and decays while increasing the heterogeneity level. From Markov chain modeling perspectives, it has been shown that there is a negative correlation between the number of states and the oscillation damping. Therefore, a large number of states is considered for modeling the homogeneous system and a small number of states is adopted for the heterogeneous system. *i.e.,* 300-states is chosen to provide a non-decaying oscillation for the homogeneous system, and three models with 40,60 and 80 states are selected for modeling the heterogeneous system. The adopted ON/OFF switching control shows high flexibility in modifying the aggregated power. However, similar to the set-point perturbation control, this technique impose high switching actions on devices which could be beyond device's capabilities.

The authors in [37],[38] have started from the coupled Fokker-Plank equations developed in [25] to build an aggregated model based on the transport load modeling theory. The finite-difference approximation is used to develop a discrete-time, discrete-state, transport model which has similar structure and characteristics as the Markov chain. The model is used for homogeneous TCL system and concluded that increasing the number of the states will improve the model accuracy. Large number of states gives the system the non-decaying nature which is similar to the actual homogeneous system performance (the same conclusion is drawn in the previous work). It has been shown that integrating the set-point control makes the system bi-linear in the state and the control variables. Therefore, a non-linear control method of sliding mode controller that is guaranteed Lyapunov stable has been utilized to provide the load following reserves.

A new modeling approach is proposed in [39]. The authors consider a second-order dynamical model for single TCL devices. These models account for not only the variations in the inside air-temperature, but also captures the temperature of the internal mass. *i.e.,* the average temperature of furniture, walls, carpet, etc. In this work, multiple groups of homogeneous systems are considered to represent the devices heterogeneity and the ON/OFF switching control is utilized to modify the aggregated power consumption. The paper highlights the effects of the resulting switching actions on devices' physical capabilities and proposes a control algorithm to avoid the frequent switching actions on single devices. However, such control restrictions will highly affect the aggregated power flexibility. Conservative control actions *i.e,* devices should operate as the nominal charging and discharging operating cycles, will leave the devices non-responsive.

The formal Markov chain abstraction method is initially introduced in [40] and further described in [41], [42], and [43]. The method is used to derive an analytic error bounds for the resulting models. Markov chains are represented in the form of finite-space stochastic dynamical system to improve the accuracy of modeling the homogeneous TCL system. The work is extended for the heterogeneous TCLs by clustering the system into many homogeneous groups. It has been shown that the stochastic Markov chain model improves the performance of the Markov model and better approximate the TCLs behavior than the deterministic version. The set-point control variation is used in this paper to turn devices ON or OFF to provide load-following reserves. However, the set-point control is similar to the one proposed in [29]. *i.e.,* small set-point adjustments force the devices to turn ON or OFF.

The authors in [44] first propose a non-uniform state transition bin structure to improve the accuracy of the predictions. The advantage of this modeling framework is that the non-uniform states' length can provide accurate results with fewer number of states. This improvement has a good practical implications since it does not require the thermostats to provide a high resolution temperature data to fit the high-state models. The direct ON/OFF switching control is used to provide load-following reserves. The effects of the resulting control actions on devices operation is highlighted in this work. Devices' short cycling protection is implemented using a randomized

priority control strategy to help reducing devices' wear and tear. The authors proposes another control algorithm in [45]-[47] for the same purposes (minimizing the short cycling) by introducing the priority stack algorithm. In addition, the work proposes new methodology to quantify the devices' aggregate flexibility by modeling the TCLs as stochastic battery with dissipation. The power limits of the battery and energy content are characterized as a function of the set-point and the ambient temperature. The direct ON/OFF switching control is the main control strategy in their work.

The work in [48],[49] proposes a safe control protocol to provide services to the grid with minimal subsequent oscillations. The work involves adopting new intelligent thermostat with timers and memory. Devices are instructed to turn OFF or ON for a given amount of time. This work is based on Monte Carlo simulation of devices and does not depend on an aggregate model. The control methodology is suitable for generate sharp power pulses but specific reference power signal is hard to obtain.

The authors in [50] adopt the transport TCL model previously developed in [38] for homogeneous TCLs and extend the work to the heterogeneous TCLs. In this work, the heterogeneous system is divided into multiple clusters such that the devices in each cluster represent an independent homogeneous system. The temperature set-point control is used in this work to modify the aggregated power consumption as in the original work [38].

The TCL devices can also be controlled in a model-free approach. Such control methodology handle the TCLs directly without developing an aggregate model. Extracting power services are achieved by implementing sequential decision-making problems subjected to the TCLs' dynamics. The solution of such methods requires adopting an improved solution algorithm. For instance, the work in [51],[52] propose a generalized battery model solved by a priority-stack-based control. The authors in [54]-[56] use a three-step control approach initially developed in [53] for plug-in hybrid vehicles to control a heterogeneous cluster of TCLs. The developed method utilizes the reinforcement learning to improve the scalability of the problem. Other approaches also consider controlling the TCLs in distributed architectures [57]-[59] to enhance the coordination between

devices and improve the solution algorithms. The main shortcoming of model-free approaches is scalability. Designing control problems subjected to devices non-linear dynamics will make it even harder to solve the problem for more than 1,000 devices. Model-free approaches are entirely based on direct ON/OFF switching control.

1.2.1 Literature Highlights

Modeling Aspects

Various statistical models are proposed in literature to describe the evolution of the ON and OFF-state devices with different state and time settings (continuous or discrete) including: the coupled Fokker-Planck equations, Markov chains, state-bin transition models, state-queuing models, and transport models. All of the statistical models share the same structure, performance, and capability. Other modeling approaches such as the system identification techniques and the Laplace transformation are also used to build an aggregate model. However, the resulting system is in the form of input/output model which only captures the changes in the aggregated power based on the set-point variations. The obtained reduced order model or transfer function obscure devices' information and the effect of the control actions can not be obtained. The major conclusions about the aggregation and TCLs' modeling are as follows:

- Statistical models are derived and identified for homogeneous TCLs. The parameters of the discrete-state models are solved analytically, while the continuous-state models are solved numerically.
- The models are derived and identified at fixed operating temperature set-point. Therefore, the models' parameters will be valid to describe the dynamics only within this temperature range. *i.e.,* the range is defined by the set-point and the dead-band length.
- The performance of the statistical models have shown high levels of accuracy in modeling the homogeneous systems. Where smaller discretization steps (Increasing the number of states) further improves the modeling accuracy.

- Parameters heterogeneity complicates deriving a single statistical model to describe the entire TCLs population.
- Parameters heterogeneity is resolved by approximating the heterogeneous system into many homogeneous clusters. The optimal number of clusters is a trade-off between the modeling accuracy and the computational tractability.
- Model-free approaches are prone to the scalability issues. The control algorithm become highly computationally intensive for large number of devices.

Control Aspects

Most of the work proposed in literature focus on demonstrating the capability of modifying the TCLs' aggregated power to provide fast load-following reserves. The aggregated power consumption at any given time is proportion to the number of the ON-state devices. Therefore, modifying the aggregated power requires regulating the number of devices in the ON-state such that the total power matches the desired value. This is done in literature with two main control techniques.

- Direct ON/OFF switching control: Devices are directly toggled between the ON and OFF states. *i.e.,* devices are controlled remotely to change the current operating cycle either from ON to OFF (load reduction) or from OFF to ON (load increase).
- Temperature set-point control: This control technique requires estimating the set-point magnitude change such that a certain number of devices will be switched ON or OFF. The change magnitude is limited to small variations. *i.e.,* much smaller than the dead-band length. This control limitation is considered for two main reasons. First, the models will not be valid to describe the TCLs dynamics since it is designed and identified at fixed set-point. Second, the aggregated power stability will be lost (large oscillations). Accordingly, the set-point control is used as a tool to switch devices either ON or OFF and the inside temperature of all devices is kept around the set-point.

  Both of the control techniques are similar, in essence, they rely on frequently switching the devices between the ON and OFF states and the average temperature inside houses is kept

as specified by the set-point. Accordingly, to extract certain services both of the control techniques impose high switching rates and short cycling on devices. The resulting control action could be way beyond the physical capability of the devices.

1.3 Proposed Approach and Contributions

Work Overview

In our work, a novel modeling approach is developed to capture the TCLs dynamics under various operating temperature set-points. The model is derived as an extension to the Markov chains developed in literature at a fixed temperature set-point. The new model is referred as the Extended Markov Model (EMM). The EMM is considered as comprehensive framework which can capture not only the steady-state dynamics but also the transient behavior in case of set-point adjustments. The model is designed to describe the dynamics for small and large set-point changes and in both directions. This modeling approach is proposed to minimize the switching actions on the devices by relying on extracting or storing thermal energy rather than relying on their switching capabilities.

The EMM is derived by restructuring and linear mapping of various Markov chains identified at fixed set-points. The development of each Markov chain is performed using statistical learning techniques. The learning process and its characteristics are described and evaluated for both the homogeneous and the heterogeneous TCLs. Extensive sensitivity analysis revealed that modeling the heterogeneous TCL system is subjected to the bias-variance trade-off and there is a specific Markov design to approximate the heterogeneous TCL system. sensitivity analysis is conducted to find the best single Markov chain to approximate the heterogeneous system.

We have found that the heterogeneous parameters previously defined in literature is limited. More comprehensive and general heterogeneous parameters are considered in our analysis. However, for such situations single Markov chains will have larger errors. Therefore, the K-mean clustering is used to divide the comprehensive heterogeneous system into multiple semi-homogeneous clusters.

Dispatching the TCL devices with new thermal energy level forces them to operate in a synchronized patterns and for relatively long time periods. This synchronization leads to large power oscillation for small set-point adjustments. While for the large changes, devices are forced to be either OFF with zero power consumption or ON with maximum power consumption. To avoid these situations, a model predictive controller with ON/OFF switching capability is proposed to determine the optimal flow toward the new temperature set-point and curb down the oscillations. Therefore, the control framework adopted in our work is hybrid. *i.e.,* the set-point adjustments and the ON/OFF switching control.

Major contributions

- A comprehensive modeling framework is developed to capture the TCLs dynamics under various operating temperature set-points. The model is able to describe the transient and steady state behavior associated with small and large set-point adjustments and in both directions.

- The Model Predictive Control (MPC) framework is proposed to resolve the problems associated with adjusting the temperature set-point. The MPC utilizes the ON/OFF switching capability to determine the optimal sequential set-point control law and curb down the power oscillations.

- Demand Response dispatching framework is proposed for utilizing the TCLs flexibility in three main ancillary services; Namely, spinning reserves, load reduction, and load shifting. The DR programs are design and integrated to the Unit commitment problem with appropriate constraints to allow devices charging and discharging without violating customers' comfort.

CHAPTER 2. THERMOSTATICALLY CONTROLLED LOADS (HOMOGENEOUS V.S. HETEROGENEOUS)

2.1 Introduction and Overview

This chapter introduces the Equivalent Thermal Parameter (ETP) model of individual Thermostatically Controlled Load (TCL) and describes the basic set of parameters used to represent its physical operation. A large group of TCLs is considered in this chapter where each device is represented by an independent ETP model. The aggregate behavior of all TCLs at fixed temperature set-point is described and analyzed. Two groups of devices are considered. First, the homogeneous TCLs, which refers to the situation where all devices under control have identical thermal parameters and power ratings. While the second case, is the heterogeneous TCLs, which indicates the case where each device in the group has distinct parameter values.

The sensitivity analysis is conducted on the aggregated power consumption and the distribution of the devices when both of the TCLs groups are subjected to temperature set-point adjustment. It has been shown that the aggregated power of the homogeneous system shows an un-damped oscillatory profile, in contrast to the heterogeneous case, where the power oscillations converges to a steady-state value. The last section discuses a control methodology to curb down the power oscillation observed in the homogeneous TCL devices. The methodology requires adopting an intelligent thermostat which is able to assign new set-point and the dead-band limits as control variables.

2.2 Equivalent Thermal Parameter Model

This section discusses the Equivalent Thermal Parameter Model (ETP) of a single TCL device. The ETP model described in this section illustrates the operation of cooling TCL devices, more specifically, air-conditioning loads. In general, the ETP model describes how the temperature

trajectories for a given device evolve with time during the cooling and the heating cycles. In discrete time settings, the ETP is shown in (2.1)-(2.2) [27]. This kind of mathematical models is described by a hybrid-state system, it comprises both continuous and discrete variables. The continuous state $T_i(k)$; reflects the air temperature inside the house i at the time instant k, while the binary state $q_i(k)$; represents the operational status of this device. *i.e.*, being ON ($q_i(k) = 1$), or OFF ($q_i(k) = 0$). The actual time can be found based on the discretization time-step h.

The temperature set-point T_s, represents the regulating temperature inside the house, while the actual switching limits are specified by bang-bang or a hysteresis controller with a dead-band of length D. The air-conditioning device will be switched ON when the inside temperature increases to the upper dead-band limit $T_s + D/2$, and will be switched OFF when the inside temperature is cooled to $T_s - D/2$.

$$T_i(k+1) = e^{\left(-h/R_iC_i\right)}T_i(k) + \left(1 - e^{\left(-h/R_iC_i\right)}\right)\left(T_a - q_i(k)R_iS_i\right) \tag{2.1}$$

$$q_i(k+1) = \begin{Bmatrix} 1 & T_i(k) > T_s + \left(D/2\right) \\ 0 & T_i(k) < T_s - \left(D/2\right) \\ q_i(k) & otherwise \end{Bmatrix}. \tag{2.2}$$

The ETP model considers two simplifying assumptions. First, the effect of noise processes such as sun exposure and door opening are insignificant to have major effects on the air temperature trajectories. Thus, devices' internal temperature is affected by the heat injected from device's rating power S_i, and the heat dissipated to the outside environment. The heat dissipation process is determined by the thermal characteristics of the buildings. *i.e.*, the thermal resistance R_i, and the thermal capacitance C_i. In the second assumption, the outside temperature T_a is considered as a time-invariant factor. Following these two assumptions, the power consumption of individual device form a periodic square waveform, and the inside temperature trajectories are regulated as specified by the set-point and the dead-band length.

As an illustration, Fig. 2.1 presents the temperature trajectories and power consumption of three identical air-conditioning devices with parameter values as listed in Table. 2.1. The devices

start from different initial conditions. *i.e.,* the initial inside temperature $T_i(k=0)$, and the initial status of the device, being ON or OFF $q_i(k=0)$.

Table 2.1 Air-conditioning ETP parameters.

Parameter	Value
T_s, Temperature set-point	20 $^\circ C$
D, Thermostat dead-band	1 $^\circ C$
S, Power Rating	5.6 kw
R, Thermal resistance	2 $^\circ C/kw$
C, Thermal capacitance	2 $kwh/^\circ C$
T_a, Ambient temperature	32 $^\circ C$

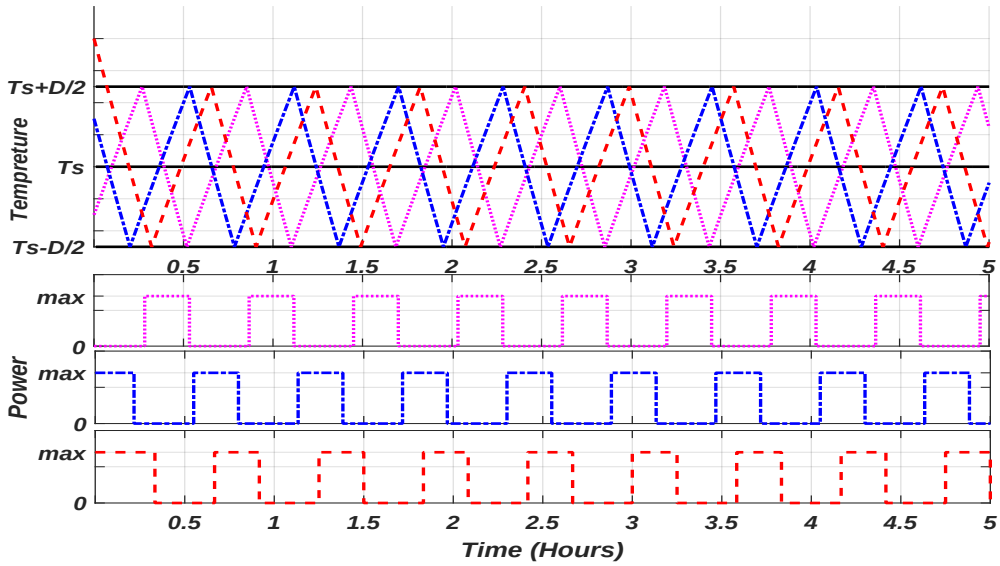

Figure 2.1 Top: Temperature trajectories. Bottom: Power consumption. Temperature and power consumption (graphs are color coded).

2.3 Homogeneous TCLs Performance

The analysis in this section considers a homogeneous group of TCLs represented by air-conditioning loads. All devices share identical set of parameters as listed in Table. 2.1. In some situations, devices might have high level of similarity. For instance, if the TCLs are representing a group of refrigerators in a residential area, the values of devices' thermal resistance and thermal capacitance

might be close. The parameters in this case will follow relatively narrow distributions. In this section, we assume the extreme situation, where all the devices are identical. It is also assumed that all devices are regulated at a common temperature set-point and is remotely accessible by a load-aggregator through a direct load control DR program.

This analysis considers a group of 10,000 device and initially operated at a fixed set-point of 20 $^{\circ}C$. The dynamics of each device is governed by Eqs. (2.1)-(2.2) and their parameter values are listed in Table 2.1. Initially, the temperature inside houses is assumed to be uniformly distributed over the dead-band, while their initial status *i.e.,* being ON or OFF, follows a Bernoulli distribution with success probability equal to the ON time duty-cycle (Eq.(2.5)). These initial conditions are required to approximate devices' steady-state distribution. Otherwise, the aggregated power consumption of all devices will have oscillations. The ON and OFF time periods, Eq.(2.3) and Eq.(2.4) respectively, can be solved for the dynamical system given that the initial and final temperature states are the upper and lower dead-band limits. The aggregated power P_{agr} can be obtained by adding up the power consumption time-series of individual devices as given in Eq.(2.6), where N_d is the total number of the air-conditioning devices.

$$t_{on} = R_i C_i \, ln\Big\{\frac{S_i R_i - T_a + T_s + 0.5D}{S_i R_i - T_a + T_s - 0.5D}\Big\} \simeq 15\, min \tag{2.3}$$

$$t_{off} = R_i C_i \, ln\Big\{\frac{T_a - T_s + 0.5D}{T_a - T_s - 0.5D}\Big\} \simeq 20\, min \tag{2.4}$$

$$P_{on} = \frac{t_{on}}{t_{on} + t_{off}} \tag{2.5}$$

$$P_{agr}(k) = \sum_{i=1}^{N_d} q_i(k) S_i \tag{2.6}$$

The simulation results for set-point change are presented in Fig. 2.2. Initially, the aggregated power consumption is almost a constant value following the initial probability assumptions. First, we consider extracting instantaneous demand reductions by increasing the set-point by 0.3 $^{\circ}C$

at $t = 5h$ (The set-point adjustment signal is shown in black in the bottom graph). Immediate load reduction is achieved. However, the control action is followed by a large undamped power oscillations. Although this temperature adjustment is small, it is considered as a large disturbance to the system and make it unstable.

For a better illustration, consider the distribution of devices before the control action as shown in Fig. 2.3. For simplicity, assume that the number of the ON and OFF state devices are equal. knowing that the upper and lower dead-band limits move along with the set-point adjustments, a small set-point increase moves the lower limit upward to a new value represented by T_1. Accordingly, all ON state devices whose temperature state fall between the old and the new lower limits will be switched OFF by their local thermostats (their inside temperature is less than the new dead-band limit). The number of devices turned OFF is proportional to the magnitude of the applied control signal. These switched devices will stay synchronized with others who are originally in the OFF-state since they have identical thermal characteristics. This synchronization causes large power oscillation as clearly seen in the top graph of Fig. 2.2. The synchronization is more obvious in the bottom graph. The temperature trajectories of these random houses come closer together after the set-point adjustment.

In the second case, we consider extracting instantaneous load increase by decreasing the set-point back to its original state at $t = 15h$, this set-point change causes additional devices to be synchronized together, and thus the amount of power oscillation is also increased. The opposite interpretation apply in this case, where decreasing the set-point forces all OFF devices whose temperature state fall between the old and the new dead-band limit (T_2) to switch ON and stay synchronized with the ON state devices.

The step adjustments completely deform the steady-state Probability Distribution Function (PDF) (shown in Fig. 2.4). This PDF represents only the ON state devices. Initially devices are uniformly distributed over the dead band ($19.5 \leq T \leq 20.5$). The control action cause spikes and gaps while PDF evolves with time which represents devices' synchronization. After the second control action, the gap is increased which means devices are more synchronized. For Multiple set-

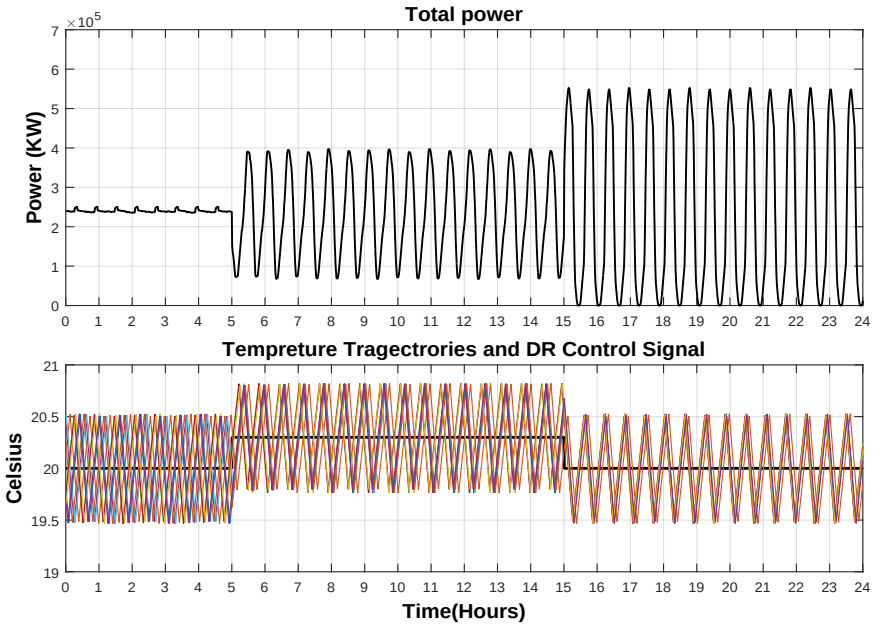

Figure 2.2 Set-point adjustment of 10,000 homogeneous air-conditioning device. Top: Aggregated power. Bottom: Temperature trajectories of ten houses.

point changes, devices will be forced eventually to converge to a dirac delta function, hence the aggregated power consumption waveform will become similar to the operation of a single device, i.e., a square waveform. In summery, temperature set-point adjustments for homogeneous TCL systems force a large number of devices to operate in a synchronized fashion, devices stay in sync since they have identical parameters. This trend is reflected in the aggregated power behavior as a large un-damped oscillations.

2.4 Heterogeneous TCLs Performance

This section discusses the effect of varying the thermal parameters and the energy transfer rates among devices. Heterogeneity has a great positive effect in improving the dynamics of the aggregated power following set-point adjustments. Heterogeneous TCLs implies that each device has distinct charging and discharging time-constants. Therefore, prolonged concurrent operations are less likely to occur, and the aggregated power consumption experience a damped oscillation.

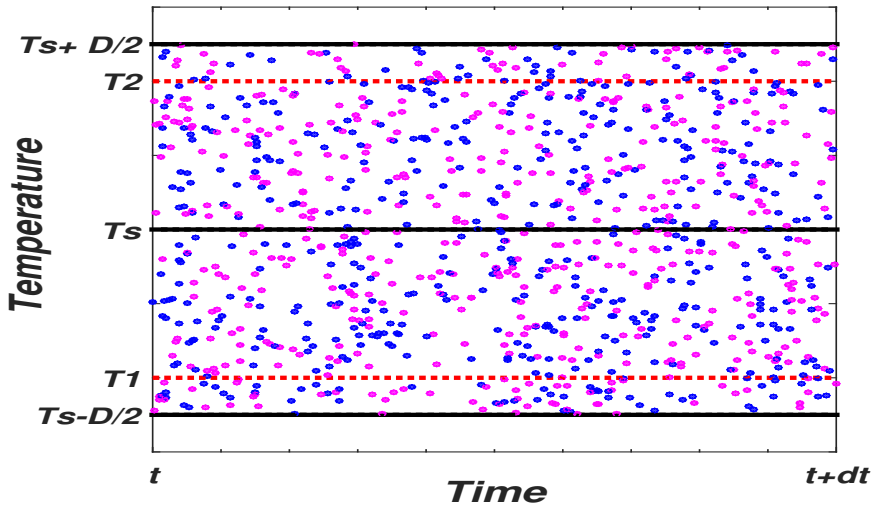

Figure 2.3 Temperature and functional states of a random sample. Blue: heating cycle moving upward. Pink: cooling cycle moving downward.

The same number of devices is considered in this case. However, the parameter values follow Gaussian distributions with mean values equal to those given in Table 2.1, and a standard deviation equal to 0.1 of their means. Heterogeneity is considered in thermal resistance, thermal capacitance, and the devices ratings.

The same control actions applied to the homogeneous system is adopted here for the heterogeneous case. Fig. 2.5 illustrates how the total power consumption oscillatory transients appeared in the homogeneous TCLs has eventually suppressed for the heterogeneous system, but with a relatively long time period until it completely converges to a steady-state value. Temperature trajectories of the random houses show that, following both of the set-point changes, temperature trajectories come closer (synchronized) but then retrieve the diverse situation (not synchronized).

The evolution of the ON-state PDF (shown in Fig. 2.6) requires a relatively long time to converges a new steady-state condition and be ready for the next control actions.

2.5 Damping Power Oscillation in TCLs.

This section describes a control methodology that maintain the stability of the aggregated power consumption by curbing down the oscillations. The new control architecture can be utilized

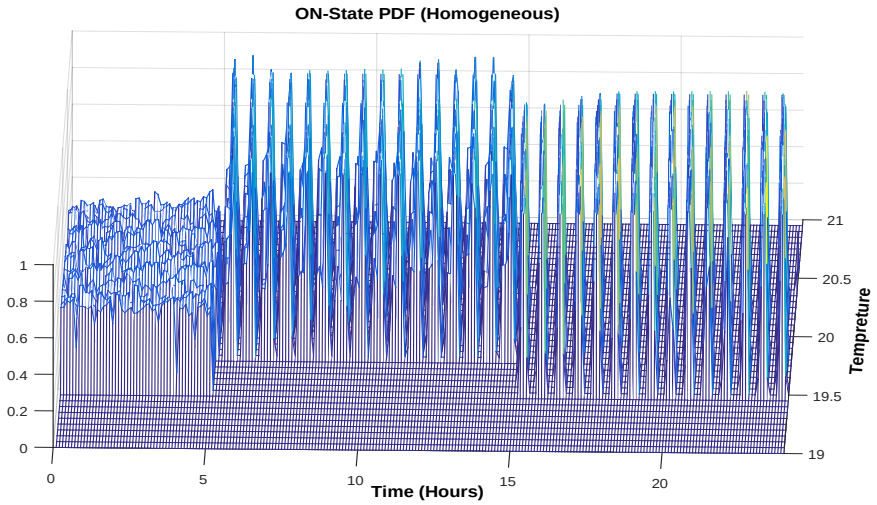

Figure 2.4 The evolution of the ON-state PDF following the set-point adjustment.

with both homogeneous and heterogeneous loads. Actual devices in the real life are heterogeneous. However, we will start from the worst case situation where all devices are homogeneous.

Thermal mass characteristics and power rating variations help creating distinct heat charging and discharging rates. Consequently, after set-point changes, synchronous operations are less probable. For a group of homogeneous TCLs, creating distinct charging and discharging rates can be forced by imposing little variation in the deadband size across the devices. To demonstrate this capability, we assume that the devices are equipped with an intelligent thermostat which is able to provide the set-point as well as devices' dead-band as accessible control variables. In our simulation, we consider that the new assigned upper and lower dead band length are randomly sampled from Gaussian distribution.

To test the control methodology, we considered extracting the maximum available power for a short time duration. This can be done by applying a global set-point control signal. The magnitude of the applied control is equal to twice the length of the deadband, a step-increase of $2°C$ implies that all the devices in the control group will be switched off. Similarly, for a step-decrease of $2°C$ all the devices will be switched ON. Fig. 2.7 illustrates the results where a maximum demand reduction is achieved and the new thermostat able to curb the oscillation. Similar characteristics govern the

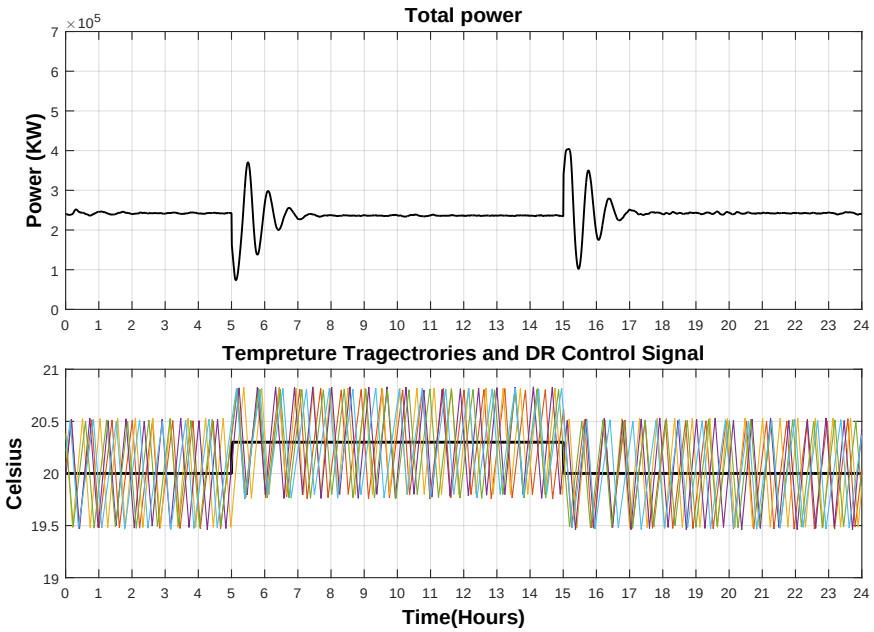

Figure 2.5 Set-point adjustment of 10,000 heterogeneous air-conditioning device. Top: Aggregated power consumption. Bottom: Temperature trajectories of ten houses.

$2°C$ set-point decrease. The effect of dead-band variation can be clearly seen in Fig. 2.8. Sensitivity analysis has shown that in order to completely curbing down the oscillation, it is required to assign a relatively narrow distribution to the new dead-band variations.

Finally, if we consider that demand changes are based on a time varying control signal, the new control methodology improves the aggregated response compared to the case when conventional thermostat settings are considered (See Fig. 2.9). The aggregated power consumption and the demand response control signal are highly correlated and suitable for applying open-loop control system. The limitation of this control architecture is that the stability of the aggregated power will be guaranteed only if we shortening the dead-band length. As consequences, devices will be forced to operate in much shorter duty-cycle with high switching rate.

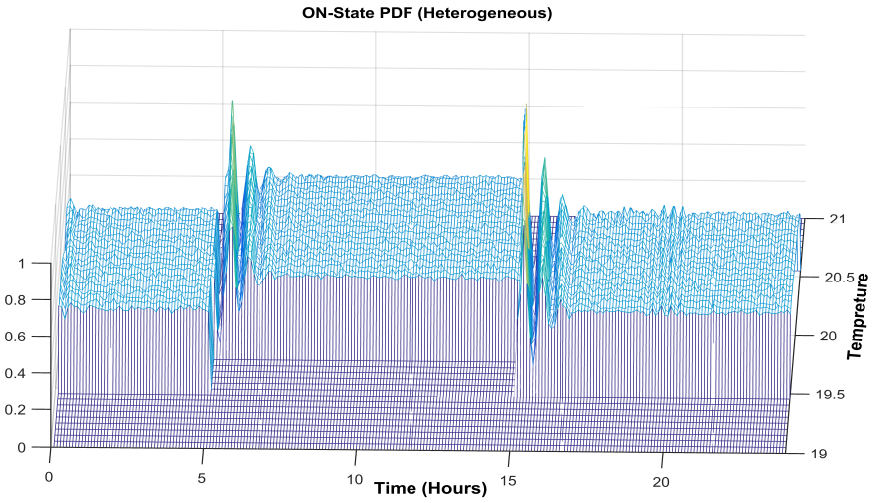

Figure 2.6 The evolution of heterogeneous TCLs (ON-state PDF) following a step-change.

2.6 Conclusion

The chapter shows that manipulating the temperature set-point of the TCLs forces a large number of devices to operate in a synchronized pattern. This synchronization causes oscillations in the aggregated power which may lead to serious stability issues. The oscillation is undamped in case of homogeneous loads, while experience a slow damping rate in the heterogeneous case. The variations in the thermal mass characteristics and power rating for heterogeneous TCLs help creating distinct heat charging and discharging rates for each device. such that, the devices will not be synchronized for long time-intervals and eventually power oscillation is suppressed.

In both cases, set-point adjustment causes oscillation. Therefore, additional control techniques are required to prevent this oscillation and guarantee a fast convergence to the new steady-state value. The chapter introduces an intelligent thermostat for theses purposes. The new thermostat is able to adjust the dead-band limits along with the operating set-point so that each device will have distinct charging and discharging rates. However, maintaining the stability of the system requires forcing the devices to operate with much shorter duty cycle and increasing their switching rates.

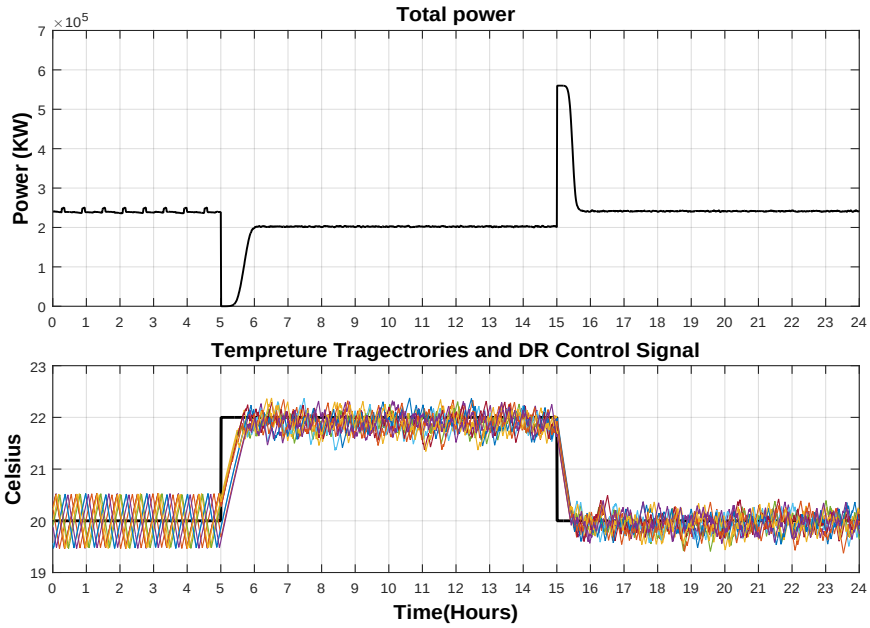

Figure 2.7 Advanced thermostat capabilities. Top: Aggregated power consumption. Bottom: Temperature trajectories of ten houses.

This highlights the need of developing an aggregate model to simplify the large-scale non-linear models into a more computationally tractable single model. This aggregate model will be used to derive an optimal control actions to achieve the services without violating the physical limitation of devices and the comfort of customers. This will be discussed in details in the next chapter.

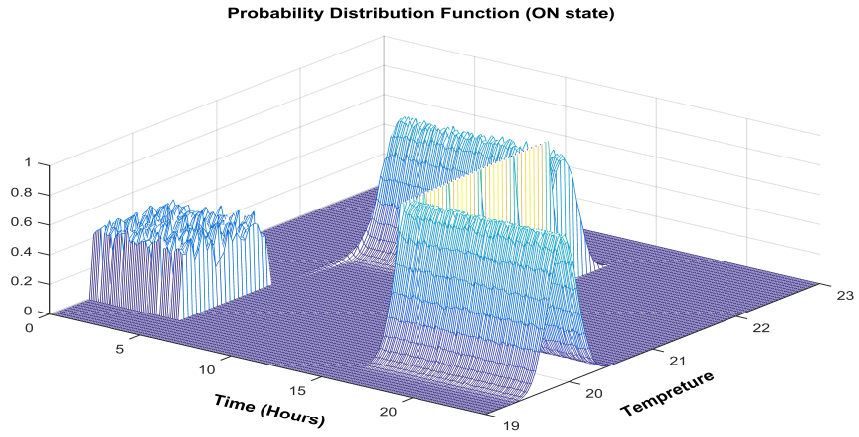

Figure 2.8 ON-state PDF evolution with the new control features.

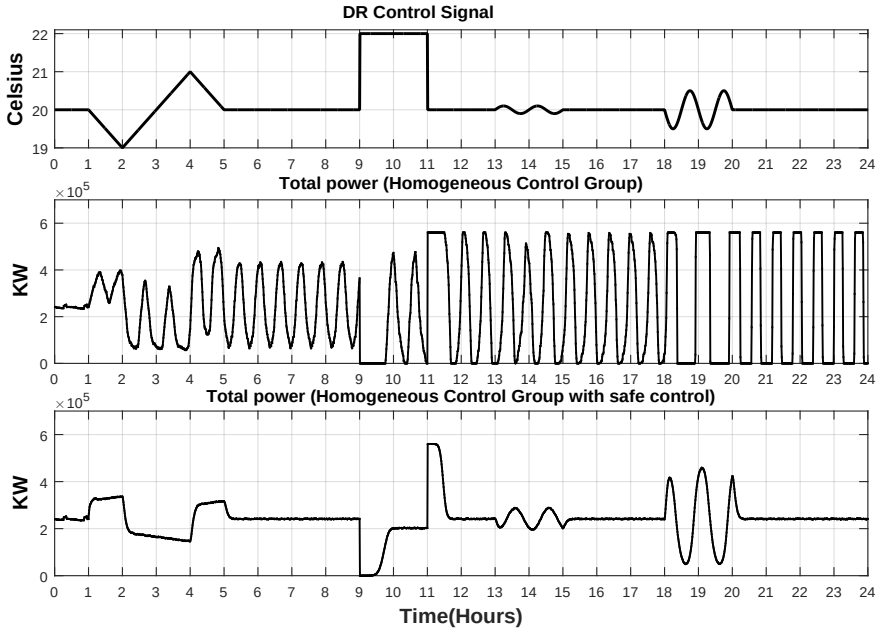

Figure 2.9 Response to a time varying DR control signal. Top: DR refrence signal. Middle: Aggregated power consumption with conventional thermostat. Bottom: Aggregated power consumption with advanced thermostate features.

CHAPTER 3. MARKOV CHAIN ABSTRACTION FOR AGGREGATING THE TCLS

3.1 Introduction and Overview

This chapter describes the fundamental aggregating methodology of a group of Thermostatically Controlled Loads (TCLs) at a fixed temperature set-point. The aggregated model is in the form of discrete-time, discrete-state Markov chain, which describes the flow and the transition probabilities of devices along the dead-band. This modeling framework can transfer the large-scale non-linear ETP models to a single model in linear systems settings. Markov chains will be used as the basis for predicting and controlling the aggregated power consumption of the TCLs in the online applications. The details of this analogy is described in Section 3.2.

Markov chains are developed in this chapter using statistical learning techniques of hidden Markov model analysis. The training data-sets are obtained using Monte Carlo simulation of individual ETP models. The data sets are used to develop various Markov chains with different number of states. Comprehensive sensitivity analysis is conducted in section 3.3 pertaining to Markov model performance against the number of the states and various initial conditions. Conclusions are drawn regarding the Markov model performance and accuracy levels for both the homogeneous and the heterogeneous TCLs. The eigenvalue analysis of the resulting transition probability Matrix is used in Section 3.5 to derive devices' distribution during the steady-state conditions. Finally, Markov models at different temperature set-points are derived in Section 3.6, and the steady-state power is evaluated using the eigenvalue analysis.

3.2 Markov chain Representation

This section discusses the details of developing a representative Markov model for a large group of air-conditioning loads. The dynamics of each individual device are described by (2.1-2.2), where

devices are regulated at a common temperature set-point. A common practice in dealing with hybrid-state models is to discretize the continuous state space, *i.e.*, the temperature space, into a number of equal intervals. In each interval, devices could be either ON or OFF, so that the intervals are grouped into two sequences of states as shown in Fig. 3.1. The nature of the temperature space of being bounded by a hysteresis loop makes the Markov chain abstraction a good candidate for representing the system dynamics as a linear system (3.1).

If we consider discretizing the temperature space into $(N/2)$ intervals, then the Markov chain has a total number of (N) states, such that device could be either in the OFF states $(1 - N/2)$, or in the ON states $(N/2 + 1 - N)$. The system-matrix A defines the transition probability between the states themselves or with adjacent states as demonstrated in Fig. 3.1. The aggregated power drawn from all devices can be estimated by adding up the average power consumption of all devices in the ON states as defined in (3.2-3.3). Where η is the average efficiency, and $\bar{S}$ is the average power rating [27],[33]. Usually in the probability theory, Markov states are defined as the probability distribution and how these distribution evolves with time. In our work, we multiply the distribution with the total number of devices, such that the physical meaning of the markov states become the number of devices at each temperature discretization.

$$x(k+1) = Ax(k), \quad A \in R^{N \times N}, \; x \in R^N \tag{3.1}$$

$$P_{agg}(k) = Cx(k), \quad C \in R^{1 \times N} \tag{3.2}$$

$$C = (\bar{S}/\eta)[\mathbf{0}_{(1 \times N/2)} \, \mathbf{1}_{(1 \times N/2)}] \tag{3.3}$$

A conceptual overview of the Markov model development process is shown in Fig. 3.2. The process involves deriving the transition probabilities by performing a training process over a set of simulated data, which is obtained using Monte Carlo simulations of the ETP models (steps 1-2). The simulated data is in the form of inside temperature and power consumption time-series data for

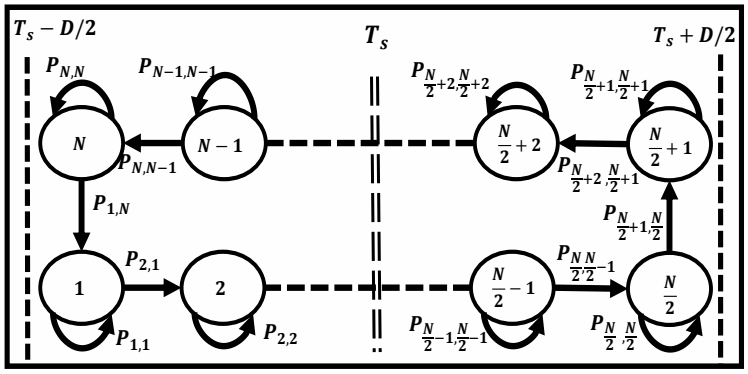

Figure 3.1 Markov chain representation for aggregated TCLs.

individual devices (step 3). For a given Markov model design, the temperature space is discretized with an increment value equal to $\Delta x = 2D/N$, which is used to assign appropriate state number for each temperature interval (Step 4). Markov state definitions and the link with the corresponding temperature intervals are shown in (3.4) and (3.5). The temperature time-series is then compared with the power consumption time-series such that the ON and OFF temperature intervals can be identified (step 5). As a result, the temperature intervals can be transformed into a sequence of Markov states as specified in (3.4- 3.5) (step-6).

$$x_m^{off} := T_s - \frac{D}{2} + (m-1)\Delta x < T_m \leqslant T_s - \frac{D}{2} + m\Delta x \tag{3.4}$$

$$m \in [1, ..., N/2]$$

$$x_n^{on} := T_s + \frac{D}{2} - (n - \frac{N}{2})\Delta x < T_n \leqslant T_s + \frac{D}{2} - (n - \frac{N}{2} - 1)\Delta x \tag{3.5}$$

$$n \in [(N/2) + 1, ..., N]$$

Representing the inside temperature as Markov state sequence is considered the major part of the learning process. The next step involves constructing states counters to measure the total number of visits in which the device remained hold or made a transition to another state. Thus,

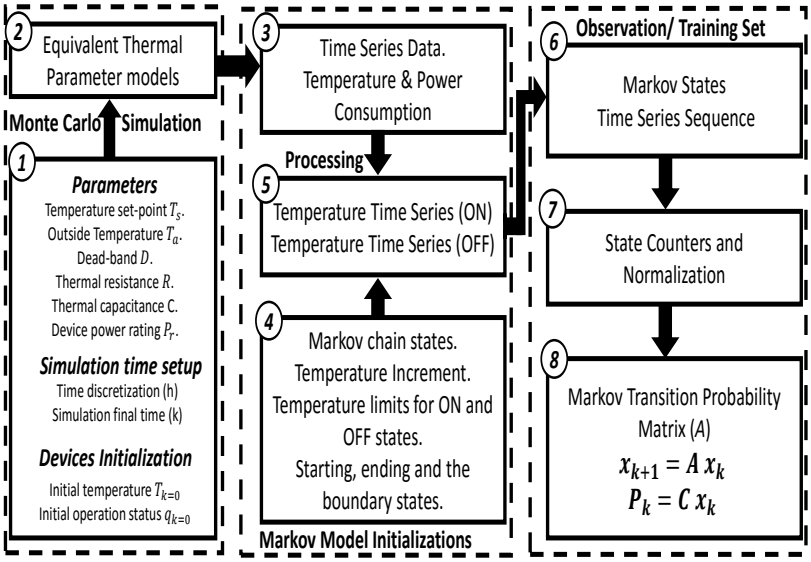

Figure 3.2 Markov chain training Process.

the transition probability between any two particular states can be found as shown in (3.6). The counter $c^d_{(i,j)}$ measures the total number of visits that have been made by device d from state j to state i in one time step. The counter is then normalized over the total number of visits made to state i to have a probability measure (step 7). Finally, the average probability across all devices N_d is found as described in (3.7).

As in linear-systems formulation, the transition probabilities are structured such that the A matrix defines a column stochastic matrix (step 8); *i.e.*, the summation of all probabilities of any particular state must add to one. For suitable discretization time selection (h), the state sequence is less likely to bypass adjacent states and the A matrix will have a general transfer probability structure as shown in Eq. 3.8.

$$P^d_{(i,j)} = P^d(x_i \mid x_j) = \frac{c^d_{(i,j)}}{\sum_{n=1}^{N} c^d_{(i,n)}} \quad d \in 1...N_d \tag{3.6}$$

$$P_{(i,j)} = \frac{\sum_{d=1}^{N_d} P^d_{(i,j)}}{N_d} \tag{3.7}$$

$$A = \begin{bmatrix} P_{(1,1)} & 0 & 0 & \dots & P_{(1,N)} \\ P_{(2,1)} & P_{(2,2)} & 0 & \dots & 0 \\ 0 & P_{(3,2)} & P_{(3,3)} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & P_{(N,N)} \end{bmatrix} \tag{3.8}$$

The training process is performed for a 10,000 air-conditioning devices with two cases. First, all devices are assumed homogeneous, while in the second case heterogeneity is considered. The training data in both systems are generated from an extreme initial condition which assumes that all devices are initially OFF and operating at the same initial temperature of 20 $^{\circ}C$. The simulation final time is set for 24 hours with time discretization of 1 second. Accordingly, the training data sets are statistically significant and each Markov state has a large number of visits to derive reliable conclusions about the actual transition probabilities.

3.2.1 Markov Models for Homogeneous system.

The learning process is initially performed for homogeneous system. The actual power consumption of all devices obtained from simulating the ETP models is shown in black in Fig. 3.3. As an extreme initial condition, all devices are selected to be in the OFF state. This assumption leads to a large oscillation in the aggregated power due to the synchronous operation of the devices. The time-series of the aggregated power experience a non-decaying oscillation due to devices homogeneity. Fig. 3.3 also illustrates the performance of different Markov chains initialized with the same conditions. In the training process, homogeneous loads are subjected to the bias error caused by the temperature discretization. As a result, Markov models with low number of states underestimate the actual system behavior (Markov chains converge to a steady-state value while the actual system is oscillating). Accordingly, increasing the number of states in the learning algorithm allows the Markov models to minimize this error and eventually improves the model predictions (see Fig. 3.3). This results support the finding in literature [34],[38], where it has been shown that increasing the number of states improves the overall model accuracy in case of homogeneous TCLs.

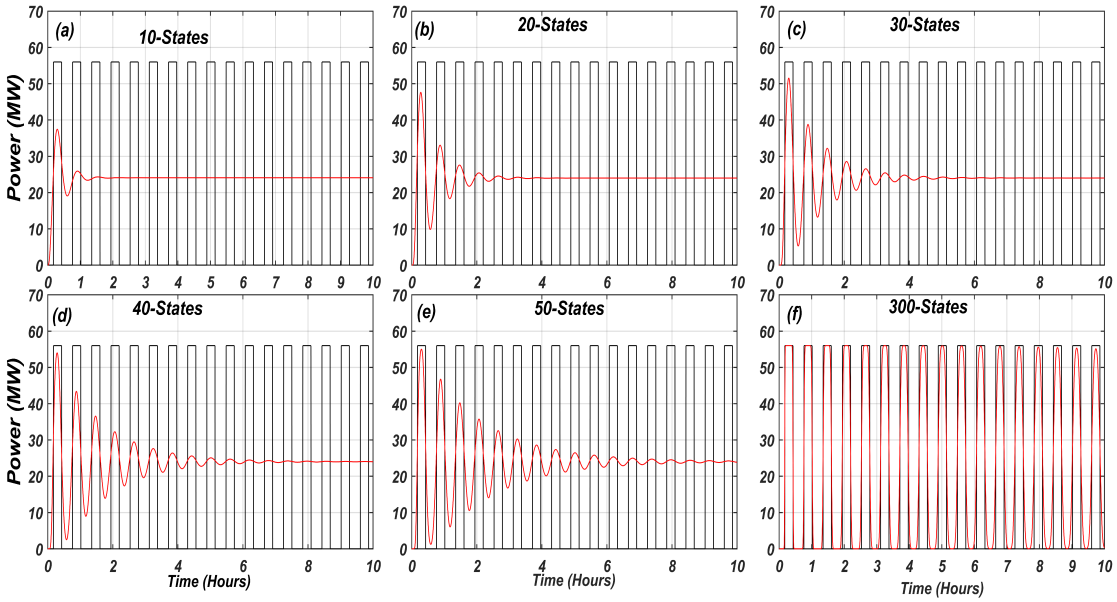

Figure 3.3 Markov chian training for homogeneous TCL system. Black: actual ETP simulation. Red: Markov model.

3.2.2 Markov Models for Heterogeneous system.

In contrast to homogeneous systems, diversity in devices' characteristics of heterogeneous systems prevents the oscillations, since each device operates at a different frequency and phase. As a result, the aggregated power converges to its steady-state condition with a convergence rate based on devices' heterogeneity. Similar illustration is also presented in this section. Fig. 3.4 demonstrates the ETP simulation in black and compare the performance of different Markov chins shown in red. Small number of states underestimate the transient behavior of the actual system but still can predict its steady-state. Increasing the number of states improves the Markov chain accuracy. However, after a specific number of states, the Markov chain starts deviating from the actual system. If we consider the same number of states used to model the homogeneous system *i.e.*, 300 states, the resulting Markov chain is not capturing the actual heterogeneous system. instead, the resulting Markov chain captures the equivalent homogeneous system *i.e.,* all devices have parameters at the mean values. (see Fig. 3.4).

It has been shown that modeling the homogeneous system is subjected to the bias error caused by the temperature discretization. However, for the heterogeneous system there is another source of error caused by the variance in the training data sets. This error is produced due to the different realizations obtained from devices, since each device has distinct parameters. For such systems, in order to minimize the variance error, the number of Markov chains must be increased along with the number of states, such that the combined effect of all of the Markov chains provides similar characteristics of the actual heterogeneous system. The optimal system representation from the accuracy point of view is to increase the number of the Markov chains up to number of devices and increase the number of states in each Markov chain. However, tractability will be lost with this huge number of linear systems. In reality, the number of Markov chains should reflect the trade-off between the computational efficiency, and acceptable accuracy levels.

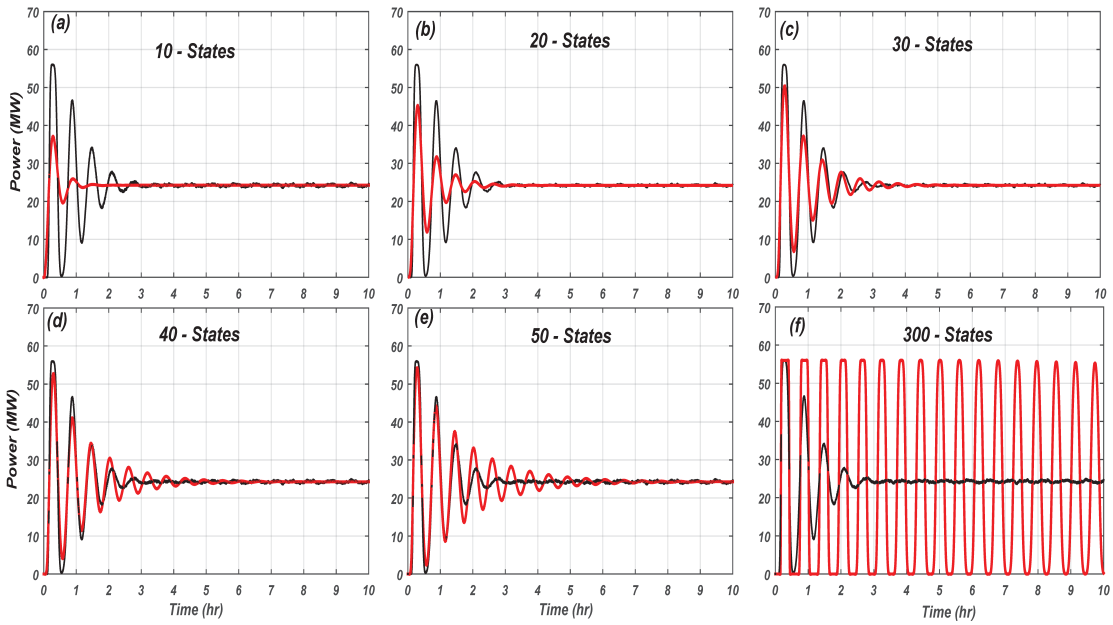

Figure 3.4 Markov chian results for heterogeneous TCL system. Black: actual ETP simulation. Red: Markov model simulation.

Modeling the heterogeneous system with single Markov chain will have this limitation, where both of errors can not be eliminated simultaneously. Single Markov chain is subjected to the bias-variance trade-off [65], and there is a specific number of states that can minimally reduce both of

the errors, but the performance of the system will not provide a perfect fit as demonstrated in the homogeneous case. In order to determine this specific Markov design, a comprehensive sensitivity analysis for the heterogeneous system is performed in the next section.

3.3 Sensitivity analysis of Heterogeneous system

The objective of this section is to evaluate the performance of the Markov chains against different validation data to reveal the best design in handling both the temperature discretization error and the variance error in the heterogeneous system. The training data set generated in the previous section is used to train various Markov chains with different number of states. The modeling process starts with two states and increases with an increment of 4 up to 200 states.

The validation data is obtained by running the ETP simulation with different initialization. In total, new 11 initial conditions are used in this sensitivity analysis. initial conditions are classified as the percentage of devices initially in the OFF state. For simplicity, the initial temperature state is considered uniformly distributed over the dead-band. The simulation of the ETP models given all initial conditions is presented in Fig. 3.5. Less oscillation is produced from initial conditions near the steady-state value (the 50% and 60% OFF devices), compared to the extreme initial conditions (0% and 100%). While it takes approximately three hours from all initial conditions to converge to the steady-state.

All of the Markov models are simulated for the given initialization, and the obtained results are compared to the actual system dynamics shown in Fig. 3.5. The Normalized Root Mean Square Error (NRMSE) presented in equation (3.9) is used to quantify the overall model performance. Where, P_{ss} represents the steady-state power consumption, $P(k)$ is the actual power obtained from ETP models, $\bar{P}(k)$ is the estimated power of any particular Markov model, and k_f is the total number of samples considered in the analysis.

$$NRMSE = \frac{1}{P_{ss}} \sqrt{\frac{\sum_{k=1}^{k_f} (P(k) - \bar{P}(k))^2}{k_f}} \tag{3.9}$$

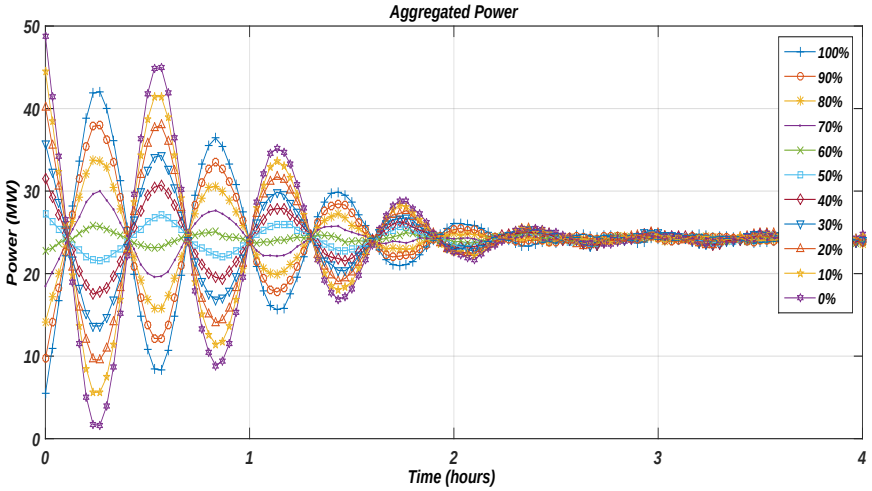

Figure 3.5 ETP simulation of all initial conditions (percentage of devices in OFF state).

The results of the analysis are shown in Fig. 3.6. Initially, the performance of the Markov chains is compared in a relatively short-term scale in Fig. 3.6.a, *i.e.,* first half-hour. It is clearly shown that increasing the number of states improves the modeling accuracy similar to the homogeneous case for this time-frame. This result is due to the fact that the variance in training data sets does not have a major impact in the considered time-scale. The operation of the TCL devices involves a relatively slow dynamics. Thus, the effect of devices' heterogeneity can not be distinguished in the short-term predictions. Therefore, modeling the heterogeneous devices can be seen as the homogeneous case where only minimizing the bias error is the dominant factor. The effect of the variance starts to take place after the first half-hour causing the aggregated power to converge to the steady-state value.

If longer time-scales are considered, for instance, the first three hours, as shown in Fig. 3.6.b, the effect of the variance in the training data sets becomes more dominant while the aggregated power starts having less oscillation. Increasing number of states in this case will not conclude in a promising results as in the homogeneous case, rather the error magnitudes start increasing for all models larger than 40-states. Therefore, for long-term predictions a trade-off between minimizing the bias and variance errors is required to find a good representative model that approximates the

actual system behavior. As shown in Fig. 3.6.b, for this load heterogeneity, the 40-state model gives the minimal error compared to the other Markov designs and for all of the initial conditions. The performance of the 40-state model for one extreme initial condition is previously shown in Fig. 3.4.

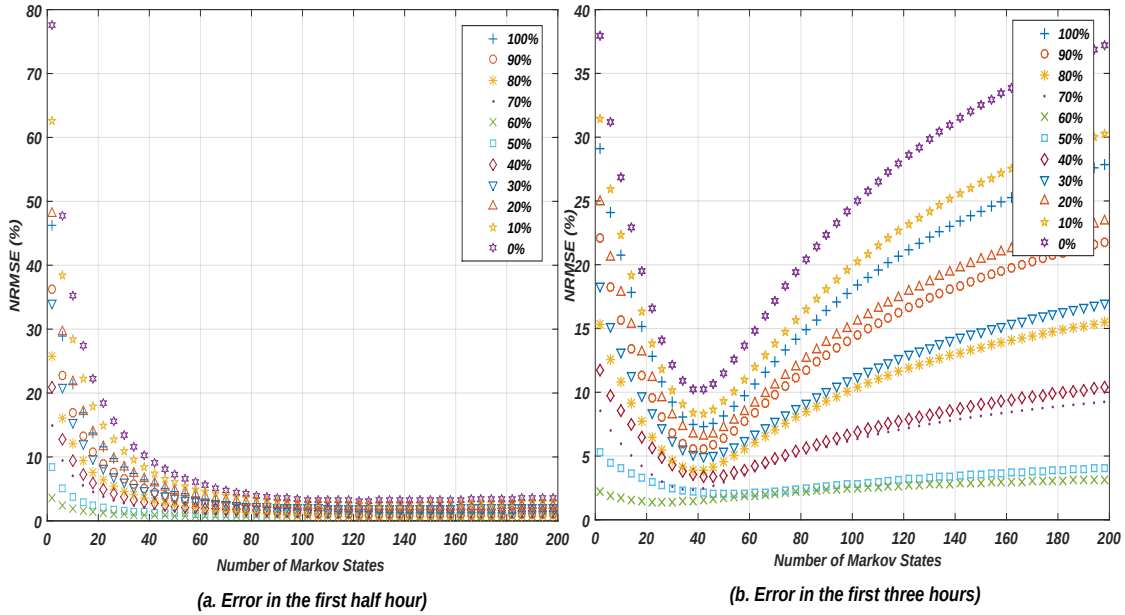

Figure 3.6 Markov model performance for various states and initial conditions.

3.4 Markov model sensitivity to simulation time-step

The development process of the Markov chains with 40-states requires that the training data have a good time resolution, whether the training data sets are obtained from simulation or collected from actual devices. Time resolutions more than 18 seconds does not yield representative models. This can be shown in Fig. 3.7, which presents the effect of varying the discretization time-step on the training process. Clearly, increasing the time-step causes the system to deviate from the actual one. Moreover, models with more than 72 seconds do not even approximate the actual system behavior.

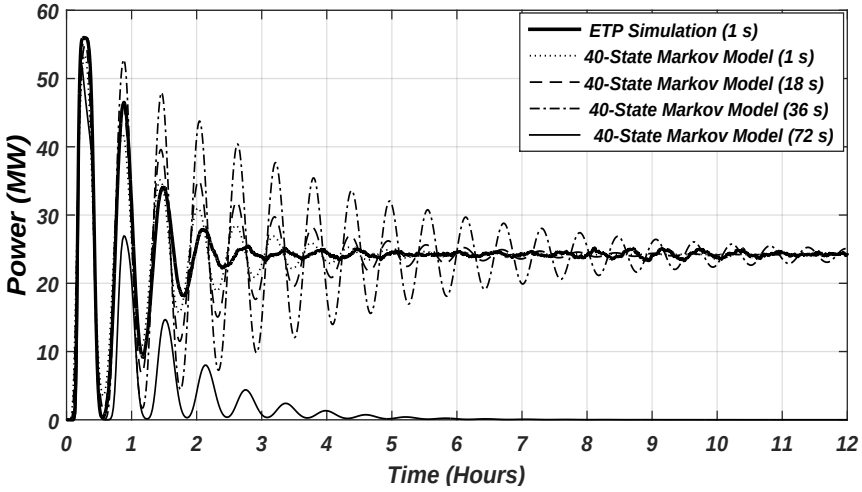

Figure 3.7 40-state Markov model training for various discretization time steps.

3.5 Eigenvalue Analysis of Markov Chain Models

The steady-state probability density functions for the OFF and ON devices have been derived analytically for homogeneous TCL system in [25],[29],[32]. In this section, the eigenvalue analysis of the Markov chains will be used to derive a representative distribution during the steady-state for the heterogeneous loads. Markov chains are described by the transition probability matrix A which constitutes a general stochastic matrix. Accordingly, the model eigenvalues are placed inside the unit circle. However, there will be a dominant eigenvalue exists on its circumference and occurs exactly at one. while all of the other eigenvalues have smaller real parts. Thus, this eigenvalue is dominant and represents the steady-state mode of the system. As the number of the states increases, the eigenvalues skewed farther to the left, causing larger complex conjugate values to appear in the system (see the outer ellipse in Fig. 3.8 which represents the 200-states). As a result, the system becomes more sensitive to perturbations and takes a longer time to converge. This is also interpret into oscillations in the homogeneous system, where increasing the number of states is essential to improve the modeling accuracy.

A detailed probability mass function during the steady-state condition can be found by determining the right-eigenvector corresponding to the dominant eigenvalue at one (v^*). Such that

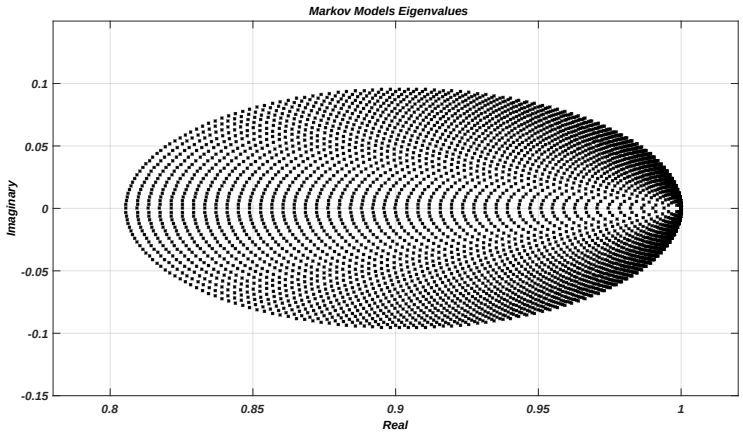

Figure 3.8 Eigenvalues of all Markov models.

devices' steady-state distribution (x^*), can be obtained as shown in (3.10), where N_d; represents the total number of TCL devices. For the 10,000 heterogeneous TCLs considered in section. 2.4 and modeled using the 40-state Markov chain, the steady-state distribution is found using (3.10) and illustrated in Table.3.1.

$$x_i^* = \frac{\mid v_i^* \mid}{\|v^*\|_1} N_d \tag{3.10}$$

Table 3.1 Steady-state distribution based on the eigenvalue analysis.

OFF States				ON States			
x_i^*	#	x_i^*	#	x_i^*	#	x_i^*	#
1	323.14	11	281.77	21	249.77	31	214.71
2	270.94	12	282.19	22	208.56	32	215.70
3	272.05	13	283.48	23	209.41	33	216.09
4	273.04	14	284.77	24	210.15	34	216.93
5	274.38	15	285.79	25	210.84	35	217.64
6	275.11	16	287.16	26	211.39	36	218.21
7	276.55	17	288.33	27	212.07	37	219.08
8	277.61	18	289.68	28	212.93	38	219.78
9	278.71	19	290.97	29	213.43	39	220.48
10	279.99	20	291.84	30	214.12	40	221.01
OFF percentage		56.68%		ON percentage		43.32%	

The distribution is numerically verified in Fig. 3.9. The ETP models are initializing with values as specified in Table. 3.1. The performance is compared against the 100% OFF initial condition to verify the results when all devices naturally converge to the steady state-value. The values listed in the table are approximated to the nearest integer and considered uniformly distributed over the corresponding temperature intervals defined in (3.4)-(3.5).

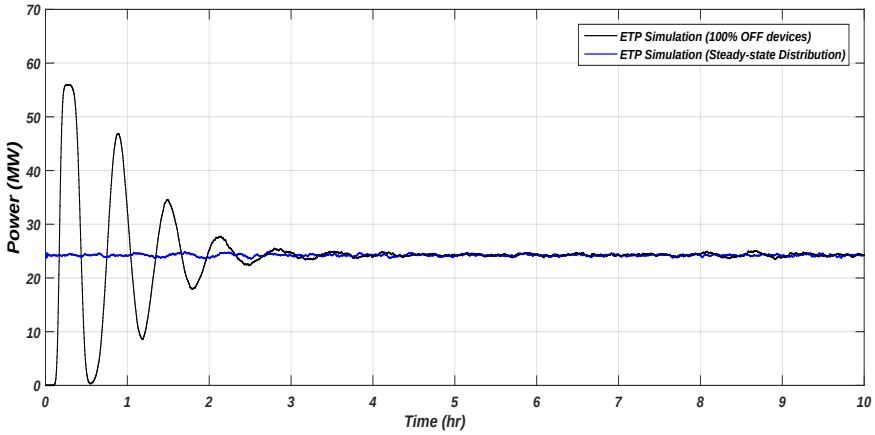

Figure 3.9 ETP simulation based on the steady-state distribution.

3.6 Markov Model development at various temperature set-points

Markov chains are developed in the previous sections at a fixed operating set-point. Set-point control will be used in our work to extract or store additional thermal energy in the TCL devices based on specific comfort levels. Therefore, it is required to evaluate the steady-state power consumption at various set-points to estimate the power capability of the devices. In this section, the ETP simulation is evaluated at different set-point settings and the corresponding Markov chains are developed accordingly. The steady-state power consumption is evaluated using both the ETP simulation and the Markov chains eigenvalue analysis discussed in the previous section.

It is assumed that the customers allow load-aggregator to modify their temperature set-point over a relatively wide range, such that the minimum set-point is $16°C$ and the maximum set-point adjustment is $24°C$. We consider discretizing the control space with $0.1°C$. Thus, the new set-point

may take 81 combination. For each set-point a Markov chain has been developed offline and saved in the database to be used in the next chapter for the control applications. Fig. 3.10 compares the Markov chains simulation with the ETP simulation. It is assumed that all devices are initially ON and uniformly distributed over the the dead-band of each temperature set-point. For better illustration, the results are shown in Fig. 3.10 for only $0.5°C$ step, while the steady-sate power consumption for each set-point is provided in Table. 3.2.

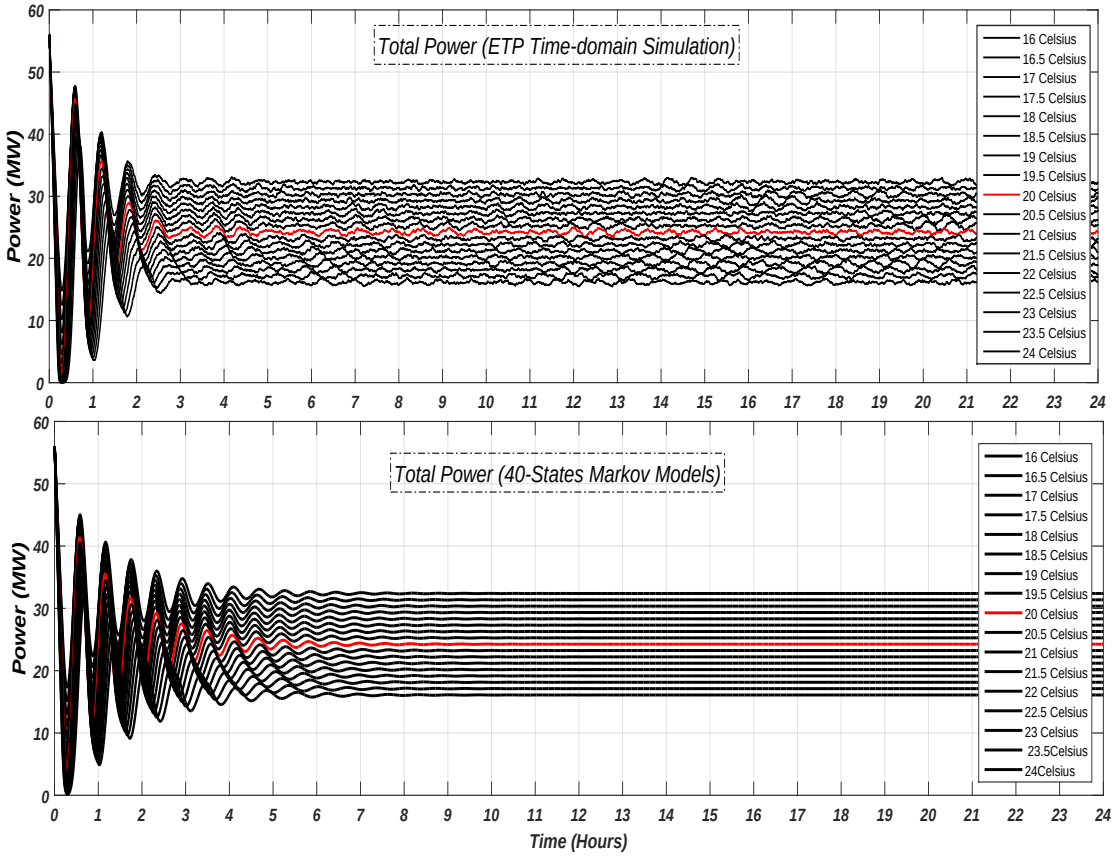

Figure 3.10 Markov Models and ETP simulation at various set-points.

3.7 Conclusion

The chapter describes the TCLs aggregation as a Markov chain. Statistical learning technique is used to derive the parameter of the aggregated model and for both homogeneous and heterogeneous

systems. It has been shown that increasing the number of states improves the Markov chains performance for the homogeneous TCLs. However, modeling the heterogeneous loads is subjected to the bias-variance trade-off. Therefore, there is a specific number of states that can reduce both of the error sources minimally. Comprehensive sensitivity analysis has shown that the 40-state Markov chain is the best model design to approximate the heterogeneous system behavior in both the transient and the steady-state. For this level of heterogeneity, the 40-state Markov chain is chosen as a base case model and is used to aggregate the TCLs at various temperature set-points.

Table 3.2 Steady-state Power consumption for various set-point settings.

Temperature	Power (MW)
16 $^{\circ}C$	32.394
16.5 $^{\circ}C$	31.375
17 $^{\circ}C$	30.356
17.5 $^{\circ}C$	29.339
18 $^{\circ}C$	28.322
18.5 $^{\circ}C$	27.305
19 $^{\circ}C$	26.287
19.5 $^{\circ}C$	25.271
20 $^{\circ}C$	24.253
20.5 $^{\circ}C$	23.235
21 $^{\circ}C$	22.217
21.5 $^{\circ}C$	21.198
22 $^{\circ}C$	20.180
22.5 $^{\circ}C$	19.160
23 $^{\circ}C$	18.139
23.5 $^{\circ}C$	17.119
24 $^{\circ}C$	16.098

CHAPTER 4. MARKOV MODEL EXTENSION FOR TEMPERATURE SET-POINT CONTROL

4.1 Introduction and Overview

This chapter provides a novel modeling framework for a heterogeneous group of TCLs. The uniqueness of this new model is the capability to capture the TCLs transient and steady-state dynamics under various set-point adjustments and in both directions. This feature is beyond the capability of the aggregated models developed in literature which is designed and thus valid at fixed set-points. The new modeling approach is based on restructuring and linking Markov chains that is previously developed at fixed set-point. thus, the model is called the Extended Markov Model (EMM).

The objective of this new model is to provide the system services through manipulating the stored thermal energy in the TCL devices, rather than relying on the devices' switching capability as discussed in literature. Regardless of the shape and the power oscillations associated with the set-point adjustments. The objective of the EMM is to accurately describe devices' transition and the aggregated power behavior in linear system framework. Solution to the power oscillation problem is proposed in the next chapter.

The details of the model development process is described for both set-point increase and decrease cases in Section 4.2. The model formulation and the corresponding results are validated against the actual ETP simulation in Section 4.3. The results have shown that the EMM can capture the TCLs' transient and the steady-state behavior for various set-point adjustments and in both directions.

4.2 The Extended Markov Model

This section discusses the extension of the Markov chain modeling approach to describe the dynamical behavior involved in temperature set-point adjustments. Modifying the operating set-point involves a slow process need to be described precisely. If this dynamics is ignored, and the Markov chains are directly switched, the transition will not be valid to describe the aggregated power waveform. For instance, Fig. 4.1 illustrates modifying the operating set-point of 10,000 air-conditioning load. Devices are initially operating at $20°C$ and started from the stead-state distribution as previously shown in Table. 3.1. At $t = 1hr$, the set-point is changed to $21°C$. The bold line demonstrates changing the set point using the ETP simulation while the dashed line illustrates switching the two Markov chains. *i.e.,* using the $20°C$ chain until $t = 1hr$ then switched to the $21°C$ chain. It is clearly shown that directly switching the two Markov chains does not describe the actual simulation from the ETP models. Therefore, broadcasting new set-points require defining a new system able to describe the actual dynamics which involved in set-point adjustments. This new system will be referred as the Extended Markov Model (EMM).

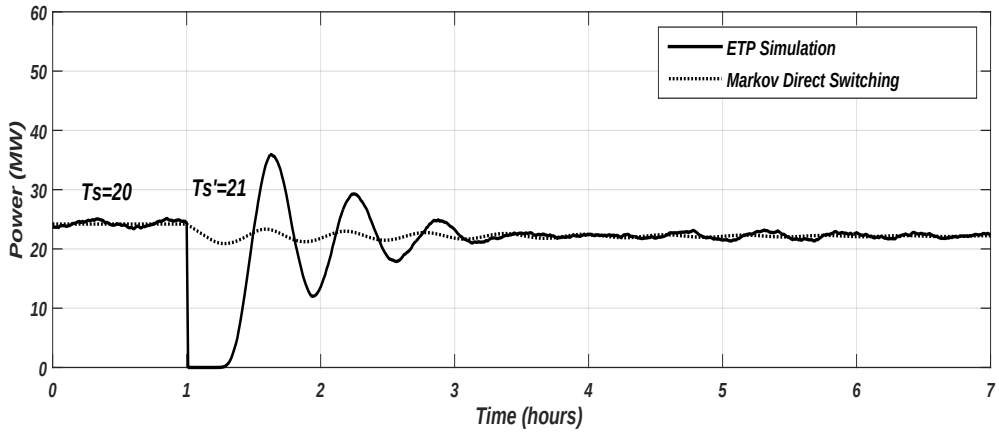

Figure 4.1 Comparison between ETP simulation and Markov direct switching.

The EMM is subjected to move devices from the old set-point setting until they are regulated at the new one. Thus, the dimension and the structure of the new model will be determined based on the sign and the magnitude of the set-point change. For instance, if we consider modifying the

set-point with a magnitude equal to the dead-band length, the Markov chain of the new set-point is formed adjacent to the old one, while its relative location is based on the direction of the set-point change, this configuration is illustrated in Fig. 4.2. For adjustment less than the dead-band, an overlap will occur between the two Markov chains. While for adjustment larger than the dead-band a gap will be formed between them. Therefore, the EMM need to be defined for all set-point magnitudes and in both directions.

It is assumed that the set-point change is defined as an integer multiple (m) of the state length (Δx), such that the new set-point can be defined as; $T_s^{'} = T_s \pm m\Delta x$. This assumption is required to prevent partial state involvements in the control actions. The EMM shown in Fig. 4.2 illustrates the case when the changes equal the dead-band, hence $m = \pm N/2$.

In general, define the EMM as shown in (4.1-4.2). The system comprises transient-states (z^{t1}, z^{t2}) and final-states (z^{off}, z^{on}). The transient-states of dimension $2m$ are modeled to describe how the devices will move form the old set-point until they reach the new one. Therefore, these states will be populated by devices only during the transition period, and will be drained during the steady-state conditions when all devices reaches the new set-point *i.e.,* the final-states.

$$z(k+1) = A_z\, z(k), \quad A_z \in R^{(2m+N)\times(2m+N)}, z \in R^{(2m+N)\times 1} \tag{4.1}$$

$$z = \left[\underset{(m\times 1)}{\mathrm{z}^{\mathrm{t1}}}, \underset{(m\times 1)}{\mathrm{z}^{\mathrm{t2}}}, \underset{(N/2\times 1)}{\mathrm{z}^{\mathrm{off}}}, \underset{(N/2\times 1)}{\mathrm{z}^{\mathrm{on}}} \right]^T. \tag{4.2}$$

Building the system-matrix A_z and initializing the states are based on the magnitude of the set-point change. There are two main cases. First, if the set-point change is less than the dead-band, both of the Markov chains are overlapped. Accordingly, transient and final-states share devices from the old Markov model. However, in case of large changes, only the transient-states are populated by devices and the final-states are initialized by zeros. The following sections provide details for set-point increase and decrease formulation for a group of air-conditioning loads.

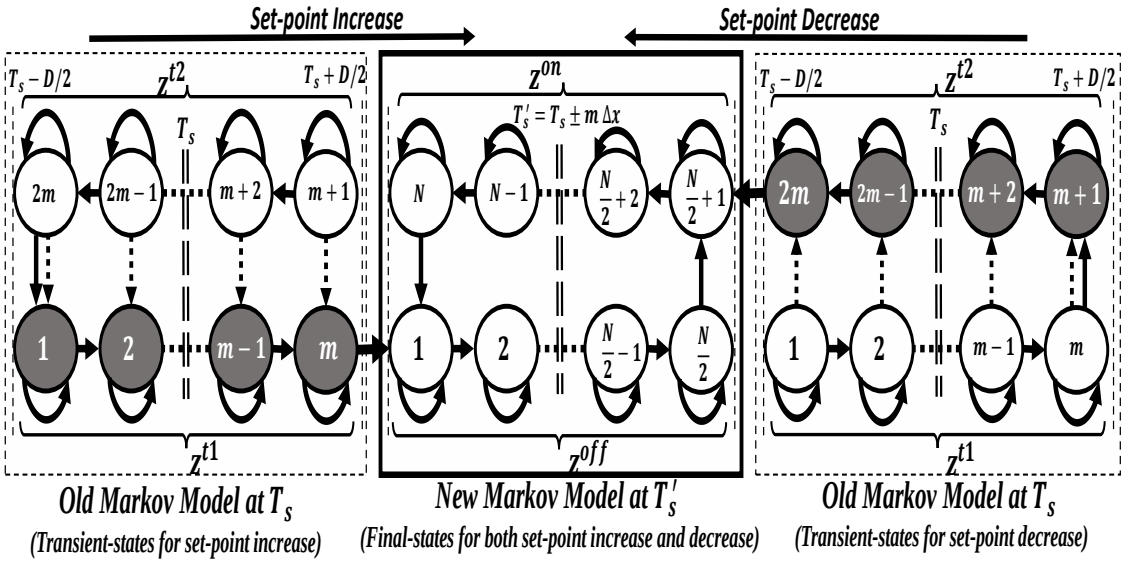

Figure 4.2 Extended Markov model structure for set-point increase and decrease cases (dark states are heavily populated by devices while the white are empty).

4.2.1 Set-point Increase Formulation

Increasing the temperature set-point for a group of air-conditioning devices forces the operating devices to turn OFF until their internal temperature increases to the new set-point setting. Therefore, all of the ON devices that are covered in the control action will change their status to OFF. The switching will occur between the corresponding states. *i.e.,* states with the same temperature representation as described earlier in (3.4-3.5). This transition is illustrated by the dotted arrows in Fig. 4.2. Accordingly, the ON transient-states z^{t2} are initialized by zeros, while the OFF transient-states z^{t1} are initialized by augmenting the ON and OFF devices. The variable x refers to the states of the old Markov chain. If the set-point change exceeds the dead-band length ($m > N/2$), then additional states need to be defined and initialized by zeros as shown in (4.3-4.4).

$$z_i^{t1} = \left\{ \begin{matrix} x_i + x_{(N-i+1)} & i \leq N/2 \\ 0 & i > N/2 \end{matrix} \right\}, \quad i \in [1, 2, ..m] \tag{4.3}$$

$$z_i^{t2} = \{0\}, \qquad i \in [1, 2, ..m] \tag{4.4}$$

$$z_i^{off} = \begin{Bmatrix} x_{(i+m)} & i \leq N/2 - m \\ 0 & i > N/2 - m \end{Bmatrix}, \quad i \in [1, 2, ..N/2] \tag{4.5}$$

$$z_i^{on} = \begin{Bmatrix} 0 & i \leq m \\ x_{(N/2+i-m)} & i > m \end{Bmatrix}, \quad i \in [1, 2, ..N/2] \tag{4.6}$$

Initializing the final-states z^{off} and z^{on} is also based on the set-point change magnitude. Large changes will initialize these states by zeros, while in case of small adjustments. *i.e.,* $(m \leq N/2)$, the new Markov chain will map values from the old one as shown in (4.5-4.6).

The system-matrix A_z has a structure as shown in (4.7); A_t represents the transient-states dynamics, and A' is the Markov transition probability matrix at T_s' . For set-point adjustments smaller than the dead-band, the matrix A_t can be obtained from the old Markov matrix A. However, for changes larger than the dead-band, a gap will be formed between the old and the new Markov chains. Therefore, it is required to define a general matrix U that traverse the entire control space. *i.e.,* the maximum and minimum set-points allowed for control ($\overline{T_s}$ and $\underline{T_s}$ respectively). The matrix U can be found by reassembling adjacent Markov matrices with a structure as previously shown in (3.8).

For any set-point control magnitude, A_t can be structured as shown in (4.8). Where, A_{t1} and A_{t2} represent the dynamics of z^{t1} and z^{t2} respectively, and can be defined as a range of elements taken from U as shown in (4.9) and (4.10). The matrix A_{t3} defines how the states in z^{t2} are linked to those in z^{t1}, these probability links are defined at only one location, hence A_{t3} constitutes a sparse matrix except only one element as shown in (4.11). Since there is no forward links between z^{t1} and z^{t2}, all of the lower diagonal elements in A_t are zeros. The linking matrix L defines the probability to move devices from the last state in z_{t1} (m) and enters the first state in the new Markov chain (see Fig. 4.2). Therefore, the matrix L is also a sparse matrix with one non-zero entry at $(1, m)$ which defined in (4.12). Once devices are regulated at the new set-point, they will stay there and will not return back to the transient-states. Hence, all of the the upper diagonal

elements in A_z are zeros. Finally, the system output-matrix C_z (4.13) is modified to account for the power consumption of all devices in the transient-states.

In this formulation, the ON transient-states z^{t2} are initialized by zeros, and the unforced system will not return devices back to the old set-point. However, these states are modeled for the purpose of the sequential control algorithm that may force devices to stay at the old set-point setting before they can make the transition to the new one.

$$A_z = \begin{bmatrix} \underset{2m\times 2m}{\mathrm{A_t}} & \underset{2m\times N}{\mathbf{0}} \\ \underset{N\times 2m}{\mathbf{L}} & \underset{N\times N}{\mathrm{A'}} \end{bmatrix}. \tag{4.7}$$

$$A_t = \begin{bmatrix} \underset{m\times m}{\mathrm{A_{t1}}} & \underset{m\times m}{\mathbf{A_{t3}}} \\ \underset{m\times m}{\mathbf{0}} & \underset{m\times m}{\mathrm{A_{t2}}} \end{bmatrix}. \tag{4.8}$$

$$A_{t1} = U\Big[(T_s - \underline{T_s})\frac{N}{2} + 1 : (T_s' - \underline{T_s})\frac{N}{2}\Big]. \tag{4.9}$$

$$A_{t2} = U\Big[(2\overline{T_s} - \underline{T_s} - T_s')\frac{N}{2} + 1 : (2\overline{T_s} - \underline{T_s} - T_s)\frac{N}{2}\Big]. \tag{4.10}$$

$$A_{t3}(1,m) = \big(1 - A_{t2}(m,m)\big). \tag{4.11}$$

$$L(1,m) = \big(1 - A_{t1}(m,m)\big). \tag{4.12}$$

$$C_z = (\bar{P}_r/\eta)[\mathbf{0}_{(1\times m)}\ \mathbf{1}_{(1\times m)}\ \mathbf{0}_{(1\times N/2)}\mathbf{1}_{(1\times N/2)}\,]. \tag{4.13}$$

4.2.2 Set-point Decrease Formulation

Decreasing the temperature set-point on the other hand forces the OFF state devices to operate and start the cooling cycle until they reach the new set-point setting. This action will leave all of the OFF transient states z^{t1} with no devices (4.14). However, the ON transient-states z^{t2} (4.15)

will be initialized by augmenting the ON and OFF devices of the old Markov chain given that the set-point is less than the dead-band. Otherwise, additional states are required to be defined and initialized by zeros. Initializing the final-states varies based on the size of the set-point adjustment. Large set-point change will initialize them by zeros. However, if the set-point change is less than the dead-band, the final-states will map devices from the old Markov chain as shown in (4.16-4.17).

$$z_i^{t1} = \{0\}, \quad i \in [1, 2, ..m] \tag{4.14}$$

$$z_i^{t2} = \begin{Bmatrix} x_{(N/2+1-i)} + x_{(N/2+i)} & i \leq N/2 \\ 0 & i > N/2 \end{Bmatrix}, \quad i \in [1, 2, ..m] \tag{4.15}$$

$$z_i^{off} = \begin{Bmatrix} 0 & i \leq m \\ x_{i-m} & i > m \end{Bmatrix}, \quad i \in [1, 2, ..N/2] \tag{4.16}$$

$$z_i^{on} = \begin{Bmatrix} x_{(N/2+i+m)} & i \leq N/2 - m \\ 0 & i > N/2 - m \end{Bmatrix}, \quad i \in [1, 2, ..N/2] \tag{4.17}$$

The system-matrix A_z has similar structure as presented earlier in (4.7). However, the transient-state matrix A_t and the linking matrix L have different representations as shown in (4.18) and (4.19) respectively. All of the upper diagonal elements in A_t are zeros since devices will be regulated at the new set-point and not allowed to return back to the old set-point. The linking matrix L defines the probability of moving from the last state in z^{t2} $(2m)$ and enters the state $(N/2+1)$ of the new Markov chain (see Fig. 4.2). A_{t1} and A_{t2} can take a range of elements from U as defined in (4.20) and (4.21) respectively. The matrix A_{t3} defines how the states in z^{t1} are linked with the states in z^{t2}. Thus, A_{t3} is defined at only one location as shown in (4.22).

$$A_t = \begin{bmatrix} \underset{m\times m}{\mathrm{A_{t1}}} & \underset{m\times m}{\mathbf{0}} \\ \underset{m\times m}{\mathbf{A_{t3}}} & \underset{m\times m}{\mathrm{A_{t2}}} \end{bmatrix}. \tag{4.18}$$

$$L((N/2+1), 2m) = \big(1 - A_{t2}(m, m)\big). \tag{4.19}$$

$$A_{t1} = U\Big[(T_s' - \underline{T_s})\frac{N}{2} + 1 : (T_s - \underline{T_s})\frac{N}{2}\Big]. \tag{4.20}$$

$$A_{t2} = U\Big[(2\overline{T_s} - \underline{T_s} - T_s)\frac{N}{2} + 1 : (2\overline{T_s} - \underline{T_s} - T_s')\frac{N}{2}\Big]. \tag{4.21}$$

$$A_{t3}(1, m) = \big(1 - A_{t1}(m, m)\big). \tag{4.22}$$

4.3 Model Verification

In order to verify the EMM formulation, it is assumed that a 10,000 air-conditioning device are initially regulated at $20°C$ and thus consuming a total aggregated power of $24.25\,MW$. Several 40-states Markov chains are developed offline at various temperature set-points and used to construct the EMM model. Fig. 4.3 compares the performance of the EMM with the ETP simulation for several set-point increase and decrease cases. For each set-point change the EMM is constructed based on the old and the new temperature set-points. The adjustments covers the cases when the shift is less, equal, or larger than the dead-band length. Clearly, the EMM can predicts to a large extent the dynamics involved in various set-point changes.

4.4 Conclusion

A new modeling approach has been derived to incorporate the set-point control for the heterogeneous TCL devices. The model derivation is based on an extension to the Markov chain abstraction method developed in literature at fixed temperature set-point. Model formulation describes the TCLs' dynamics involved in small or large set-point adjustments. The model is proposed to achieve the system ancillary services through extracting or storing thermal-energy in the TCLs by modifying the operating temperature set-point. It is shown in the validation section that the new model can capture the TCL transitions between various set-points with acceptable level of accuracy. Small error is naturally propagated in the EMM dynamics due to the bias-variance

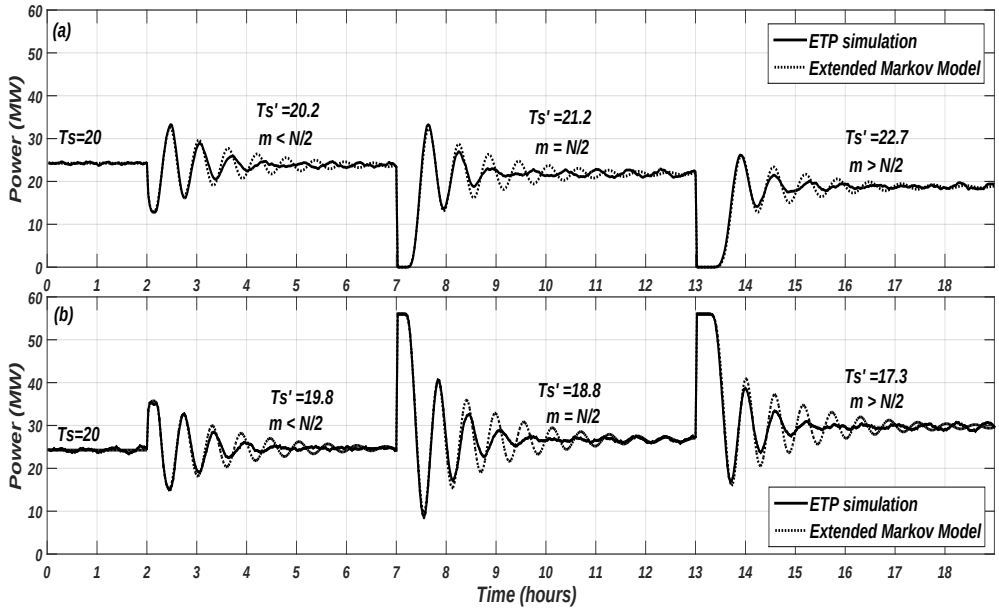

Figure 4.3 Extended Markov Model Validation (a) set-point increase. (b) set-point decrease.

trade-off error discussed previously in the chapter 3. Future work will address further improving the modeling accuracy.

It is shown in the validation step that temperature set-point adjustment forces the devices to operate in a synchronized pattern causing the aggregated power to oscillate and overshoot to extreme values before it finally converges to a new steady-state value. The power oscillation last for long time-intervals and decays as devices' heterogeneity recovers the steady-state conditions. Therefore, the set-point control requires additional control technique which is able to prevent power oscillations and guarantees a fast convergence to the new seat-state value. These issues will be discussed in the next chapter.

CHAPTER 5. CONTROL DEVELOPMENT FOR THE EXTENDED MARKOV MODEL

5.1 Introduction and Overview

It is shown in the previous chapter that changing the operating set-point achieve steady-state load-reduction or increase services. However, the aggregated power shows unstable transients before it finally converges to the new steady-state value. For set-point changes larger than the dead-band, devices are forced to operate in only one state for a considerable amount of time. *i.e.,* all devices are either OFF with zero power consumption, or ON with maximum power consumption. Therefore, this oscillatory transients need to be prevented through additional control technique which is able to avoid such extreme oscillation and guarantee a fast convergence to the new steady-state value.

This chapter thus proposes the Model Predictive Control (MPC) framework to resolve the aforementioned issues. The MPC is subjected to find the optimal sequential set-point adjustments such that devices synchronous operation is prevented. The MPC determines the optimal devices' flow toward the new set-point by the mean of ON/OFF switching capability. Some devices are forced to stay at the current set-point until the other reach the new set-point setting. The ON/OFF switching control will also be used to suppress the oscillations and force a fast convergence to the new steady-state. When all devices reach the new set-point, the system is considered naturally stable and no further control actions are needed.

The performance of the proposed control framework is compared with models proposed in literature which rely only on switching the devices ON or OFF without adjusting the stored thermal-energy level.

5.2 Model Predictive Control with the EMM model

The direct load control paradigm adopted in this work allows load aggregators to selectivity dispatch the set-point control, and modify the status of individual devices by switching them between the ON and the OFF states. The EMM will be used to model devices transition between the old and the new temperature set-points. However, if the set-point is applied to all devices at once, devices will be synchronized. Therefore, the MPC controller is proposed to determine which devices should accept the set-point change immediately and which devices should wait with specific delay-time.

Accordingly, the scope of the MPC is to achieve two main objectives. First, is to determine the optimal number of devices and their distribution that are required to stay at the old-set point before they can make a transition to the new set-point. This will be determined at the beginning of each planning horizon *i.e.,* at the instant of applying the set-point change. The second control objective is to prevent the power oscillations while devices are making the transition to the new set-point. Some devices are required to change their status until they are finally converge to the new steady-state value. Once convergence is achieved, the system will be stable and the new aggregated power will be as desired.

The direct ON/OFF switching control is augmented to the EMM model and represented by a control input $u \in R^{(m+N/2)}$. The EMM formulation with the new control feature is described in (5.1-5.3). Switching the devices is done between the corresponding states which have the same temperature representation. For instance, in 40-state Markov chain, states 1 and 40 have one control input u_1 that describes the number of devices that is required to switch between these two states. The relation between the corresponding states are presented in the structure of the B matrix (5.2). Therefore, this relation has to be defined for all of the EMM states as described in (5.3), where the diagonal elements represent changing the status of devices during the transient

and the final-states. The system-matrix A_z is constructed as described earlier in (4.7) while the output matrix C_z is defined in (4.13).

$$z(k+1) = A_z\, z(k) + B_z\, u(k) \tag{5.1}$$

$$B = \begin{bmatrix} -1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & -1 \\ 0 & \dots & 1 \\ \vdots & \iddots & \vdots \\ 1 & \dots & 0 \\ & & \end{bmatrix}, \tag{5.2}$$

$$B_z = \begin{bmatrix} \underset{2m\times m}{\mathrm{B}} & \underset{2m\times N/2}{\mathbf{0}} \\ \underset{N\times m}{\mathbf{0}} & \underset{N\times N/2}{\mathrm{B}} \end{bmatrix}, \tag{5.3}$$

The performance index of the optimal control problem (5.4) is set to track the aggregated power associated with the new set-point P^{ref} with a positive weighting factors Q, while minimizing the ON/OFF switching actions based on the weights specified by the positive-definite matrix R. Problem formulation includes the non-negativity constraint (5.7) to all of the system states. This constraint implies that the control actions exist only for states that are populated with devices. Otherwise, the control actions will be forced to zero.

$$MinJ = \sum_{k=k_0}^{k_f} Q(P(k) - P^{ref})^2 + u^T(k) R u(k) \tag{5.4}$$

Such that:

$$z(k+1) = A_z\, z(k) + B_z\, u(k) \tag{5.5}$$

$$P(k) = C_z\, z(k) \tag{5.6}$$

$$z_k \geq 0 \tag{5.7}$$

5.3 Model Performance and Comparison

This section demonstrates the advantages of using the EMM when steady-state long-term services are requested. First, these services are extracted from a modeling framework proposed in literature which rely on the direct ON/OFF switching control but without modifying the operating set-point (section. 5.3.1). Model performance and the associated negative consequences are analyzed and highlighted. Second, the same reference signal is used to extract the services using the proposed modeling framework. Where, the set-point is adjusted to a new thermal energy level and the EMM is used to represent devices transition. The results and performance of the proposed model are compared with the first modeling technique. Both of the models are tested with the same number of devices of 10,000 air-conditioning loads. Devices are initially regulated at $20°C$ and following the same initial-conditions and parameters heterogeneity as previously discussed in Section. 2.4.

5.3.1 Direct ON/OFF Switching Control at Fixed Temperature Set-point

The ON/OFF switching control does not rely on adjusting the temperature set-point. Therefore, a single Markov chain at $20°C$ is used to model the devices as previously shown in Fig. 3.1. Devices are initially following the $20°C$ steady-state distribution. Thus, the aggregated power experience no oscillations. The considered reference signal increases the power consumption by $4\,MW$ at $t = 1h$, and provides a $4\,MW$ load-reduction at $t = 13h$. Both of the load adjustments are set for 6-hour time-intervals as illustrated in Fig. (5.1.a). This control technique extracts the desired services by regulating the number of devices in the ON states by switching the devices between the corresponding states. The MPC is employed to find the optimal switching sequence that is required to achieve these services.

The overall performance is illustrated in Fig. 5.1. The simulation results show that the aggregated power can be modified to track the reference signal by applying a total number of switching actions as in Fig. (5.1.b). Initially, increasing the aggregated power by $4\,MW$ requires switching 725 device from OFF to ON. This is not considered as a huge control effort since it is required at

only one time-instant. However, in order to maintain the service, the total number of switching actions start increases to reach a total of 265 device at a continuous switching rate. At the end of the load-increase service, devices are instructed to return back to the steady-state power consumption value. This action required to turn OFF the same number of devices initially turned ON. However, devices are forced to switch from OFF to ON with a high switching rate that eventually reaches zero even though no services are requested. Same argument applies for the load-decrease case.

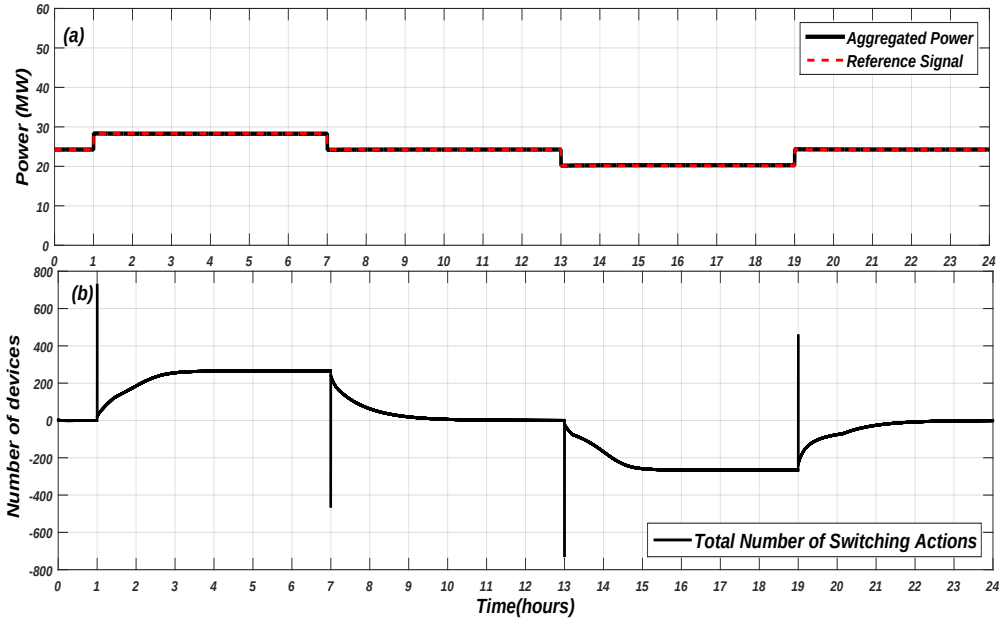

Figure 5.1 ON/OFF switching control without set-point adjustment. (a) Aggregated power consumption and the reference signal. (b)Total number of devices' switching actions.

The control inputs and the number of devices in each state over the control horizon are shown in Fig. 5.2. Clearly, most of the control efforts are exerted at the boundary states. *i.e.,* states 1 & 40 in the case of load-increase, and at states 20 & 21 in the case of load-decrease. During the time of extracting the services, the boundary states are highly congested in devices while a large number of states are completely empty. This is due to the high switching rates imposed by the control actions that is much higher than the unforced system dynamics. If we consider dividing the Markov chain into two equal halves, the lower-half covers all states with a temperature representation less than

the set-point (states 1-10 and 31-40). While the upper-half represents the states 11-20 and 21-30. Fig. 5.3 gives an illustration for the operating conditions during the load-decrease case. The control actions have moved all devices to be congested in only some of the upper-half states, while all states in the lower-half are completely drained. The same argument applies for the load-decrease case but devices will be congested in the lower-half. Therefore, this control approach imposes high switching frequencies and forces the devices to operate with much shorter duty-cycles.

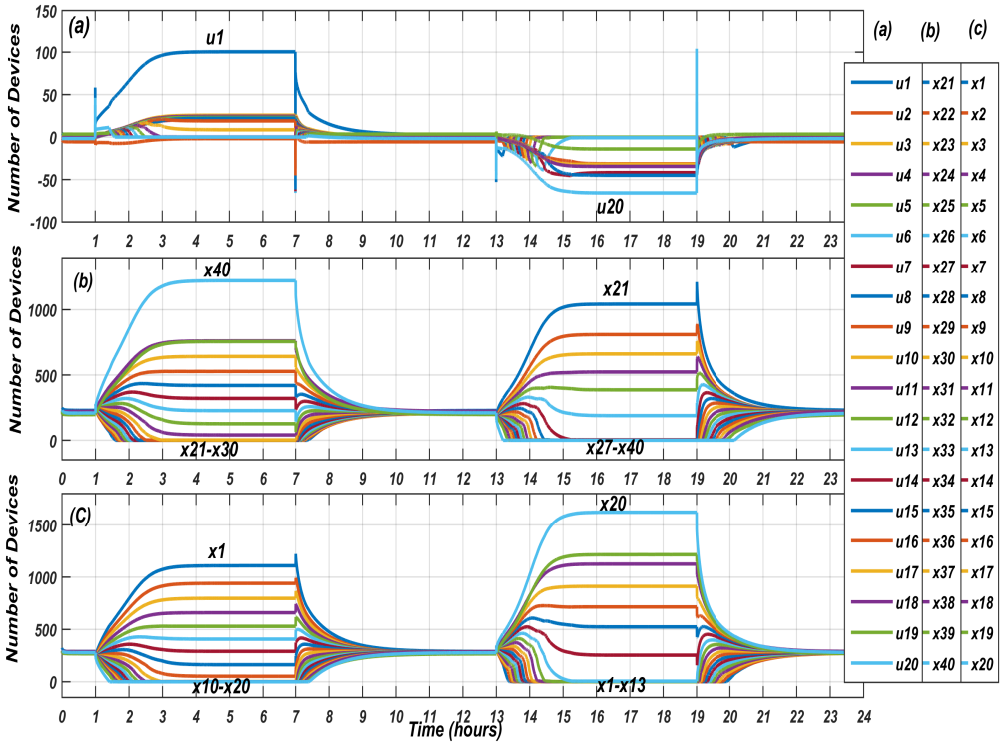

Figure 5.2 ON/OFF Switching Control. (a) States control actions. (b) Evolution of the ON state devices. (c) Evolution of the OFF state devices.

The worst case situation leads to an extreme switching frequencies at the boundary states. For instance, in the case of load-decrease, when temperature inside houses increases to state 20, the local controllers switch the devices to state 21 to start the cooling cycle. However, the central controller forces them to return back to state 20 in order to contribute in the the load-decrease service. Therefore, the local and the centralized controllers are working against each other. This action may cover a considerable number of devices and would last for the entire service-time.

However, programmable thermostats are usually equipped with a time delay unit as a short cycling protection. This time is required to allow the compressors equalize the internal pressure and be able to start again. Otherwise, the motors would stall or damaged [62]. Therefore, implementing this control technique for such services requires an extensive design consideration while in some cases it may cause devices damage.

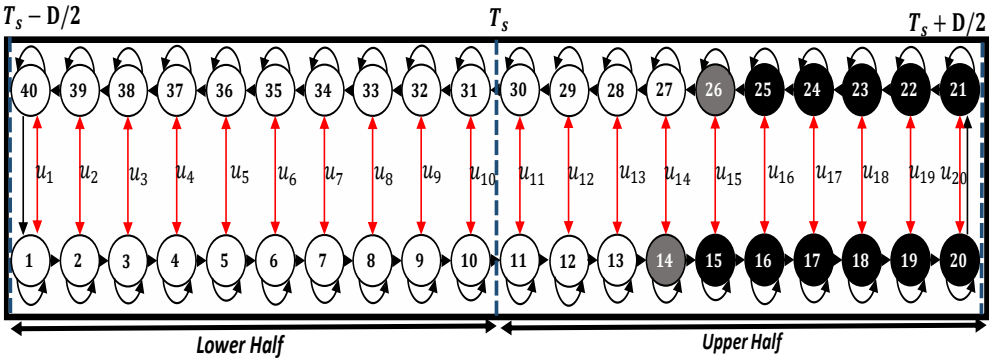

Figure 5.3 Devices distribution during the load-decrease case.

5.3.2 Direct ON/OFF Switching Control with Set-point Adjustment

This section analyze the performance of the EMM for the same system services considered in the previous section. Sensitivity analysis on the 10,000 air-conditioning devices has shown that in steady-state the aggregated power consumption changes approximately 2 MW for each 1°C variation. Therefore, achieving a 4 MW load changes requires shifting the set-point by 2°C. *i.e.*, $T_s^{'}$ is 18°C for the load-increase and 22°C for the load-decrease, given that they are initially regulated at 20°C.

Initially, we demonstrate the effect of the set-point change in extracting the services without introducing the ON/OFF control of the MPC. Fig. 5.4 illustrates the simulation for set-point adjustments as discussed before. Clearly, We can see that the aggregated power follows the reference value only in steady-state. However, during transient, set-point change causes power overshooting and oscillations. The next discusses developing the EMM model for these set-point changes and implementing the MPC controller to eliminate the oscillation.

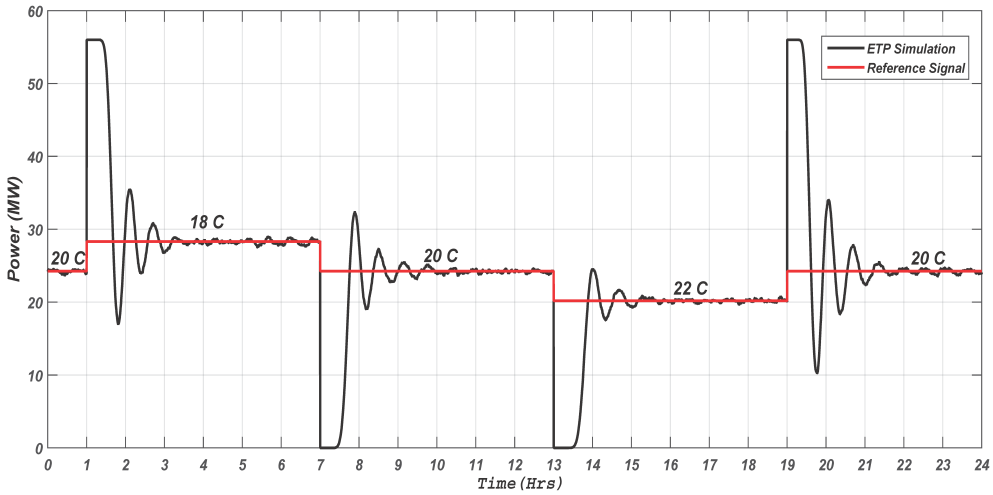

Figure 5.4 ETP simulation for adjusting the set-point (without MPC)

For each set-point change, the EMM shown in (5.1-5.3) is constructed as discussed in the previous chapter. The EMM comprises 80 transient-states and 40 final-states, the transient states represent two adjacent Markov chains. The order of the transient and final-states is based on the direction of the set-point adjustment. For instance, a set-point change from $20°C$ to $18°C$ dictates that the $20°C$ and $19°C$ are the transient-states while the $18°C$ Markov chain represents the final-states.

The EMM models are implemented with the MPC controller and the overall results are shown in Fig. 5.5. The aggregated power consumption of the new model is able to precisely follow the load variation services as instructed by the reference signal (Fig. (5.5.a)). All power overshooting and oscillations have been eliminated by applying a total switching actions as shown in (Fig. (5.5.b)). The results indicate that a large number of devices are required to switch their status when the EMM is applied, this is shown as spikes at the instant of the set-point changes. However, the total switching actions then converges to zero while the services are still provided.

The switching spikes in the EMM have different interpretation than in the previous case. Switching spikes in this case demonstrate the optimal sequential set-point control that should be applied across the devices. Initially, the set-point change is applied to all device at the same time instant,

and the EMM is formulated to describe this situation. However, this action leads to devices synchronization and the aggregated power accordingly traverse extreme conditions. To prevent such synchronization, a large number of devices are forced to stay at the old temperature set-point and wait until they finish the current operating cycle. The concept is further illustrated in following example which describes the first switching spike in Fig. (5.5.b).

For example, consider the operation at $t = 1h$, where a set-point adjustment is applied to all devices from $20°C$ to $18°C$. This set-point change instructs devices to cool the houses down. Accordingly, ON state devices will continue their cooling cycle. However, all of the OFF state devices will turn ON and starts the cooling cycles. This action causes the power to overshoot to the maximum value. *i.e.,* all devices become ON. Therefore, the MPC forces devices that are recently changed to ON to return back to their original stats in the OFF state. Devices are forced to stay regulating the temperature at the old set-point until they finish the OFF cycle, afterward, the new set-point can be applied.

Therefore, the implication of the switching spikes is to determine how to initially apply the set-point change, although they are not required to be physically implemented. The following steps summarize the set-point control actions. First, obtain the switching spike for the next time step. Then, in the current time step, apply the set-point change to all devices except those in the switching spike. Third, keep all devices included in the switching spike at the current set-point and apply the set-point change when they finish the current operating cycle.

Tracking devices' distribution in all states and the corresponding control actions are shown in Fig. 5.6 and Fig. 5.7 respectively. Devices are initially modeled using one Markov chain at $20°C$. At $t = 1h$ the EMM is formulated to move the devices to the new set-point. The devices thus exist the $20°C$ chain in Fig. (5.6.a) and enter the $19°C$ Markov chain in Fig. (5.6.b). Time is required until all devices pass through the $19°C$ Markov chain and enter the final-states at $18°C$ in Fig. (5.6.c). The figure enumerates the EMM states based on the direction of the set-point change. During the transition, control efforts are also required to prevent power oscillations and guarantee a fast convergence to the new steady-state value. Once devices enter the final-states, the system

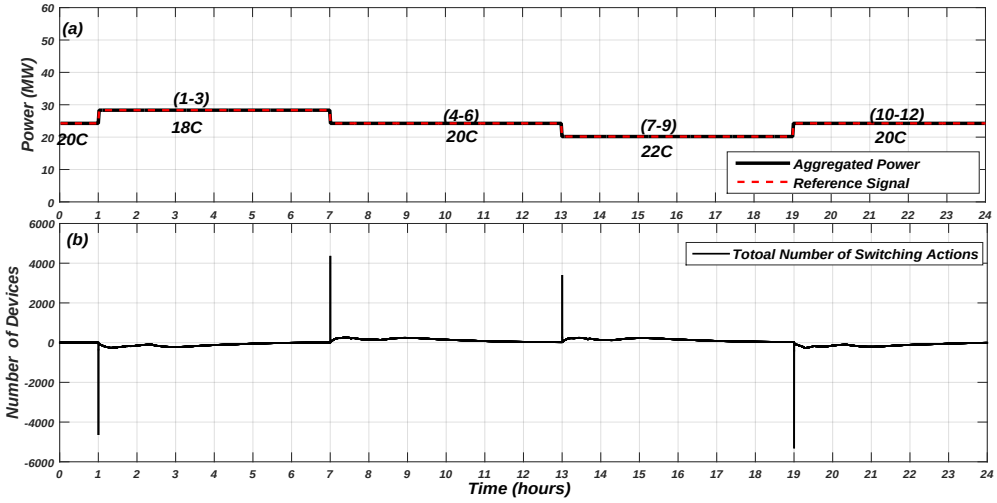

Figure 5.5 Extended Markov model. (a) Aggregated power consumption and the reference signal. (b) Total switching actions.

reaches the steady-state conditions and thus it is considered naturally stable and no further control efforts are required. At this point there will be no dead-band contraction and devices will operate according to the normal dead-band length.

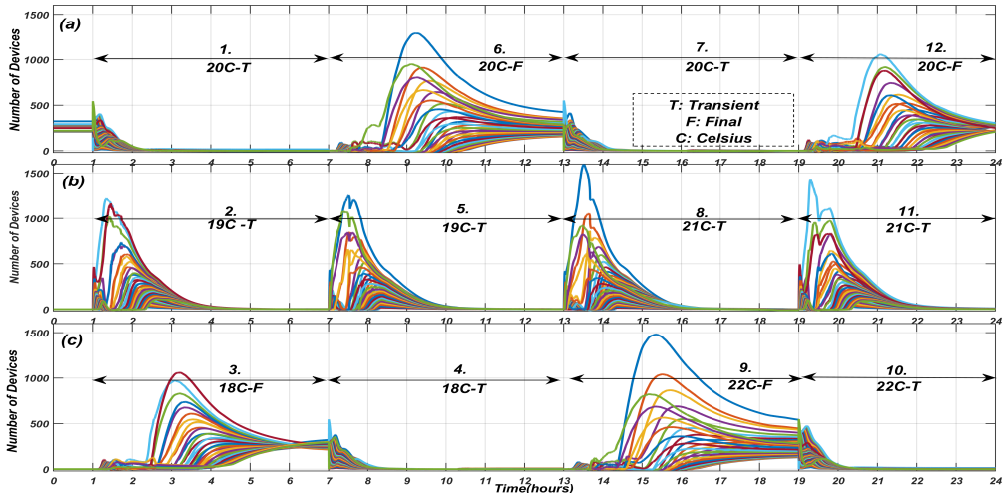

Figure 5.6 Extended Markov model devices evolution. Transient and final-states are enumerated based on set-point change direction

In both Fig. 5.5.b and Fig. 5.7 the number of devices postponed by the sequential control and the associated distribution are not shown clearly since all of the control actions happened at the same time-instant. Table. 5.1 provides better illustration about the exact number of devices and their state distribution at each time instant of applying the extended Markov model.

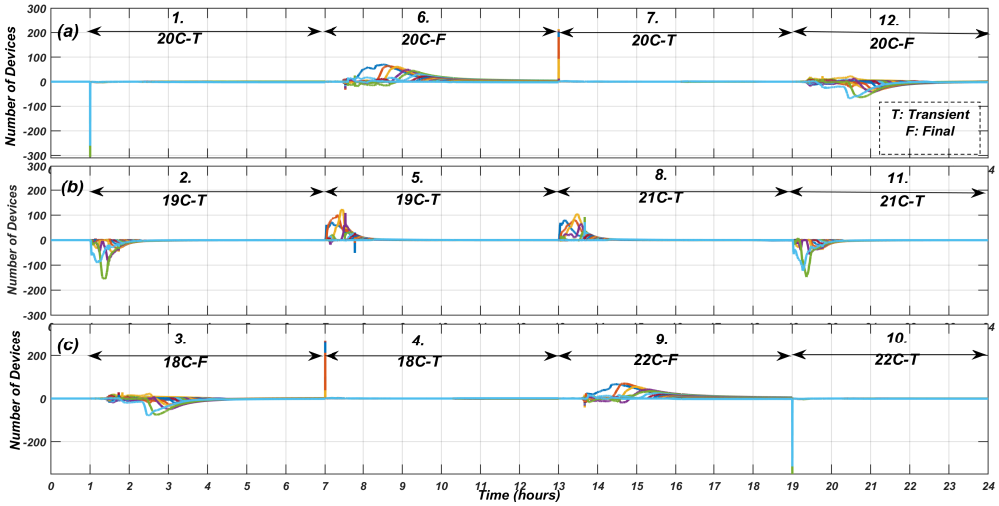

Figure 5.7 Extended Markov model control inputs for all EMM states.

5.4 Conclusion

The work in this chapter proposed a new control architecture for the heterogeneous thermostatically controlled loads (TCLs). The control is utilized to provide support to the power systems in the form steady-state load-decrease or increase services. The services are achieved through extracting or storing thermal-energy in the TCLs by modifying the operating temperature set-point. Therefore, the Extended Markov model developed in the previous chapter is utilized here to describe devices transition between the temperature set-points. It has been shown that the instantaneous set-point adjustment causes unstable power profile with weak convergence rates which may cause serious stability issues.

The Model Predictive Control framework with ON/OFF switching capability is proposed to perform two main objectives. First, provide the set-point control sequentially, and second curb

down any possible power oscillation. The results have shown that the new modeling and control techniques minimizes the devices' switching rates and short cycling compared to other models that do not rely on adjusting the thermal-energy.

Table 5.1 Sequential control algorithm (devices forced to stay at old set-point by the).

$t = 1h$		$t = 7h$		$t = 13h$		$t = 19h$	
u_i^*	# Dev.	u_i^*	# Dev.	u_i^*	# Dev.	u_i^*	# Dev.
1	1	1	266	1	172	1	2
2	2	2	272	2	214	2	2
3	1	3	266	3	213	3	2
4	1	4	267	4	207	4	1
5	197	5	269	5	205	5	120
6	275	6	269	6	204	6	294
7	291	7	269	7	204	7	337
8	292	8	269	8	204	8	337
9	293	9	269	9	204	9	343
10	293	10	269	10	204	10	344
11	293	11	269	11	204	11	344
12	293	12	269	12	204	12	344
13	292	13	270	13	204	13	344
14	293	14	268	14	203	14	344
15	292	15	261	15	201	15	344
16	293	16	212	16	182	16	344
17	295	17	39	17	94	17	345
18	301	18	1	18	0	18	351
19	306	19	0	19	0	19	358
20	360	20	0	20	0	20	315

CHAPTER 6. MARKOV CHAINS MODELING IMPROVEMENTS

6.1 Introduction and Overview

To this end, Markov chain aggregation experience three main limitations. The limitations are investigated and improved in this chapter as the followings:

- First, the heterogeneous parameters are not perfectly established to cover wide variety of devices. In this Chapter, we consider performing the analysis on a more generic and comprehensive set of parameters.
- Second, the work presented in the previous chapters is built upon single Markov chains for the entire heterogeneous system. Such representation experience a percentage of error which can goes up to 10%. The error will naturally propagate to the EMM performance and the corresponding control actions. In this chapter, clustering approach is used to divided the heterogeneous system into multiple semi-homogeneous groups, such that, the collective behavior of all clusters be able to characterize the comprehensive heterogeneous system accurately.
- Third, Markov chain aggregation assumes a constant outside temperature. Thus, the aggregated power appears as a constant value. In this chapter, Markov chains and the EMM modeling approach are improved to capture the outside temperature. This step is achieved through discretizing the outside temperature variation into small intervals and performing the identification process at each interval. As a result, the Markov chains are represented in linear time-varying settings.

6.2 Heterogeneous Parameters Estimation

There are different assumptions regards the parameters heterogeneity. A general assumption usually made by considering Gaussian distributions with specific standard deviation to thermal

resistance, thermal capacitance, and power ratings. These assumptions are initially made to contradict the homogeneous systems which consider all of these parameters are the same across devices.

The previous heterogeneous parameters are not comprehensive which can be seen as only a subgroup or single category of the actual heterogeneous TCLs. Therefore, it is required to provide more generic assumption regards heterogeneity. For instance, if we tried to increase the standard deviation of the power rating distribution to cover a wider range of devices (similar to what can be found at the distribution side), the ETP models will no longer be realistic due to the correlation between these parameters. *i.e.,* devices would be extremely oversized with very short ON cycling-time or undersized with very long operating time intervals.

In this chapter, we consider performing the analysis on a more comprehensive set of parameters. Such that, devices' power ratings are more realistic and covers a wider range of possible power categories. To perform this analysis, the heterogeneity is not assigned for the parameters themselves, instead, is assigned for the ON and OFF cycling times. This step gives a great advantage since it's much easier to specify reasonable cycling times than in assigning appropriate thermal resistances or thermal capacitance. In addition, this gives a more practical procedure, since it is much easier to measure the ON and OFF cycling time than measuring the thermal parameters themselves. The process illustrated in Algorithm. 1 gives an overview of this analysis.

The analysis starts by running a Monte Carlo sampling to specify legitimate power rating for each device. Five Gaussian distributions are considered for the power ratings as specified in Table. 6.1. For a more conservative approach, the parameters need to be evaluated at extreme operating conditions. *i.e,* relatively high outside temperature and low set-point. Accordingly, the resulting ETP models will remain valid at all other conditions. However, the ON and OFF cycling times need to be appropriately estimated when the operating condition is selected. For instance, at high outside, low set-point temperatures, the air-conditioner requires more time to cool the house. Thus, the ON cycling time is higher than the OFF time. In contrast to the low outside, high set-point temperatures, where the OFF cycling time is the higher.

Algorithm 1 Paremeter estimations for a generic hetrogenous TCLs

1: **Stage 1:** Sampling: specify the device power rating S_i.
2: **Stage 2:** Specify appropriate operating conditions T_s,T_a.
3: **Stage 4:** Sampling: Assign cycling times t_i^{off},t_i^{on}.
4: **for** device i is violating the minimum or maximum cycling times **do**
5: Replace t_i^{off},t_i^{on} with more appropriate values.
6: **Stage 5:** Parameter estimation based on the on and off cycling times equations derived from the ETP model.

$$t_i^{off} = R_i C_i ln \frac{T_a - T_s + 0.5D}{T_a - T_s - 0.5D}.$$

$$t_i^{on} = R_i C_i ln \frac{R_i S_i - T_a + T_s + 0.5D}{R_i S_i - T_a + T_s - 0.5D}.$$

7: **Output:** R_i, C_i

Table. 6.1 demonstrates the heterogeneous assumption considered in our analysis. Five power ratings categories are considered which covers small (1KW) to large (5KW) air-conditioners. Small standard deviation of 0.1 is imposed for all of the groups for more practical assumption. For each power category the charging and discharging times are specified as shown in Table. 6.1. Relatively large standard deviation of 8 minutes is considered for all of the groups. This step will ensure that the heterogeneous system covers houses with good and bad thermal characteristics. As can be seen in the table, the mean value of the charging time is higher than the discharging time since these values are assumed at relatively high out-side temperature and low set-point values. *i.e.*, $T_a = 35°C$, and $T_s = 17°C$.

Table 6.1 Assumptions of heterogeneous parameters (mean vlues are provided in minutes)
.

Category		Rating (KW)		t^{ON} (minutes)		t^{off} (minutes)	
#	# of A/Cs	Mean	SD	Mean	SD	Mean	SD
G1	5,000	1	0.1	20	8	15	8
G2	5,000	2	0.1	25	8	20	8
G3	5,000	3	0.1	30	8	25	8
G4	5,000	4	0.1	35	8	30	8
G5	5,000	5	0.1	40	8	35	8

The results of performing the procedures specified in Algorithm. 1 are illustrated in Fig. 6.1. The same operating conditions assigned in the initial assumptions are also used in the procedure ($T_a = 35°C$, $T_s = 17°C$). Clearly, we can see the correlation between devices' power ratings and their thermal characteristic. As the device's power rating increases, the value of its thermal resistance decreases, while its thermal capacitance increases. The obtained values are fixed throughout our analysis and used for the other operating conditions.

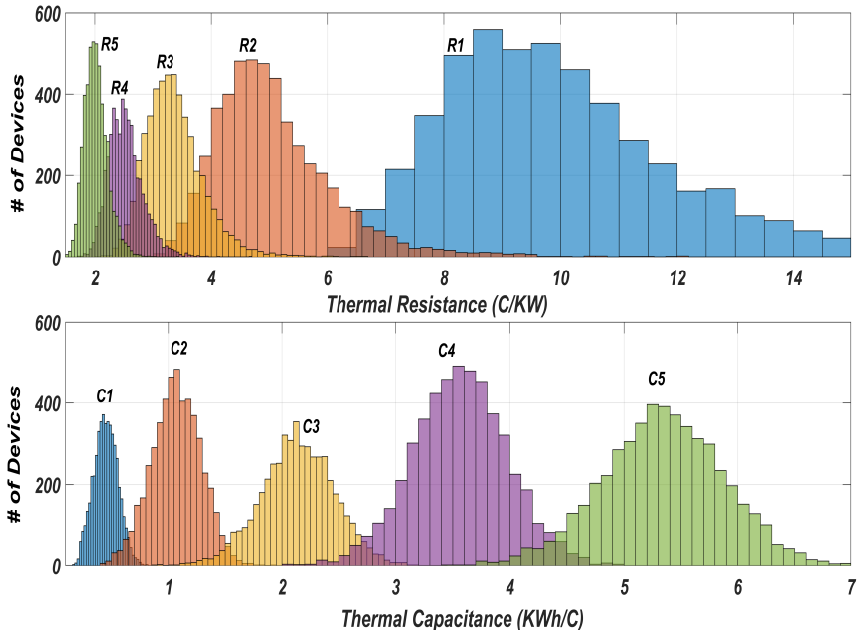

Figure 6.1 Heterogeneous parameter results (Top: Thermal Resistance $°C/kw$).(Bottom: Thermal Capacitance ($kwh/°C$). (R_i, C_i corresponds to G_i in Table.6.1).

6.3 Initial Clustering Based on Power Ratings

Under the new heterogeneity assumption, it is not practical to represent all devices in a single Markov chain, higher errors are expected due to the wide variety of the power ratings. Accordingly, more accurate performance would result if we represent each power group shown in Table.6.1 with an independent Markov chain as shown in (6.1)-(6.3). Where, N_g is the optimal number of states representing each power group, and A_g is the corresponding Markov chain. The statistical learning process and the optimal number of state analysis described in the previous chapters are performed

here for each power rating group. The normalized root mean square error is found for each group. It is assumed that all devices are initially ON to compare the performance under the worst case scenario. As illustrated in Fig. 6.2, the results show that each power group has different optimal number of states to obtain the best performance with minimal error.

$$x_g(k+1) = A_g x_g(k), \quad A_g \in R^{N_g \times N_g}, \; x \in R^{N_g} \tag{6.1}$$

$$P_{agg}(k) = \sum_g C_g x_g(k), \quad C_g \in R^{1 \times N_g} \tag{6.2}$$

$$C_g = (\bar{S}_g)[\mathbf{0}_{(1\times N_g/2)} \; \mathbf{1}_{(1\times N_g/2)}] \tag{6.3}$$

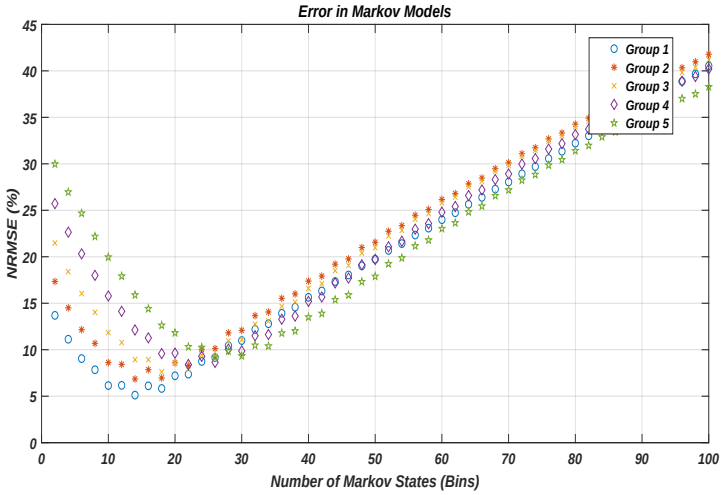

Figure 6.2 Optimal number of state for each power group.

The time domain simulation of each power group is shown in Fig. 6.3. The performance of the Markov chins are compared with the ETP simulation at the extreme initial conditions (worst case scenario). The result of combining the five Markov chains is also shown in the right-bottom graph. Clearly, we can see that the performance of the Markov chins has a percentage of error. Relatively large deviation resulted when comparing the performance of the Markov chains during the transient period. The Models do not have perfect match with the ETP simulation, while in

some cases the dynamics is out of phase. The details of the normalized root mean square error for all of the groups are shown in Table.6.2. The combined performance,*i.e.,* the Markov chains of the five groups together, has a total error of 6.6.39 %.

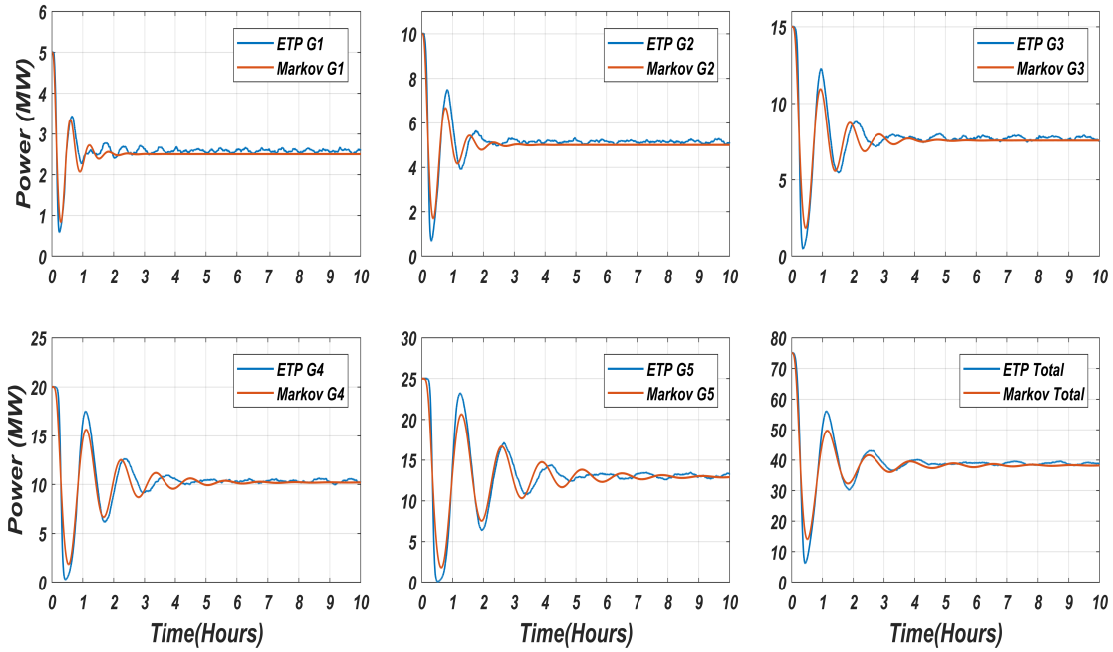

Figure 6.3 Markov chain perofmance at the optimal number of states.

In the next section, our objective is to further divide the heterogeneous TCLs into cluster in order to reduce this error and improve the Markov chain representation.

Table 6.2 Optimal number of bins and the resulting error.

Group Rating	optimal bins	% error
G1	14	5.1169
G2	14	6.8525
G3	18	7.6087
G4	22	8.4093
G5	26	9.1961
Combined	-	6.3904

6.4 Devices Clustering Based on the Charging and Discharging Time Characteristics

Markov chains are initially developed for homogeneous systems. such that, all devices under control share the same power and thermal ratings. Accordingly, the results of the corresponding Markov chains are very accurate. However, for heterogeneous systems, due to the variations in the speed of charging and discharging rates. Single Markov chains will only capture the mean behavior of all devices. Therefore, in this section, the K-mean clustering is utilized to systematically divide the devices into multiple clusters with similar charging and discharging characteristics. such that, the entire heterogeneous system is approximated with multiple semi-homogeneous systems. In this analysis, each device is represented by a single data point $v_i = (t_i^{on}, t_i^{off})$, which combines the time required for the cooling and the hating cycles. The objective in Eq.(6.4) is set to minimize the squared Euclidean distances between data points and the corresponding nearest mean μ_k; where K is the number of the required clusters. A binary variable $x_{i,j}$ is used to designate single devices in only one cluster as demonstrated in Eq.(6.5). while in Eq.(6.6), the mean value of each cluster is updated. Sensitivity analysis has shown that dividing each power group into 10 clusters, (see the results in Fig. 6.4), highly improves the accuracy and provides satisfactory performance. The validation results are provided in the next section.

$$MinJ = \sum_{i=1}^{N_g} \sum_{j=1}^{K} x_{i,j} \|v_i - \mu_j\|^2 \tag{6.4}$$

$$\sum_{j=1}^{K} x_{i,j} = 1, \quad \forall i \in N_g \tag{6.5}$$

$$\mu_j = \frac{\sum_i x_{i,j} v_i}{\sum_i x_{i,j}}, \quad \forall j \in K \tag{6.6}$$

$$x_{i,j} \in \{0, 1\} \tag{6.7}$$

The detailed results of the k-mean clustering technique is demonstrated in Table. 6.3 (K=10). The outcome of the clustering optimization are the clusters' mean (μ_k) and the classifier index

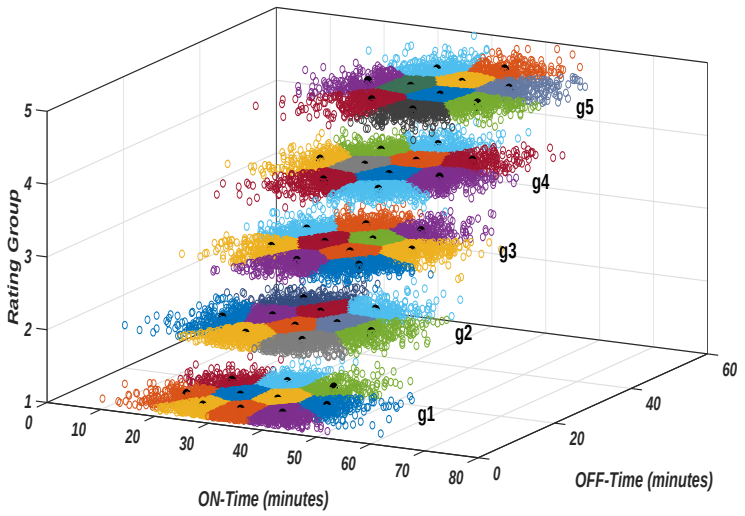

Figure 6.4 Illustration of the clustering results (10 CLusters/Group).

$(x_{i,j})$ which designates the devices to the corresponding clusters. It is important to emphasis here that the clustering optimization is a non-convex problem. Therefore, the outcome is a local optimal solution. Repeating the clustering algorithm will result in a different local optimal solution. The clusters' mean may slightly move. Accordingly, Some devices previously considered in one specific cluster would immigrate to another adjacent cluster, especially those devices located far from the mean at the boundaries. Therefore, it is important to perform this analysis only once and fix the solution for all other operating conditions. The results presented in Table. 6.3 illustrate the number of devices in each cluster and the corresponding clusters' mean (t^{on},t^{off})

6.5 Markov Model Performance under clustering

The heterogeneous devices are now clustered into multiple groups, such that, devices belongs to each cluster have almost similar charging and discharging characteristic. To evaluate the performance of the new devices representation, each cluster is processed independently. Single Markov chain is identified to each cluster based on the inside temperature and power consumption time series data (see Fig. 3.2). The data are generated by running the corresponding ETP models.

Table 6.3 Clusters' mean and associated number of devices (Time is provided in minutes)

	Group G1			*Group G2*			*Group G3*			*Group G4*			*Group G5*		
	#	ton	toff	#	ton	toff	#	ton	toff	#	ton	toff	#	ton	tof
C1	472	21.0	27.2	412	18.9	31.5	576	19.6	20.0	664	31.9	46.7	649	34.0	44.7
C2	646	14.4	14.1	919	19.7	24.1	409	28.2	48.6	194	33.4	63.3	466	35.7	63.1
C3	714	18.1	23.0	153	25.6	48.8	416	23.5	42.0	667	25.5	30.4	424	30.7	55.8
C4	414	8.8	6.7	277	10.6	8.1	773	26.9	36.2	482	33.2	54.2	184	24.4	25.4
C5	180	22.5	41.3	588	24.4	33.6	241	14.7	12.4	897	27.3	35.8	734	34.8	50.6
C6	565	11.9	10.5	287	20.9	39.8	496	21.7	33.5	368	27.5	50.8	545	28.2	33.8
C7	927	17.3	18.1	744	17.9	19.0	793	22.1	25.9	142	18.3	16.6	172	37.6	71.9
C8	261	15.5	30.0	742	22.8	27.7	669	25.4	30.4	398	23.0	24.3	762	30.3	40.0
C9	321	21.9	33.5	569	15.0	14.0	141	30.0	56.7	634	31.0	40.5	564	29.2	47.0
C10	500	13.7	18.9	309	26.9	40.0	486	29.6	42.0	554	26.3	42.9	500	36.8	56.3

In order to gain the best performance, the sensitivity analysis performed earlier in Chapter.3 is repeated here to find the best number of states representing the individual clusters. The Markov chain performance is evaluated for various number of states. The evaluated is found only under the worst-case initial condition (all devices are initially in the ON-state). Clearly, we can see in Fig. 6.5 that for all of the clusters, there is only one optimal number of states to represent the cluster.

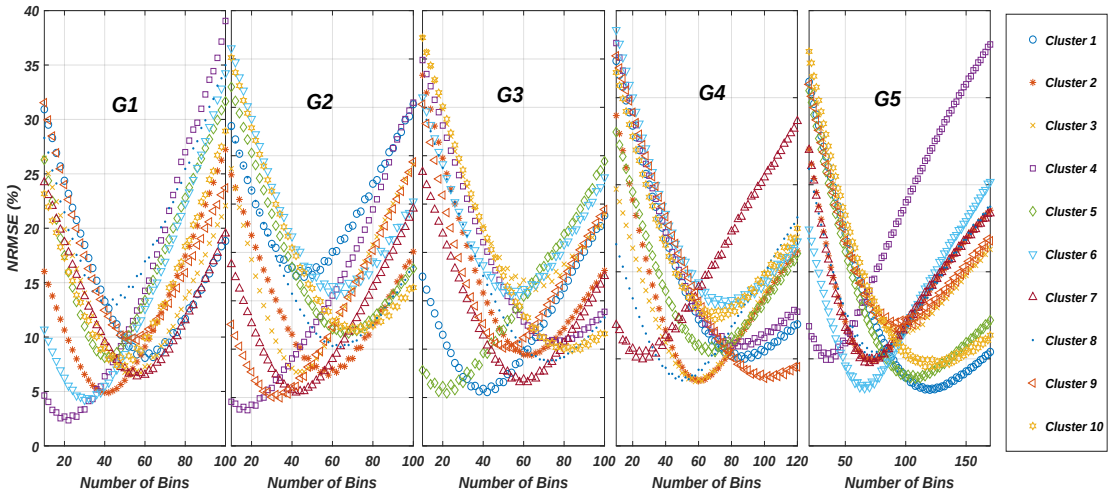

Figure 6.5 Optimal number of states for individual clusters.

The clustering approach highly improves the modeling performance. It is clear in Fig. 6.6 that the error is decreased from individual groups point of view and the combined response point view

(shown in the right-bottom graph). Under clustering, the resulting normalized root mean square errors are shown in Table.6.4. From individual groups point of view, the error is significantly reduced in all of the cases, while in the combined response the error is reduced to 2.087%, compared to 6.39% without clustering.

The system in this case is represented by 50 Markov chains (10 clusters for each power group) as shown in Eqs. (6.8)-(6.10). The index g represents the group, while the index ℓ indicates the cluster. Considering more clusters would further improves the results. However, this step will increase the computation complexity for both the predictions and the control applications. Simulating all of the clusters for the 10-hours horizon, with 10 seconds time-step, last approximately 5.32 seconds.

$$x_{g,l}(k+1) = A_{g,l}\, x_{g,l}(k), \quad A_{g,l}(k) \in R^{N_{g,l} \times N_{g,l}},\ x_{g,l} \in R^{N_{g,l}} \tag{6.8}$$

$$P_{agg}(k) = \sum_{g} \sum_{l} C_{g,l} x_{g,l}(k), \quad C_{g,l} \in R^{1 \times N_{g,l}} \tag{6.9}$$

$$C_{g,l} = (\bar{S_g})[\mathbf{0}_{(1 \times N_{g,l}/2)}\ \mathbf{1}_{(1 \times N_{g,l}/2)}] \tag{6.10}$$

Table 6.4 Optimal number of bins and the resulting error.

Group Rating	optimal bins (10 clusters)	% error
G1	(60,42,54,22,40,30,56,42,56,50)	3.0315
G2	(46,58,42,18,68,62,42,64,34,72)	3.8188
G3	(42,64,66,76,22,54,58,72,50,84)	3.078
G4	(88,60,58,88,70,78,26,50,100,70)	3.7215
G5	(120,72,94,38,108,66,70,74,94,120)	3.5291
Combined	-	2.087

6.6 Outside temperature variations

The outside temperature (T_a) plays a significant role in determining the TCLs' aggregated power. Summer peaks are always considered as a big challenge to systems' operators which is

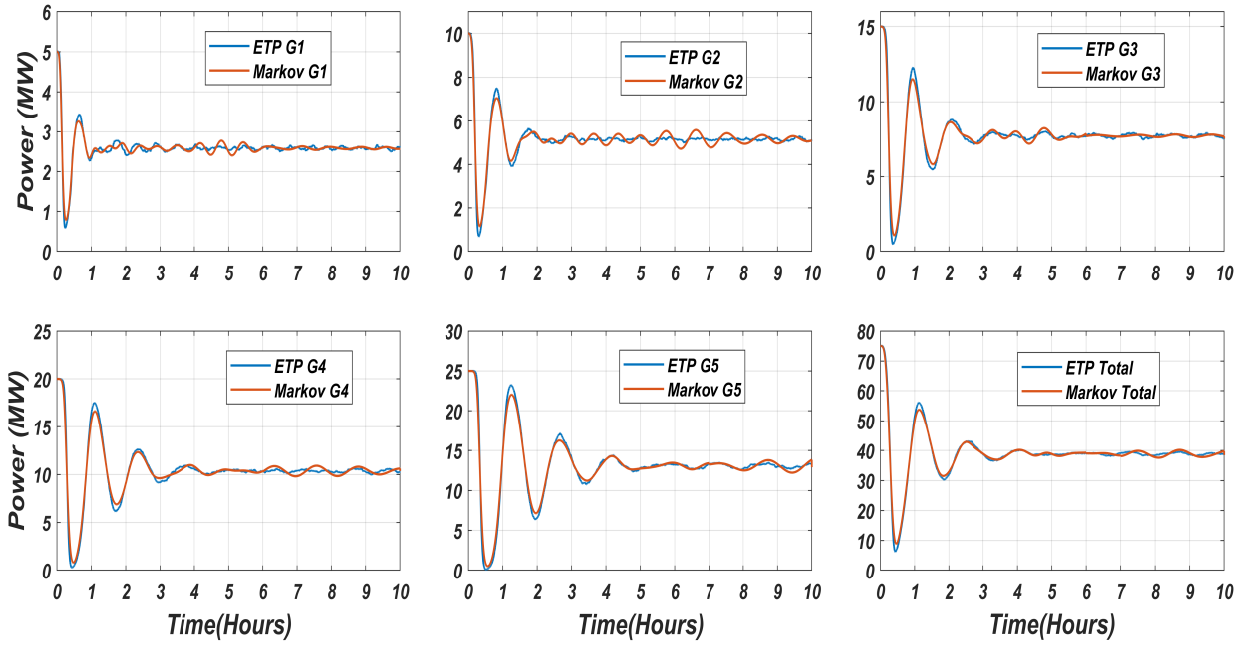

Figure 6.6 Markov models performance under clustering

mainly caused by the air-conditioners penetration. Therefore, for more realistic results, the aggregated model should appropriately consider the variation of $T_a(k)$, and captures the changes in the aggregated power accordingly. Markov chain identification techniques are incompetent in determining time-varying transition probabilities. In general, the techniques are based on calculating the average probabilities over a time-series observations generated by running the ETP models. However, for more practical results, it is important to improve the Markov chain representation and its identification process to capture this variation. One possible way to conduct this analysis by discretizing $T_a(k)$ into small intervals and repeat the identification process at each small interval. Such that, the resulting model will be time-invariant inside the small intervals while time-varying outside. In this way, the Markov chains will be able to capture the low frequency component of $T_a(k)$.

Temperature set-point control will add another dimension to the problem complexity. The characteristics of the Markov chain and the internal transition probabilities will also vary as the

temperature set-point changes. Therefore, we considered building a comprehensive database which covers all possible combinations between the set-points and out-side temperature variations. This database is more generic and can be used to track any out-side and set-point temperature profiles. The database considers 0.1 Celsius discretization-step for both the set-point and the outside temperature. An illustration is shown in Fig. 6.7, Each single point represents 50 Markov chain since the heterogeneous system is now modeled by 5 power groups and 10 clusters for each group.

Eigenvalue analysis is performed to all of the Markov chains as described previously in Section. 3.5. This analysis yields the invariant distribution for each temperature setting. The invariant distribution which refers to the devices' distribution during steady-state can be used to find the aggregate power of each independent cluster. The aggregate power is found using Eq. 6.9. Accordingly, the database now contains the Markov chains and the corresponding aggregated power for each cluster. The database information will be used later for the control applications.

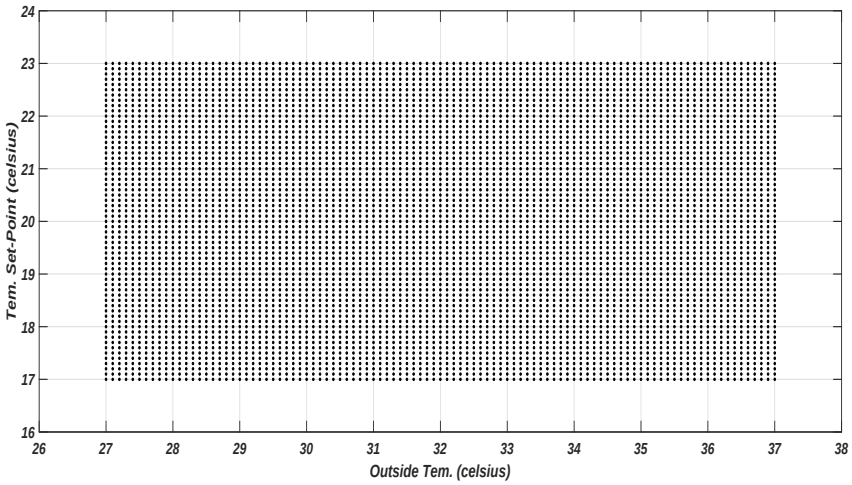

Figure 6.7 Illustration of Markov Chains Database. Direct switching for transitions along the x-axis, EMM for transitions along the y-axis

6.7 Markov Model Performance under out-side temperature variations

This section validates the Markov chain performance under two temperature variations scenarios. First, only the outside temperature variation is considered. In this case, the Markov chins

make direct switching along the x-axis of Fig. 6.7. Second, the outside temperature variation is considered with set-point adjustment. Therefore, the Extended Markov model will be used to make the transitions along the y-axis of Fig. 6.7.

In the first scenario, two outside temperature profiles are considered as shown in Fig. 6.8(a). The actual temperature variation is approximated by 0.1 Celsius discretization to agree with the database developed in the previous section. Based on this discretization, the Markov chains need to make in total 60 switching action to represent Ta_1, and 80 times to capture Ta_2. Transitions happen along the x-axis. The Markov chain approximation is compared with the actual ETP simulation in Fig. 6.8. Clearly, this new time-varying representation accurately captures the out-side temperature variations with a total error less than 1% in both cases.

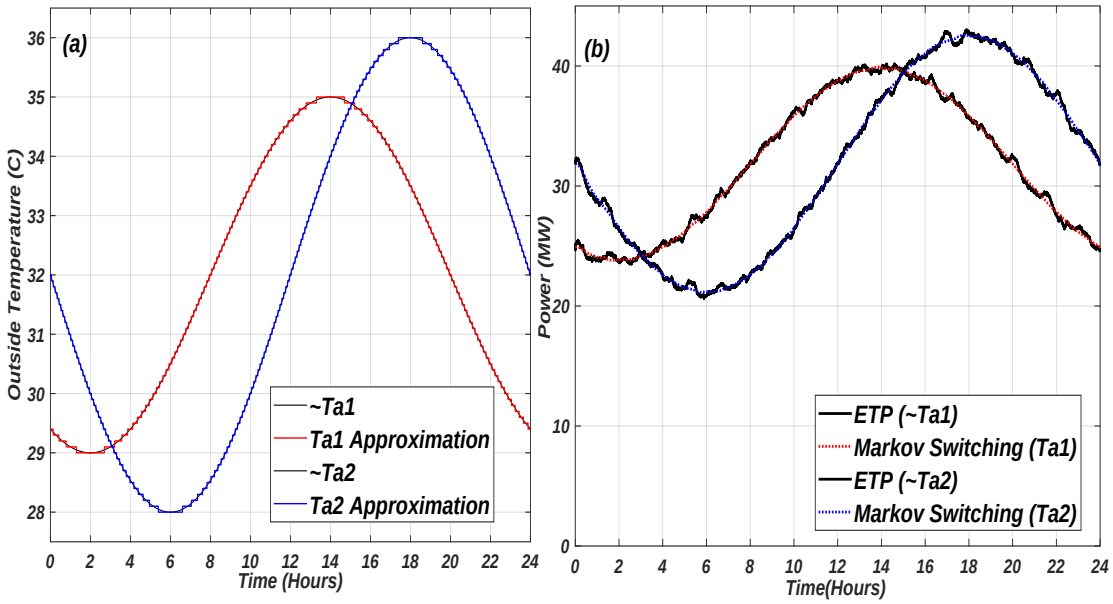

Figure 6.8 Markov chain performance under outside temperature variations.

In the second scenario, we consider validating the Markov chains under both the set-point adjustments and the outside variations. To align this section with our objectives in the next chapter; mainly, utilizing the TCLs to improve the power system flexibility. The Markov chains are validated to describe the TCLs in providing three main ancillary services. Namely, emergency

spinning reserves, load reduction, and load shifting. Set-point adjustment is the main tool to extract the three ancillary services. For the emergency reserves, we consider increasing the set-point at $t = 8hr$ from $20°C$ to $22°C$ for only one hour as shown in Fig. 6.9. This set-point change is order to achieve fast load reduction which can help the system during emergencies. From the Markov modeling point of view we can see that the Extended Markov Model (EMM) can accurately capture both, the out-side temperature variations, and the set-point adjustment. In the next chapter, the EMM model and the MPC controller will be used to derive appropriate control actions to smooth the oscillations and prevent all possible synchronous operations.

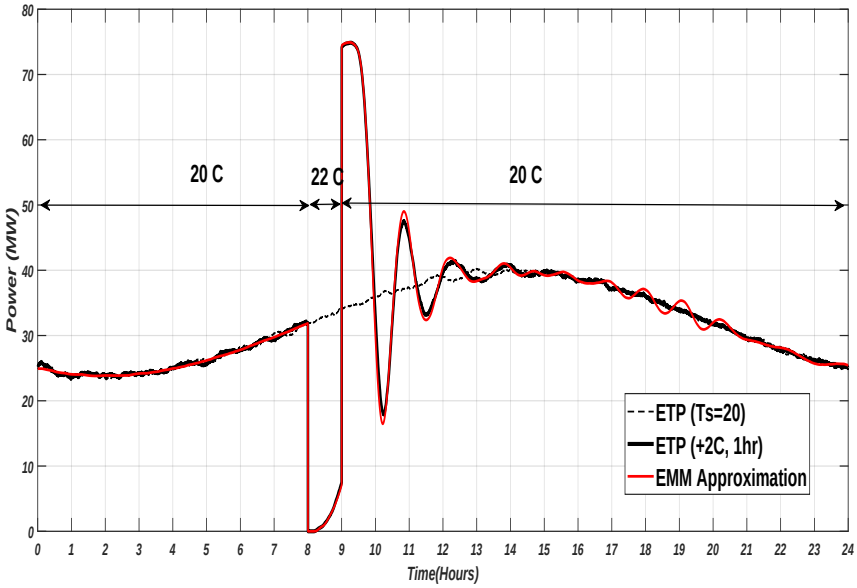

Figure 6.9 Extended markov model performance for spinning reserves ancillary services.

The second temperature set-point control is designed for load reduction applications. In this case, the set-point is also adjusted from $20°C$ to $22°C$ but is implemented for a longer duration. All devices are instructed to increase the operating set-point during the peak-hours. The service starts at hour 10 and finishes at hour 16. Smoothing the fast charge and discharge will be discussed in the next chapter. Clearly, we can see that the EMM approximation, to a high accuracy level, captures both the set-point change and the outside temperature variations.

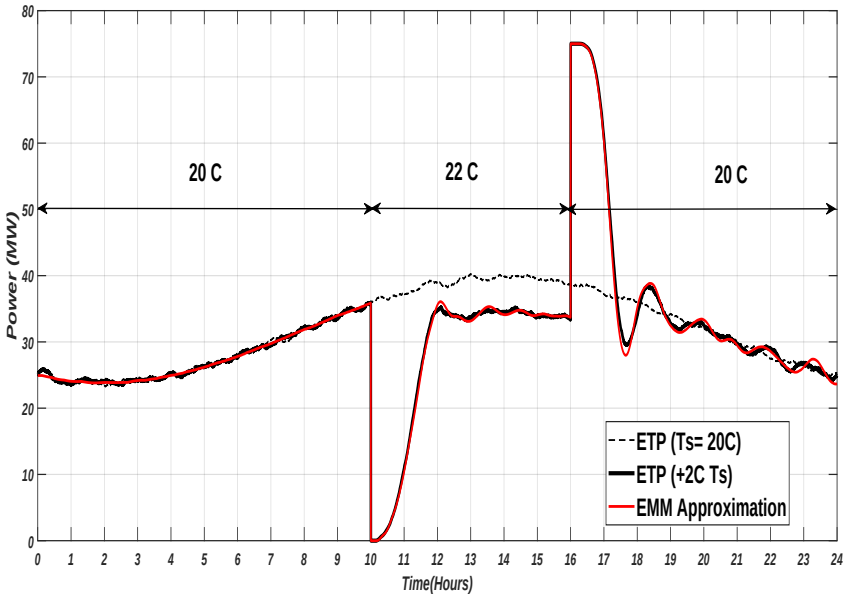

Figure 6.10 Extended markov model performance for load reduction ancillary services.

The last set-point change is designed to provide load shifting aspects. *i.e.,* devices are pre-cooled before the peak hours in order to be able to provide load reduction during the peak hours. Devices are instructed to decrease their operating set-point to $18°C$ before the peak hours and resume their normal operations at $20°C$ during the peak hours. Regardless the fast charging and discharging rates, the EMM approximation performs well in approximating the set-point change and the ambient temperature variations.

6.8 Conclusion

This chapter has improved the Markov chains performance under two main aspects. First, the Markov chain representation has been improved to capture more comprehensive heterogeneous parameters. The results have shown that Markov chains are more accurate if the heterogeneous devices are divided into multiple semi-homogeneous clusters. *i.e.,* devices share similar power ratings, charging time, and discharging time characteristics. Under clustering, the collective behavior of all clusters is able to accurately characterize the entire heterogeneous system. In the second aspect,

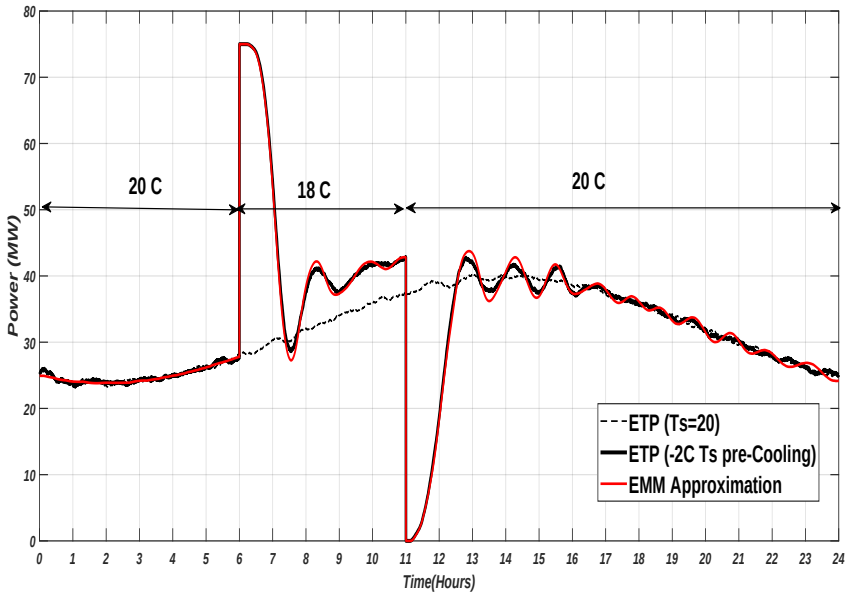

Figure 6.11 Extended markov model performance for load shifting ancillary services.

the outside temperature variations is considered. A database is developed based on discretizing both the set-point control and possible outside temperature variations. The results have shown the extended Markov model can accurately capture both the set-point adjustments and ambient temperature variations.

To attain this accuracy, two level of complexity are added to the problem. First, the number of Markov chains and their corresponding states have been increased. Second, to capture the outside temperature variations, the extended Markov model become a time-varying system. To account for these challenges, a decentralized time-varying model predictive control is proposed and discussed in the next chapter.

CHAPTER 7. AGGREGATION IN THE DAY-AHEAD MARKETS FOR ANCILLARY SERVICES

7.1 Introduction and Overview

This chapter investigates the capability of the Thermostatically Controlled Loads (TCLs) in providing three main ancillary services. The services are designed as Demand Response (DR) programs and integrated to the Day-head energy and reserves market. The Security constrained Unit Commitment (SCUC) problem is used to represent the day-ahead market. The three ancillary services are; the emergency spinning reserves, the load reduction, and the load shifting DR programs. System operators conduct the SCUC as a tool to dispatch the available conventional thermal units including the DR resources in hourly basis for the next day.

The DR ancillary services are provided through charging or discharging the TCLs with additional thermal energy proportional to the required power performance. Temperature set-point adjustment is the main control tool to modify the aggregate power. Both the Extended Markov Model (EMM) and the Model Predictive Control frameworks (MPC) are utilized in this chapter to modify the aggregated power based on the requirements of each ancillary service. Quantifying the TCLs' capability is made based on the assumption that the customers agreed to allow their temperature set-point being adjusted within $\pm 2^{\circ}C$.

The heterogeneous parameters and the clustering results obtained in the previous chapter are used in our analysis in this chapter. In case of load increase or decrease dispatch, each cluster is requested to modify the set-point and achieve certain power services proportional to the number of devices belongs to the cluster. The DR programs are designed as Mixed integer linear models with appropriate capability and charging/discharging constraints.

7.2 Time-varying Model Predictive Control Framework

Load aggregators are the entities responsible for controlling the devices at the distribution side. Under perfect monitoring and control conditions, aggregators are required to quantify the capability of the underlying devices based on the comfort level specified by the customers' contracts. Under each aggregator, devices are clustered based on their power ratings and the charging and discharging rates as discussed in chapter.6. In order to simplify the complexity of the control problem, it is more practical to move from the centralized to a decentralized approach. Such that, the control problem specifies each cluster (l) as an independent entity. Accordingly, the advancements in parallel computing can make the solution of such problems very fast and suitable for online environments.

The control problem is modified as follows; The performance index shown in (7.1) is subjected to minimize the deviation between the cluster's aggregated power $P_l(k)$ (7.3) and the associated reference signal P_l^{ref}. There is no limitation governing the reference signal. However, in order to completely eliminate the ON/OFF switching actions and rely solely on the stored thermal energy, the reference signal should be the expected power value at the new temperature set-point. The expected power can be inferred by performing eigenvalue analysis to the Markov chains identified previously in the offline process (Section. 6.6).

During the transition between the set-points, devices switching (ON or OFF) is needed $u_l(k)$ to maintain the stability of the aggregated power by preventing synchronous operations. However, during steady state, when all devices reach the new set-point, no further ON or OFF switching actions are required, and devices will naturally achieve the load increase or load decrease services. The EMM model shown in (7.2) is formulated to describe devices transition between the old and the new set-points. Details regards the control input matrix (B) and the physical meaning of the

none-negativity constraint (7.4) can be found in Section.(5.2). Note that the EMM formulation (7.2) holds the time-varying setting to account for the outside temperature variations.

$$MinJ = \sum_{k=k_i}^{k_f} Q(P_l(k) - P_l^{ref})^2 + u_l(k)^T R u_l(k) \tag{7.1}$$

$$z_l(k+1) = A_l(k) z_l(k) + B_l u_l(k) \tag{7.2}$$

$$P_l^{agg}(k) = C_l z_l(k) \tag{7.3}$$

$$z_l(k) \geq 0 \tag{7.4}$$

The control problem is used to assess the TCLs under the three DR ancillary services and discussed in the following sections.

7.3 Spinning Reserves Ancillary Services

The balance between generation and load has to be always maintained. Generation deficiency causes the power frequency to drop significantly making the system vulnerable to brown or black-outs. In case of emergencies, load shedding schemes are implemented to disconnect large portion of the load and save the system. Therefore, systems operators have to maintain enough amount of spinning reserves available online in case of emergencies. Spinning reserves are required to ramp very fast to supplement any possible large deficiency in generation. To maintain the system reliability, it is usually required to have enough amount of reserves equal to the largest unit committed in the system. The spinning reserves can also be provided by the load instead of generation. Fast load reduction in case of emergency has the same effect as the power provided by the high ramping units. This section discusses how such reserves can be extracting by implementing temperature set-point control on TCLs.

The firs DR scenario (shown in Fig. 7.1) assumes fast thermal energy discharge, the operating set-point is suddenly increased from $20°C$ to $22°C$ for a period of one-hour. Since the new operating set-point is higher than the actual indoor temperature, devices react to this instruction by turning to off-state, while their inside temperature start slowly increasing to $22°C$. This fast transition from ON to OFF causes the aggregated power to experience an abrupt fall similar to the load-shedding schemes implemented by the operators in case of contingencies. This is considered as a great advantage the TCLs can provide. *i.e.,* the power is lost for a certain amount of time while devices are not completely disconnected. In this case, the stored thermal energy is utilized to make this service. The thermal energy dissipation is relatively slow, devices will take long time until the internal temperature reaches the $22°C$.

However, at the end of the service time, when all devices are instructed to charge again by resuming their normal operations at $20°C$, the aggregated power overshoot to a very high value and starts to oscillate due to devices' synchronous operation (can be seen by the ETP simulation with T_s control and its EMM approximation). This situation can be harmful to the power system and must be mitigated in a coordinated control. Therefore, the transition from $22°C$ back to $20°C$ must be implemented in a sequential pattern which revealed by implementing the MPC.

The blue curve in Fig. 7.1 demonstrates the aggregated power behavior after implementing the MPC control actions (shown in the bottom graph). The result shows that the aggregated power stability can be maintained if suitable set-point time delay is specified for $11,849$ devices. Since the EMM combines both the Markov chains at the old and the new set-points, the MPC forces large number of devices to stay one more cycle at the old set-point before they can make transition to the new set-point. The time required for each device to finish the current operating cycle is the sequential control time delay. In addition to the optimal sequential set-point adjustments, the MPC performs ON/OFF switching actions to curb the oscillations and provide smother responses. Once all devices return to $20°C$, the switching actions almost converges to zero.

During the service time, zero power consumption is given as the reference signal to all clusters *i.e.,* $P_l^{ref} = 0,\ 11hr \leq k \leq 12hr$. Since the unforced response can not maintain zero power

consumption for the entire service time, ON/OFF switching actions are exerted by switching devices to OFF (the negative sign indicates a forced transition from ON to OFF). At the end of the service time ($k = 12hr$), the MPC controller prevents all devices to return back simultaneously to $20°C$, large number of devices are delayed in order to give smooth return, otherwise, the power will overshoot and oscillate afterward. The reference signal given at ($k = 12hr$) is the cluster's nominal power at that time which is identified using the eigenvalue analysis.

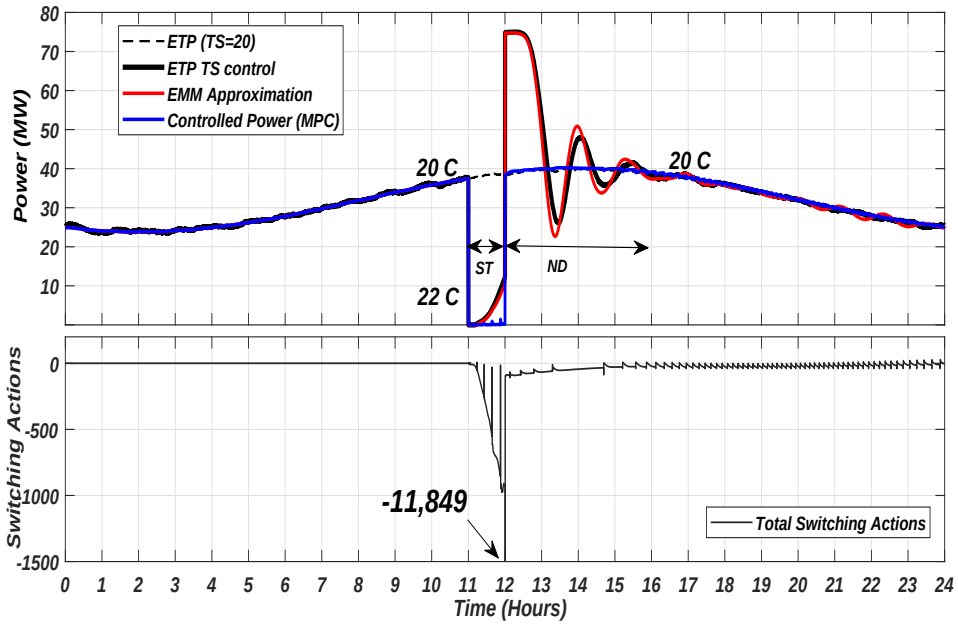

Figure 7.1 Demand Response Emergency reserves. Top: Aggregated Power. Bottom: total control action perfomred by the MPC.

7.3.1 Demand Response Program for Spinning Reserves

Dispatching the TCLs to provide such services requires specifying the time constraints associated with the ancillary service. ST is defined here as the service time which is equal to one hour as illustrated in Fig. 7.1. At the end of the service time, the permitted thermal energy is already discharged. Therefore, it is required to charge the devices again at $20°C$ to allow another dispatching signal. The charging time is relatively slow due to both the natural slow temperature trajectories and the time delay imposed by the sequential control. The no dispatch time (shown as ND in

Fig. 7.1) is estimated by the load aggregators and notified to the operators to prevent violating the customers' comfort.

Based on the previous discussion, the model shown in (7.5-7.9) allows system operators to dispatch the aggregate (i) at hour (t) to supply an amount of reserves equal to (DR^{Res}). This amount is thus bounded by the maximum load reduction capability under the fast discharge scenario. This value represented by $P^{Max}_{(i,t)}$ in (7.5) which represents the average aggregated nominal power at each time interval. The binary variable (X^{Res}) is used as commitment indicator.

Once an aggregator is dispatched to discharge the devices at any given hour, aggregators need time to apply the control and bring the temperature back to the nominal value. The no dispatch time ND is imposed by constraint in (7.6). For coordination purposes, a flag F^{Res} is introduced in (7.7) to indicate all time intervals the aggregator is committed to provide reserves (This includes both the service time and the no dispatch time). This flag will prevent the same aggregator being dispatched for other DR services during the discharge and the charging modes. Similarly, in (7.8) aggregators will not be able to provide the emergency reserves while providing other DR services. *i.e.,* load reduction F^{Lr} and load shifting F^{Ls}, both will be defined in their programs in the next sections. (Note: constraint (7.8) is repeated twice for each flag but combined for simplicity).

$$0 \leq DR^{Res}_{(i,t)} \leq P^{Max}_{(i,t)} X^{Res}_{(i,t)} \tag{7.5}$$

$$1 - X^{Res}_{(i,t)} \geq \sum_{t-ND_i}^{t-1} X^{Res}_{(i,tx)} \tag{7.6}$$

$$\sum_{t}^{t+ND_i} F^{Res}_{(i,tx)} \geq (ND_i + 1) X^{Res}_{(i,t)} \tag{7.7}$$

$$X^{Res}_{(i,t)} \leq 1 - F^{Lr/Sh}_{(i,t)} \tag{7.8}$$

$$X^{Res}, F^{Res}, F^{Lr}, F^{LS} \in \{0, 1\} \tag{7.9}$$

7.4 Load Reduction Ancillary Services

Other ancillary services do not require the fast thermal energy discharge. For instance, for peak clipping applications, longer time frames are involved. Such load reduction can be performed by increasing the operating set-point during the service time and return it back to the nominal value afterwards. However, the fast thermal energy discharge and charge are curbed by implementing the sequential set-point control, such that, the aggregated power stability can be maintained as dictated by the reference signal. The difference in the load reduction program compared to the emergency reserves is that devices are instructed to operate at higher set-point for a longer time span.

An example of this scenario is provided in Fig. 7.2. The analysis here assumes that the set-point adjustment can be made for 5 hours (The value of service time ST is usually constraint to customers' contracts. *i.e.,* for how long it is allowed to stay at the maximum set-point). The temperature set-point is adjusted at $t = 11\,hr$ from $20°C$ to $22°C$, while at $t = 16\,hr$ is brought back to $20°C$. The simulation of the ETP models for this set-point adjustments and the EMM approximation are also shown in sold black and red lines respectively. The effects of the fast thermal discharge and thermal charge are obvious at $t = 11\,hr$ and at $t = 16\,hr$.

The blue curve in Fig. 7.2 shows the aggregated power after implementing the MPC control actions (shown in the bottom graph). We can see that in order to curb down the fast thermal discharge at $t = 11\,hr$, the MPC implement the sequential set-point control for 10,943 devices. While to avoid the fast thermal energy charge, the sequential set-point control is implemented for 11,908 devices.

During the service time, the reference power is specified as the clusters' power at the new temperature set-point ($22°C$), while during the no dispatch time, the reference power is specified as the clusters' power at the nominal temperature set-point ($20°C$). In this example, the load reduction ancillary service is mainly achieved by extracting some of the stored thermal energy. Clearly, we can see that the amount of the ON/OFF switching actions are minimal and converges to zero during and after the service time.

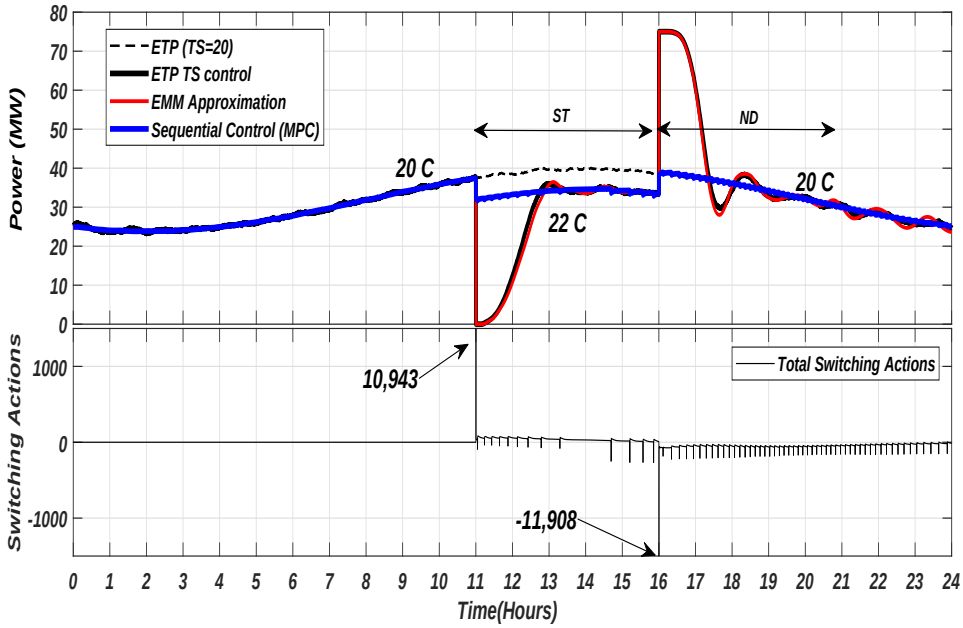

Figure 7.2 load reduction demand response. Top: Aggregated Power. Bottom: total control actions perfomred by the MPC

7.4.1 Demand Response Program for Load Reduction

In the load reduction program, load aggregators are requested to maintain the ancillary service (DR^{Lr}) for consecutive time intervals (ST). The minimum and maximum load reduction capabilities, $\Delta LR^{Min}, \Delta LR^{Max}$ respectively, are specified by the aggregators based on the minimum and maximum set-point adjustments allowed for control. The capability limits are imposed in (7.10). The model defines in (7.11) a dispatch indicator $S^{Lr}_{(i,t)}$ to identify the dispatch starting time (binary variable will be set to one once the aggregator i is requested to provide load reduction). This indicator is used in (7.12) to insure that the commitment is performed for the entire service time (ST). The constraint defined in (7.13) forces the no dispatch time (ND) similar to constraint (7.6) in the spinning reserves model. The load reduction flag F^{Lr} is set in constraint (7.14) to indicate that the aggregator is committed for load reduction and not available during both the service time

and the no dispatch time. Finally, the model in (7.15) verifies that the aggregator is not committed for other demand response services.

$$LR^{Min}_{(i,t)} X^{Lr}_{(i,t)} \leq DR^{Lr}_{(i,t)} \leq LR^{Max}_{(i,t)} X^{Lr}_{(i,t)} \tag{7.10}$$

$$S^{Lr}_{(i,t)} \geq X^{Lr}_{(i,t)} - X^{Lr}_{(i,t-1)} \tag{7.11}$$

$$X^{Lr}_{(i,t)} \geq \sum_{t-ST_i+1}^{t-1} S^{Lr}_{(i,tx)} \tag{7.12}$$

$$1 - X^{Lr}_{(i,t)} \geq \sum_{t-ST_i-ND_i+1}^{t-ST_i} S^{Lr}_{(i,tx)} \tag{7.13}$$

$$\sum_{t}^{t+ND_i} F^{Lr}_{(i,tx)} \geq (ND_i + 1) X^{Lr}_{(i,t)} \tag{7.14}$$

$$X^{Lr}_{(i,t)} \leq 1 - F^{Res/Sh}_{(i,t)} \tag{7.15}$$

$$X^{Lr}, S^{Lr} \in \{0, 1\} \tag{7.16}$$

7.5 Load-Shifting Ancillary Services

Load shifting program allows the aggregators to pre-cool the houses at a lower temperature set-point during the off-peak hours. Charging the houses with more thermal energy gives the potential to release this energy during the peak hours, such that, a load reduction can be achieved. The difference between load shifting and load reduction programs is that, in load shifting, houses are pre-cooled then returned back to the preference set-point, while in the load reduction program, houses are heated during the service time (relative to the preference set-point) then returned back to the preference value after the service.

An example of load shifting is demonstrated in Fig. 7.3. All devices are instructed to operate at $18°C$ before the peak hours. The order of storing additional thermal energy should not be instructed at the same time instant. Otherwise, the aggregated power will experience a sharp increase (See the ETP with set control and the EMM approximation). Therefore, the MPC determines the optimal sequential set-point control such that the aggregated power will move smoothly to the new steady state value. In this case, the sequential set-point control is applied for $13,775$ devices. During pre-cooling, the reference signal is specified as the clusters' steady state power at $18°C$.

During the peak hours, devices are instructed to release the stored thermal energy to provide sustained load reduction. The sequential set-point control is needed to avoid the fast thermal energy discharge. In this case, the sequential set-point control is assigned for $10,943$ devices. It is important here to provide a reference signal which can guarantee achieving a load reduction. For instance, if we provide the clusters' aggregated power at $20°C$ as the reference signal, then the aggregated power will go back to the nominal value without net load reduction. Therefore, the clusters' aggregated power at $22°C$ is used here as a reference signal.

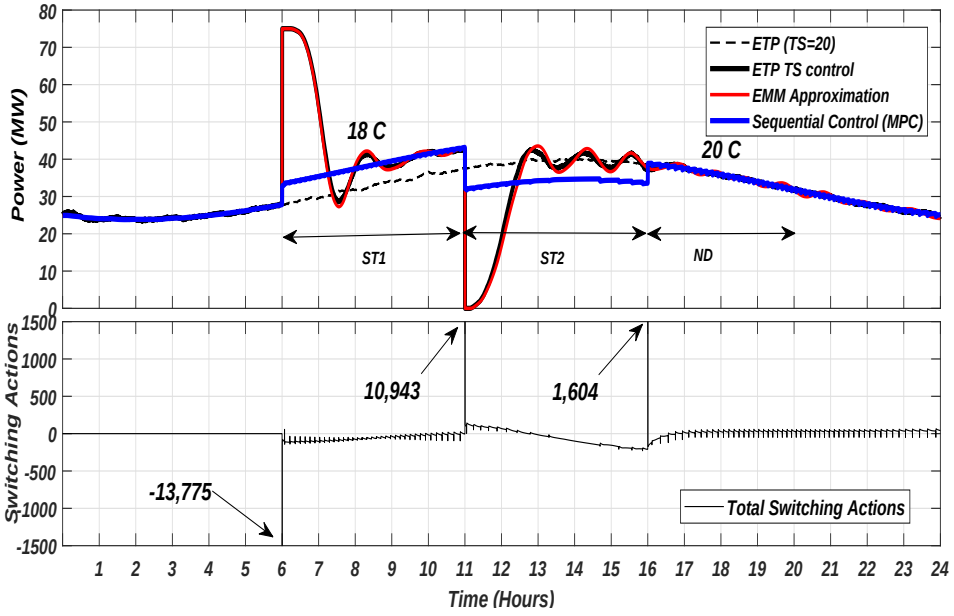

Figure 7.3 load shifting demand response. Top: Aggregated Power. Bottom: total control action perfomred by the MPC

7.5.1 Demand Response Program for Load-Shifting

To model the load shifting capability, load aggregators are required to specify several parameters. Those parameters are determined based on the flexibility in adjusting the customers' set-points and their comforts boundaries. *i.e.,* the minimum set-point adjustment specifies the load increase capability (Max^{Up}), and the maximum set-point value determines the load decrease capability (Max^{Dn}). The limits are demonstrated in Eq.(7.17). The constraint impose the limits for both, the load increase variable $DR^{Up}_{(i,t)}$, and the load decrease variable $DR^{Dn}_{(i,t)}$.

The indicators ($S^{Up/Dn}$) are defined in (7.18) for the load increase and the load decrease to specify the starting time for both cases. The indecators are then used in (7.19) to force the minimum service times (ST_1, ST_2) as shown in Fig.7.3. The no dispatch constraint is defined in (7.20) to verify that all devices are charging back to the normal operating conditions. The model in (7.21) prevents giving discharging orders while devices are not previously charged. While the constraint defined in (7.22) verifies that there will be no charging and discharging orders at the same time instant. Finally, the constraint in (7.23) sets the load shifting flag, and (7.24) verifies the other demand response flags.

$$Min^{Up/Dn}_{(i,t)} X^{Up/Dn}_{(i,t)} \leq DR^{Up/Dn}_{(i,t)} \leq Max^{Up/Dn}_{(i,t)} X^{Up/Dn}_{(i,t)} \tag{7.17}$$

$$S^{Up/Dn}_{(i,t)} \geq X^{Up/Dn}_{(i,t)} - X^{Up/Dn}_{(i,t-1)} \tag{7.18}$$

$$X^{Up/Dn}_{(i,t)} \geq \sum_{t-ST_{1/2}+1}^{t-1} S^{Up/Dn}_{(i,tx)} \tag{7.19}$$

$$1 - X^{Dn}_{(i,t)} \geq \sum_{t-ST_2-ND_i+1}^{t-ST_2} S^{Dn}_{(i,tx)} \tag{7.20}$$

$$X^{Dn}_{(i,t)} \leq \sum_{t-ST_2+1}^{t-1} X^{Up}_{(i,tx)} \tag{7.21}$$

$$1 - X^{Dn}_{(i,t)} = X^{Up}_{(i,t)} \tag{7.22}$$

$$\sum_{t}^{t+ND_i} F^{Up/Dn}_{(i,tx)} \geq (1 + ND_i)(X^{Up}_{(i,t)} + X^{Dn}_{(i,t)}) \tag{7.23}$$

$$X^{Up/Dn}_{(i,t)} \leq 1 - F^{Res/Pc}_{(i,t)} \tag{7.24}$$

$$X^{Up}, S^{Up}, X^{Dn}, S^{Dn} \in \{0, 1\} \tag{7.25}$$

7.6 Security Constraint Unit Commitment Co-Optimization Problem

The Unit Commitment (UC) problem over a specified planning horizon is usually conducted by system operators to determine the optimal dispatch of the thermal generation units. Conventional units are dispatched to satisfy both the energy and the reserve requirements with minimum cost. Problem formulation constitutes a Mixed Integer Linear Programming (MILP) problem. Our objective in this analysis is to integrate the three demand response programs with the unit commitment problem and regard the flexibility of the TCLs as virtual power plants. Dispatching the TCLs to provide spinning reserves, load reduction, and load shifting is expected to provide valuable economical and environmental benefits. Demand response resources can displace the expensive peaking and thus reducing the Market clearing price. Moreover, shifting loads or providing load reduction services can minimize the system ramping requirements and increase the opportunity to serve the load with more economic base-case generations.

The main focus of this chapter is the TCLs modeling and their integration with UC problem. These models will facilitate the dispatching process based on the true TCLs characteristics in the charging and discharging capabilities once their operating set-point is modified. This chapter considers the conventional UC problem which includes only the costs of the thermal power plants with no demand bidding [67]-[70].

7.6.1 Objective Function

The conventional objective function 7.26 of the UC problem is to minimize the costs associated with the thermal units, these costs are mainly; the energy production cost C_i^g, the fixed operational cost C_i^f, the start up cost C_i^u, the shutdown cost C_i^d, and the spinning reserve cost C_i^s. The problem involves binary variables. (*i.e.,* $X_{(i,t)}$ as a commitment variable, $Y_{(i,t)}$ as a start up variable, and Z as a shutdown variable), and continuous variables. (*i.e.,* $Pg_{(i,t)}$ indicates the generation output level, and $Pg^s_{(i,t)}$ represents the amount of the spinning reserves). The index i is used for the thermal units, while the index t is used for the planning horizon.

$$Min\ \ J = \sum_{i}^{T} \left\{ \sum_{i}^{G} \left\{ C_i^g Pg_{(i,t)} + C_i^f X_{(i,t)} + C_i^u Y_{(i,t)} + C_i^d Z_{(i,t)} + C^s_{(i} Pg^s_{(i,t)} \right\} \right. \tag{7.26}$$

7.6.2 Thermal Generation Units Constraints

A set of constraints governs the performance of the conventional units. These constraints are imposed to respect the generation physical capabilities in providing the energy or the spinning reserves in the day-ahead market. The interpretation of the constraints is represented as follows; the minimum and maximum generation limits (Pg_i^{min} and Pg_i^{max}) are imposed in (7.27). The start up Y, and the shutdown Z binary variables are defined in (7.28) and (7.29) respectively. These variables are defined to indicate the time interval the unit turned ON or OFF and account for the associated costs.

The minimum up time (UT) represents the time required for the unit to stay ON once committed and is defined in (7.30). While the minimum down time (DT) indicates the time required for the units to stay OFF once decommitted and is imposed in (7.31). The ramping up (R^U) and ramping down (R^D) capabilities are defined in (7.32) and (7.33) respectively. In order to provide spinning reserves, the units must be committed to produce an amount of power less than the maximum

generation limit (7.34); this amount can be utilized as the spinning reserves if the generator has enough ramping capability (7.35).

$$Pg_i^{min} X_{(i,t)} \leq Pg_{(i,t)} \leq Pg_i^{max} X_{(i,t)} \tag{7.27}$$

$$Y_{(i,t)} \geq X_{(i,t)} - X_{(i,t-1)} \tag{7.28}$$

$$Z_{(i,t)} \geq X_{(i,t-1)} - X_{(i,t)} \tag{7.29}$$

$$X_{(i,t)} \geq \sum_{t-UT_{(i)}-1}^{t-1} Y_{(i,t)} \tag{7.30}$$

$$1 - X_{(i,t)} \geq \sum_{t-DT_{(i)}-1}^{t-1} Z_{(i,t)} \tag{7.31}$$

$$Pg_{(i,t)} - Pg_{(i,t-1)} \leq X_{(i,t-1)} R_{(i)}^U + \big(1 - X_{(i,t-1)}\big) Pg_i^{min} \tag{7.32}$$

$$Pg_{(i,t-1)} - Pg_{(i,t)} \leq X_{(i,t)} R_{(i)}^D + \big(1 - X_{(i,t)}\big) Pg_i^{min} \tag{7.33}$$

$$Pg_{g(i,t)} + Pg_{(i,t)}^s \leq Pg_{(i)}^{max} X_{(i,t)} \tag{7.34}$$

$$Pg_{(i,t)}^s \leq (10/60)\ R_i^U \tag{7.35}$$

7.6.3 System-level Constraints

7.6.3.1 Power Balance

The power balance constraint (7.36) need to be satisfied at each time interval in the planning horizon. The constraint guarantees that there are enough resources committed to fulfill the net-load

requirement. Net-load is defined as the difference between the expected load D^{ex} and the expected renewable generation. This analysis is limited to consider only wind generation at the bulk power system side P_w. Power resources constitute the power supplies from all committed thermal units Pg, and the power decrements or increments achieved by the DR participation. The DR^{Lr} of the load reduction program and the DR^{Dn} of the load shifting program are added to the generation side since they participate in decreasing the load. However, the DR^{Up} of the load shifting program is augmented to the load (D^{ex}) side since it provides load increase (see (7.36)). N_b refers to the number of load buses, while N_w is the number of wind farms.

$$\sum_{i}^{N_g} Pg_{(i,t)} + \sum_{i}^{N_b} DR^{Lr}_{(i,t)} + \sum_{i}^{N_b} DR^{Dn}_{(i,t)} = \sum_{i}^{N_b} D^{ex}_{(i,t)} - \sum_{i}^{N_w} P_{w(i,t)} + \sum_{i}^{N_b} DR^{Up}_{(i,t)} \tag{7.36}$$

7.6.3.2 Reserves Requirement

Spinning reserves represent the online capacity synchronized to the power system and ready to be utilized within ten minutes following dispatch orders. Such reserves are needed to maintain the power system frequency within stability limits in case of emergencies (generation trip). The overall contribution from all units must be at least equal to the system SR requirement (7.37). The amount of spinning reserves varies from one utility to another. Usually, it's proportional to the output power of the largest generating unit committed online. Such that, if the largest unit experiences an outage, the system can recover safely without being forced to load interruptions. Costs associated with sniping reserves are based on the opportunity cost lost for not participating in the energy market. In this model, the reserves requirement constraint is augmented by the reserves provided by demand response participation DR^{Res}. Where, Res is the system reserves requirement.

$$\sum_{i}^{N_g} Pg^{s}_{(i,t)} + \sum_{i}^{N_b} DR^{Res}_{(i,t)} \geq Res, \tag{7.37}$$

7.6.3.3 Transmission Lines Thermal Limits and Security Constraints

Under normal operating conditions, the power flow limit for each transmission line F_ℓ is preserved using (7.38). Each line-flow limit should not be violated at any given time over the planning horizon. The power transfer distribution factor $PT_{(\ell,b)}$ (assuming random slack distribution among all buses) is used to determine the power line flow based on total bus injections. The map $\Lambda_{(i,j)}$ is used to location of thermal power plants and wind generation on the buses.*i.e.,* $\Lambda_{(i,j)} = 1$, if unit i is located on bus j.

$$-F_\ell \leq \sum_{i}^{N_b} PT_{(\ell,i)} \left\{ \sum_{j}^{N_G} P_{g(j,t)} \Lambda_{g(i,j)} + \sum_{j}^{N_w} P_{w(j,t)} \Lambda_{w(i,j)} + DR^{Lr}_{(i,t)} + DR^{Dn}_{(i,t)} - DR^{Up}_{(i,t)} - D^{ex}_{(i,t)} \right\} \leq F_\ell \qquad \forall \ell \in N_L, \forall t \in T \quad (7.38)$$

The security constraint is demonstrated in (7.39); under any given line contingency, the thermal limit of the lines should not be violated at any given time over the planning horizon. The line outage distribution factor $LO_{(\ell,\ell)}$ is used to find the relative change in each line flow after any given contingency c.

$$-F_\ell \leq \sum_{i}^{N_b} \{PT_{(\ell,i)} + LO_{(\ell,c)} PT_{(c,i)}\} \left\{ \sum_{j}^{N_G} P_{g(j,t)} \Lambda_{g(i,j)} + \sum_{j}^{N_w} P_{w(j,t)} \Lambda_{w(i,j)} + DR^{Lr}_{(i,t)} + DR^{Dn}_{(i,t)} - DR^{Up}_{(i,t)} - D^{ex}_{(i,t)} \right\} \leq F_\ell \qquad \forall c \in N_L, \forall \ell \in N_L, \forall t \in T \quad (7.39)$$

7.7 Case Study

The proposed DR models are demonstrated using IEEE Reliability test system (RTS) (see Fig. 7.4), information regards system topology, thermal units incremental heat rates, fuel price, startup and shutdown costs, cycling restriction, and ramping rates are presented in [66]. Tuesday of week 51 is the selected simulation day with system peak-load of 2850 MW, distribution of load among load buses and load temporal characteristic are also presented in [66]. Hydro power generation are assumed committed for only the spinning reserves adequacy with a total capacity

of 25 MW. Wind generation with relatively high penetration level is assumed for the simulated day. Wind farms provides 25% of the total system load. The inverse peaking characteristic of wind production shown in Fig.7.5 adversely affect the load curve by producing a net-load curve with much higher ramping requirement than originally required by the load curve. Spinning reserve is assumed to be constant over the planning horizon hours with a value of 375 MW.

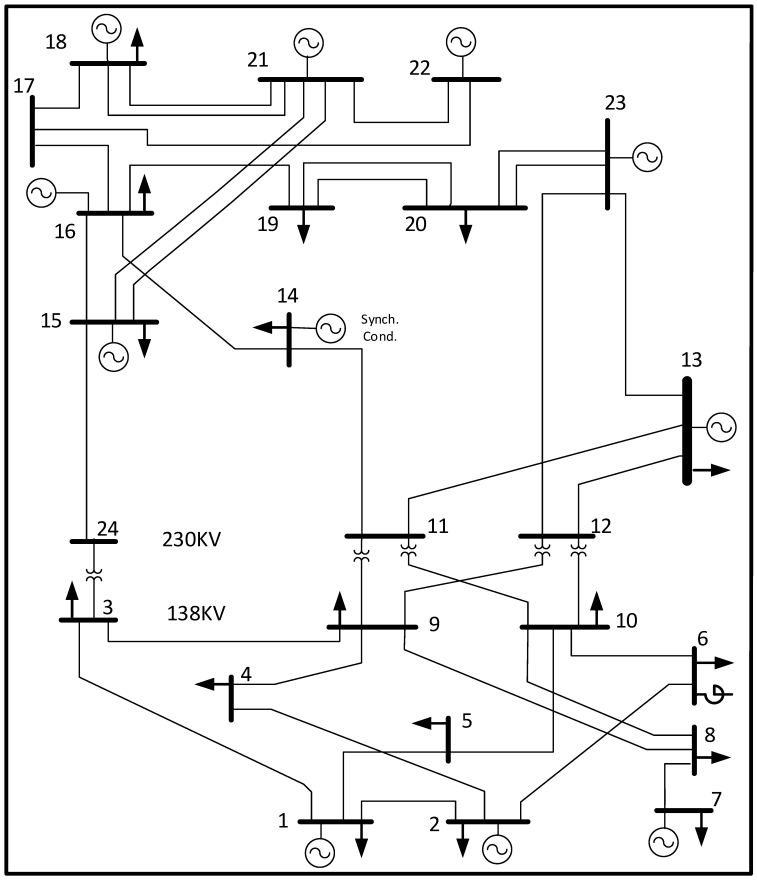

Figure 7.4 IEEE-RTS 24-Bus system.

7.8 Base-case scenario

Base-Case scenario provides the solution of the Security constraint unit commitment problem without DR participation. All of the committed generators are shown in Fig. 7.6. Most of the units are de-committed in the morning hours except for the two nuclear power plants at bus 18 and one

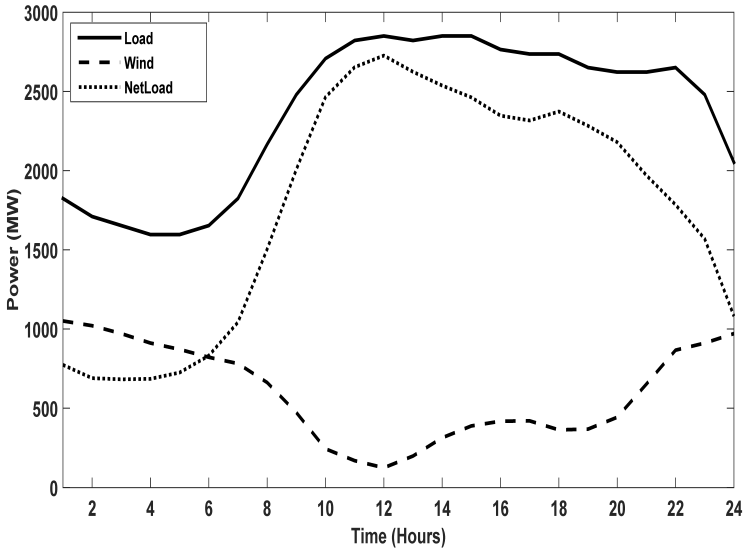

Figure 7.5 System load, wind generation, and net-load

coal power plant at bus 23. During the noon and afternoon hours, mostly all of the power plants are committed in addition to the expensive ones (Combustion turbines) at bus 1 and bus 2 in order to supply both energy and reserves. The same units are also committed to provide the spinning reserves requirements of 375 MW at each hour in the planning horizon (24 hour). The commitment of the expensive units causes the market clearing price to jump to a high value as shown in Fig. 7.8.

7.8.1 Assumptions of TCLs Participation

The analysis conducted in the Sections (7.3 - 7.5) assume 25,000 air-conditioning load. The average of the aggregated power without control is 31.851 MW. In this section, it is assumed that there is a TCLs aggregator at each load bus.*i.e.,* 17 aggregators. The number of devices and the associated power consumption varies according to the size of the load bus. Three participation scenarios are considered as shown in Table. 7.1. It is assumed that the average amount of flexible loads (average TCLs' power without control) in the three scenarios is 30%,35%, and 40% of the load at that bus. The resulting amounts of flexible power are compared with the reference aggregator

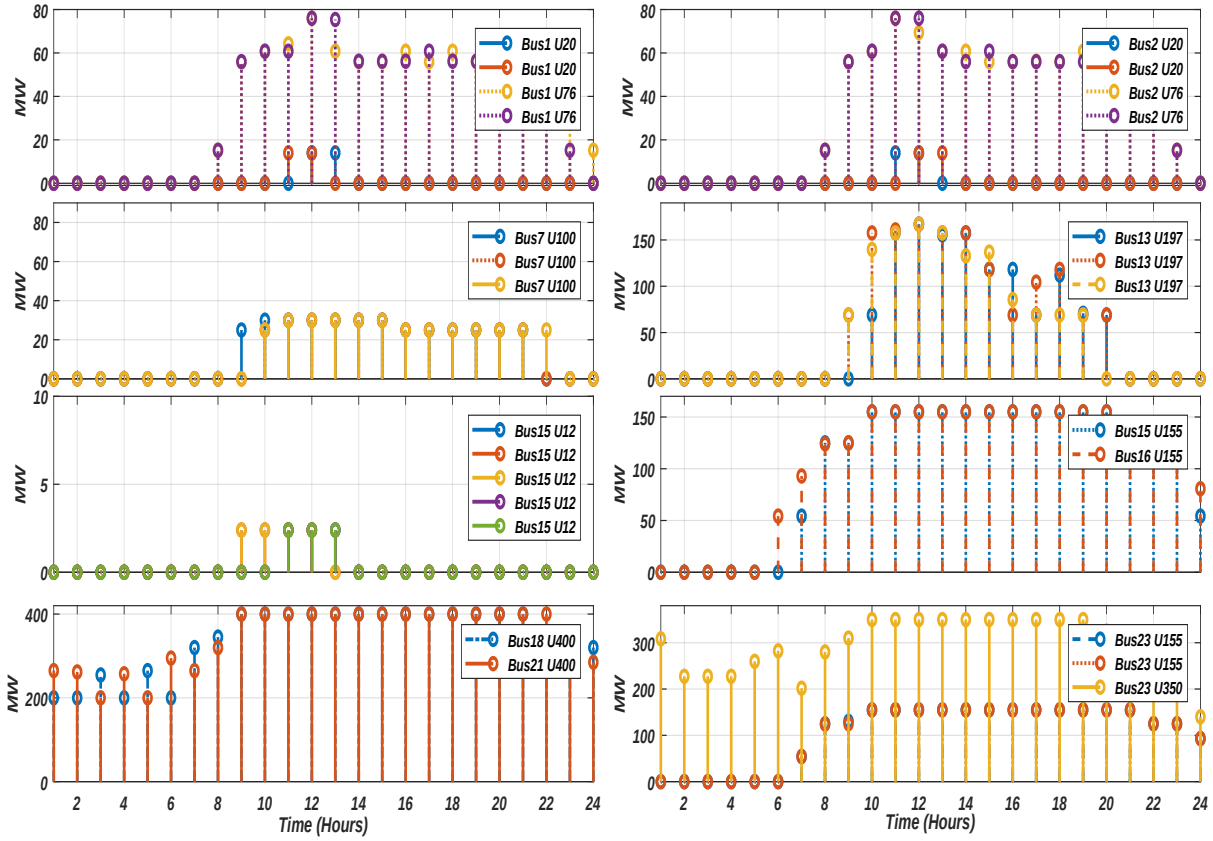

Figure 7.6 Thermal units commitment for energy adequacy (Base-Case scenario)

size (31.851 MW average power with 25,000 device) and the number of devices under control are assigned accordingly. The details are shown in Table. 7.1.

7.8.2 Demand Response Participation in Spinning Reserves Program

This section provides the results of the unit commitment problem when all DR aggregatros are willing to supply and participate in providing the systems' spinning reserves. It is assumed that all aggregators are willing to adjust their aggregated power as previously shown in Fig. 7.1. A No dispatch time of 3-hours is imposed for all aggregatros once they are committed at any specific hour. This time is required to recover the set-point back to the preference value and maintain the customers' comfort level. In this analysis, it is assumed that there are no costs associated with the

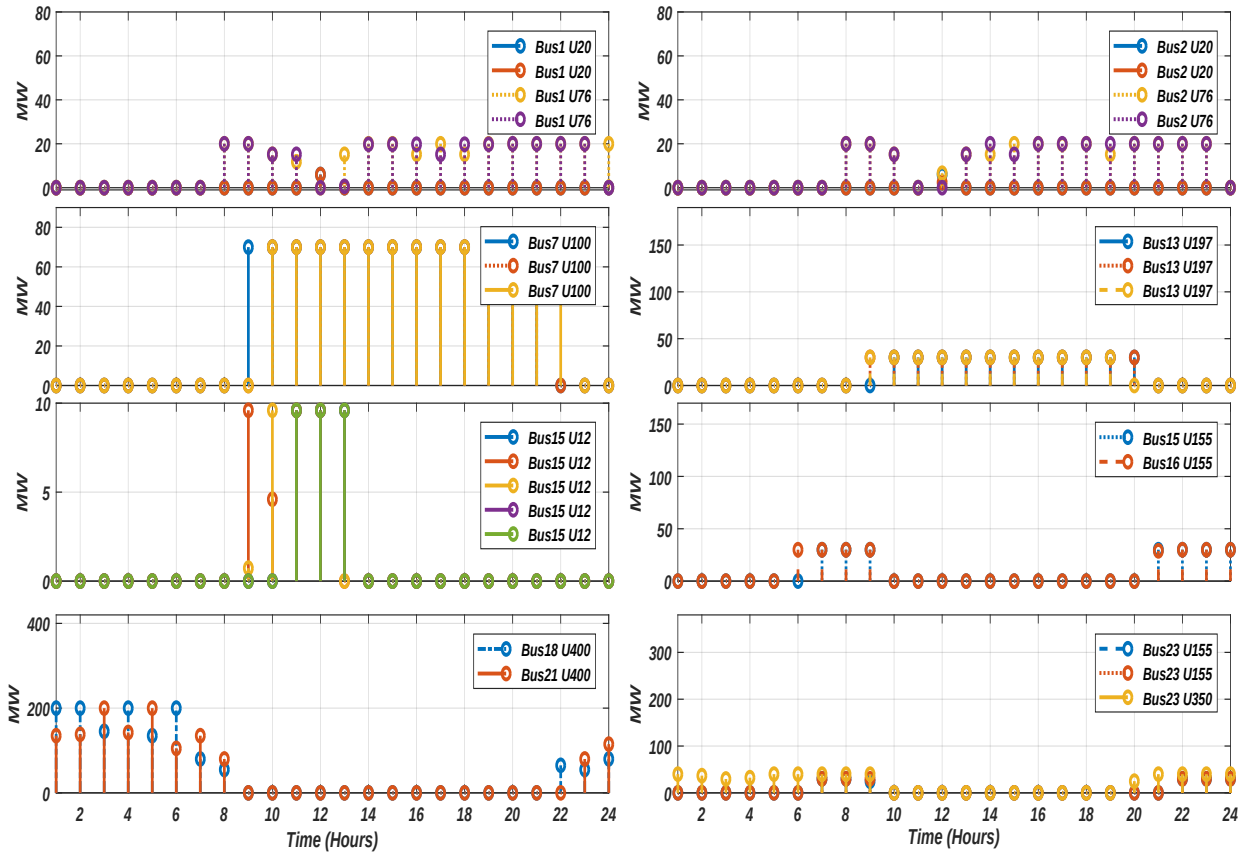

Figure 7.7 Thermal units commitment for reserves adequacy (Base-Case scenario)

DR participation. The amount of cost reductions can be used as a basis to estimate the amount of customers' incentives in participating in such DR programs.

A comparison between the three DR participation scenarios is shown in Fig. 7.9. In the base case, only thermal power plants are committed to provide the spinning reserves requirements. Therefore, expensive units are committed especially during the peak hours to provide the reserves. Under DR participation, all peaking units are de-committed. Thus, the market clearing price has more flat profiles. The over all cost to supply both energy and reserves has been reduced from \$617,974.462 in the base case scenario to \$525,674.3571 for scenario 1, and to \$510,855.2468 for scenario 2, and to \$497,005.558 for scenario 3. A large cost reduction can be seen between the base case and scenario 1 due to decommitting all of the expensive units. As an example, Table. 7.2

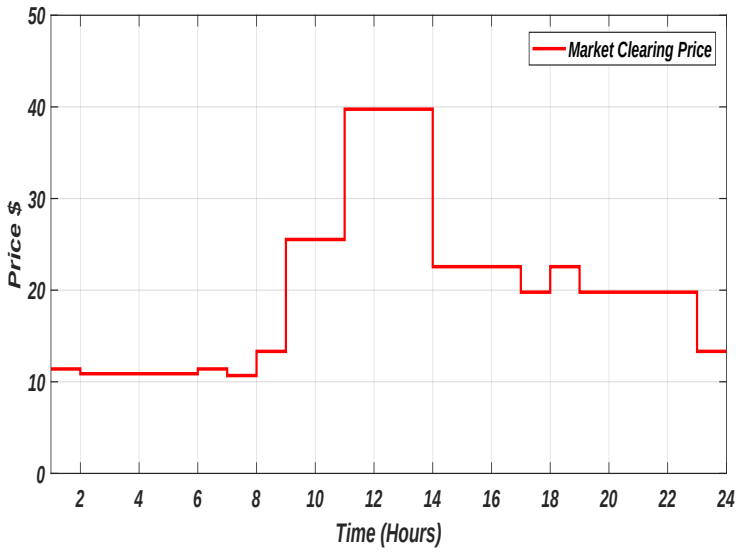

Figure 7.8 Market Clearing Price (Base-Case scenario)

provides the details of the dispatching results for all DR aggregators under scenario 1. As we can see, each aggregator is given at least 3-hours to enable smoothly charging the devices without violations (overshooting or oscillations). The aggregators are committed and arranged to provide the maximum market benefits with minimum possible cost.

7.8.3 Demand Response Participation in the Load Reduction Program

In this section, all DR aggregators are participating in the load reduction program. Similar to the spinning reserves program, it is assumed that the aggregators are offering the load reduction services with no associated costs. The obtained cost reductions can be used to estimate the amount of rewards and other incentives the aggregatores deserve for providing such ancillary service. Aggregators are expected to provide the load reduction as previously described in Fig. 7.2, the load reduction is offered with specific service-time. *i.e.,* the amount of time the customers are welling to control their set-point, and the recovery time or the no dispatch-time which indicates the time required for the aggregators to bring the customers back to the preference set-point (charging time). In this test case, its assumed that both the service-time and the no dispatch-time are 3-hours.

Table 7.1 TCLs Participation Scenarios

Bus	AVG Load	%			AVG Flex MW			Relative Size			# of TCLs		
		S1	S2	S3	S1	S2	S3	S1	S2	S3	S1	S2	S3
1	89.80	30	35	40	26.9	31.4	35.9	0.85	0.99	1.13	21,145	24,669	28,193
2	80.35	30	35	40	24.1	28.1	32.1	0.76	0.88	1.01	18,919	22,072	25,226
3	148.88	30	35	40	44.7	52.1	59.6	1.40	1.64	1.87	35,056	40,899	46,742
4	61.44	30	35	40	18.4	21.5	24.6	0.58	0.68	0.77	14,468	16,879	19,290
5	59.08	30	35	40	17.7	20.7	23.6	0.56	0.65	0.74	13,911	16,230	18,548
6	113.43	30	35	40	34.0	39.7	45.4	1.07	1.25	1.42	26,710	31,161	35,613
7	103.98	30	35	40	31.2	36.4	41.6	0.98	1.14	1.31	24,484	28,564	32,645
8	141.79	30	35	40	42.5	49.6	56.7	1.34	1.56	1.78	33,387	38,951	44,516
9	144.15	30	35	40	43.2	50.5	57.7	1.36	1.58	1.81	33,943	39,601	45,258
10	160.69	30	35	40	48.2	56.2	64.3	1.51	1.77	2.02	37,838	44,145	50,451
11	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	0
13	219.77	30	35	40	65.9	76.9	87.9	2.07	2.41	2.76	51,750	60,375	69,000
14	160.69	30	35	40	48.2	56.2	64.3	1.51	1.77	2.02	37,838	44,145	50,451
15	262.31	30	35	40	78.7	91.8	104.9	2.47	2.88	3.29	61,766	72,060	82,354
16	82.71	30	35	40	24.8	28.9	33.1	0.78	0.91	1.04	19,476	22,722	25,968
17	0	0	0	0	0	0	0	0	0	0	0	0	0
18	276.49	30	35	40	82.9	96.8	110.6	2.60	3.04	3.47	65,104	75,955	86,806
19	151.24	30	35	40	45.4	52.9	60.5	1.42	1.66	1.90	35,613	41,548	47,484
20	106.34	30	35	40	31.9	37.2	42.5	1.00	1.17	1.34	25,040	29,214	33,387
21	0	0	0	0	0	0	0	0	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0	0	0	0	0	0
24	0	0	0	0	0	0	0	0	0	0	0	0	0

The maximum amount of load reduction that can be achieved by the aggregators is the difference between the aggregated power when all devices are regulated at the preference set-point ($20°C$) and the aggregated power when the set-point is adjusted to $22°C$ (based on the assumption that the maximum set-point adjustment is limited to $+2°C$). The eigenvalue analysis is used to obtain the aggregated power value at both set-points.

Fig. 7.10 demonstrates the results of the load reduction program. The effects on the net-load shape can be seen in the right plot, while the changes in the market clearing price is shown in the left graph. Participating in such DR programs can provide an overall energy efficiency from

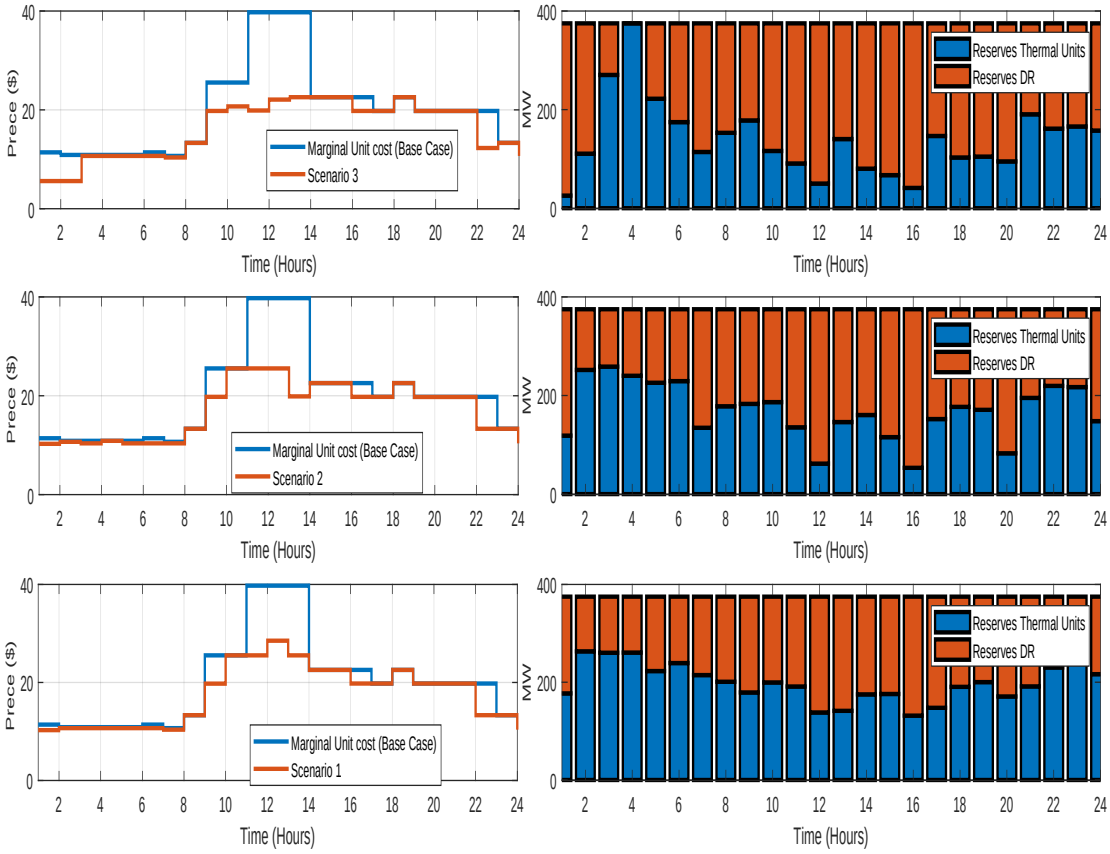

Figure 7.9 Aggregators Participating in the spinning reserves requirement.(Bottom: Scenario 1). (Middle: Scenario 2). (Top: Scenario 3).

the net-load point of view, and as a tool to avoid committing the expensive units from the market clearing price point of view.

Expensive units are also decommitted in this case due to the load reduction offered during the peak hours. Since there is no restriction on the commitment time, the dispatch also utilize the aggregators during the low load periods *i.e,* night hours, then the aggregators are allowed to charge (bring the set-point back to $20°C$) in order to be able to provide the load reduction during the peak hours. The objective function is reduced from being \$617,974.462 in the base case scenario to \$592,435.0677 in scenario 1, and to \$588,803.3014 in scenario 2, and to \$585,031.4221 in scenario 3. As an example, Table. 7.3 provides the details of the aggregators dispatch under scenario 1.

Table 7.2 Aggregators participation in spinning reserves (scenario 1)

time	*Agg 1*	*Agg 2*	*Agg 3*	*Agg 4*	*Agg 5*	*Agg 6*	*Agg 7*	*Agg 8*	*Agg 9*	*Agg 10*	*Agg 11*	*Agg 12*	*Agg 13*	*Agg 14*	*Agg 15*	*Agg 16*	*Agg 17*	*Total*
1	0	0	34.28149	14.20233	0	0	0	32.81228	0	0	0	36.97503	60.48234	19.09968	0	0	0	**197.8531**
2	0	18.19795	0	0	0	0	23.46578	0	0	36.15646	0	0	0	0	0	34.00144	0	**111.8216**
3	19.15633	0	0	0	13.40943	0	0	0	32.56575	0	49.56699	0	0	0	0	0	0	**114.6985**
4	0	0	0	0	0	26.20321	0	0	0	0	0	0	0	0	63.67135	0	24.48898	**114.3635**
5	0	0	35.74205	14.80742	0	0	0	0	0	0	0	38.55035	63.05918	0	0	0	0	**152.159**
6	0	20.51892	0	0	0	0	0	36.1781	0	40.76785	0	0	0	0	0	38.33798	0	**135.8029**
7	23.0362	0	0	0	16.12534	0	0	0	39.16155	0	59.60618	0	0	22.4603	0	0	0	**160.3896**
8	0	0	0	0	0	32.95752	30.1854	0	0	0	0	0	0	0	80.0837	0	30.80142	**174.028**
9	0	0	46.02748	19.06853	0	0	0	0	0	0	0	49.64392	81.20562	0	0	0	0	**195.9455**
10	0	26.51142	0	0	0	0	0	46.74383	0	52.67401	0	0	0	0	0	49.5345	0	**175.4638**
11	29.34183	0	0	0	0	0	0	0	49.88111	0	75.92198	0	0	28.60828	0	0	0	**183.7532**
12	0	0	0	0	21.36145	40.81563	37.38254	0	0	0	0	0	0	0	99.17818	0	38.14545	**236.8833**
13	0	0	54.85747	22.72667	0	0	0	0	0	0	0	59.1677	96.78426	0	0	0	0	**233.5361**
14	0	30.19418	0	0	0	0	0	53.23711	0	59.99107	0	0	0	0	0	56.41544	0	**199.8378**
15	31.78276	0	0	0	0	0	0	0	54.03068	0	82.23788	0	0	30.98819	0	0	0	**199.0395**
16	0	0	0	0	21.94175	41.92441	38.39806	0	0	0	0	0	0	0	101.8724	0	39.18169	**243.3183**
17	0	0	53.39847	22.12222	0	0	0	0	0	0	0	57.59407	94.21016	0	0	0	0	***227.3249***
18	0	27.87137	0	0	0	0	0	49.14163	0	55.37601	0	0	0	0	0	52.07546	0	**184.4645**
19	27.90263	0	0	0	0	0	0	0	47.43447	0	72.19805	0	0	27.20506	0	0	0	**174.7402**
20	0	0	0	0	18.40768	35.17182	32.21344	0	0	0	0	0	0	0	85.46423	0	32.87086	**204.128**
21	0	0	43.11374	17.86141	0	0	0	0	0	0	0	46.50125	76.06496	0	0	0	0	**183.5414**
22	0	21.88045	0	0	0	0	0	38.5787	0	43.47301	0	0	0	0	0	40.8819	0	**144.8141**
23	21.59533	0	0	0	0	0	0	0	36.71206	0	55.87792	0	0	21.05545	0	0	0	***135.2408***
24	0	0	0	0	14.29474	27.31317	25.0158	0	0	0	0	0	0	0	66.36845	0	25.52633	***158.5185***

Because of the charging and discharging time restrictions, aggregators are dispatched throughout the day to provide the load reduction service in only four time intervals.

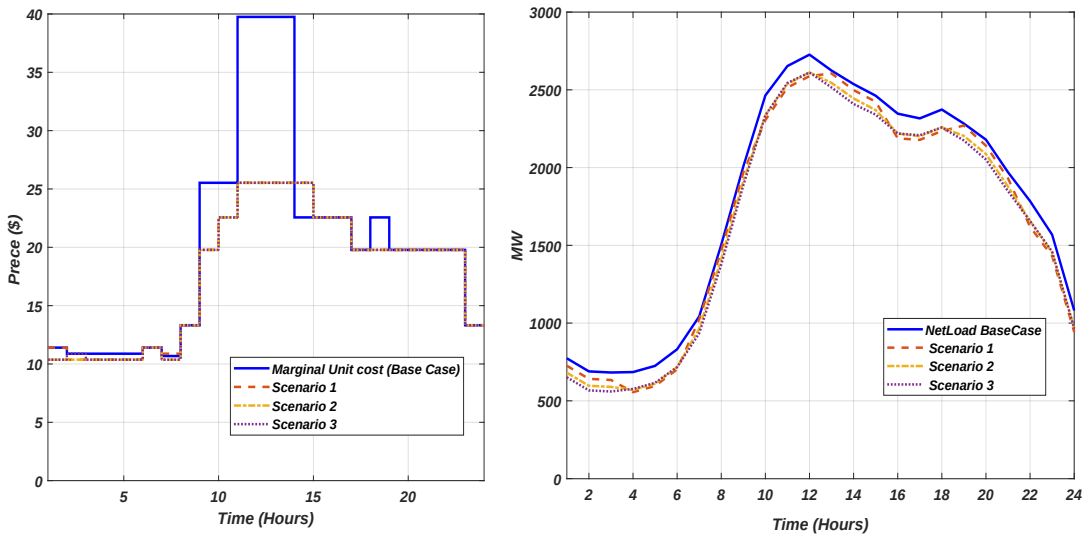

Figure 7.10 Aggregators Participating in the load reduction program.

7.8.4 Demand Response Participation in the Load Shifting Program

Load shifting program is offered by DR aggregators with three main time constraints as previously discussed in Fig. 7.3. First, the charging time which is required to move the operating

Table 7.3 Aggregators participation in the load reduction program (scenario 1)

time	*Agg 1*	*Agg 2*	*Agg 3*	*Agg 4*	*Agg 5*	*Agg 6*	*Agg 7*	*Agg 8*	*Agg 9*	*Agg 10*	*Agg 11*	*Agg 12*	*Agg 13*	*Agg 14*	*Agg 15*	*Agg 16*	*Agg 17*	*Total*
1	0	4.039681	0	3.082915	2.976607	0	0	0	0	0	0	8.026208	0	0	13.81996	0	0	**31.94537**
2	0	4.043853	0	3.086098	2.979681	0	0	0	0	0	0	8.034497	0	0	13.83423	0	0	**31.97836**
3	0	4.043848	0	3.086094	2.979677	0	0	0	0	0	0	8.034487	0	0	13.83422	0	0	**31.97832**
4	4.252284	0	7.441497	0	0	5.68743	5.209048	7.122575	7.228882	8.026186	11.00278	0	13.12893	4.145977	0	7.547804	5.315355	**86.10875**
5	4.254512	0	7.445396	0	0	5.69041	5.211777	7.126308	7.23267	8.030391	11.00855	0	13.13581	4.148149	0	7.551759	5.31814	**86.15387**
6	4.252595	0	7.442041	0	0	5.687846	5.209429	7.123097	7.229412	8.026773	11.00359	0	13.12989	4.14628	0	7.548356	5.315744	**86.11505**
7	0	4.040458	0	0	0	0	0	0	0	0	0	8.027751	0	0	0	0	0	**12.06821**
8	0	4.038693	0	0	0	0	0	0	0	0	0	8.024246	0	0	13.81658	0	0	**25.87952**
9	0	4.040666	0	0	0	0	0	0	0	0	0	8.028166	0	0	13.82333	0	0	**25.89216**
10	4.251516	0	7.440153	3.082349	2.976061	5.686402	5.208107	7.121289	7.227577	8.024736	11.0008	0	13.12656	4.145228	13.81743	7.546441	5.314395	**105.969**
11	4.252815	0	7.442427	3.083291	2.976971	5.688141	5.209699	7.123466	7.229786	8.027189	11.00416	0	13.13057	4.146495	0	7.548747	5.316019	**92.17977**
12	4.250312	0	7.438046	3.081476	2.975218	5.684792	5.206632	7.119273	7.22553	8.022464	10.99768	0	13.12284	4.144054	0	7.544304	5.31289	**92.12551**
13	0	4.037192	0	0	0	0	0	0	0	0	0	8.021263	0	0	0	0	0	**12.05845**
14	0	4.039063	0	0	0	0	0	0	0	0	0	8.02498	0	0	13.81785	0	0	**25.88189**
15	0	4.039056	0	0	0	0	0	0	0	0	0	8.024967	0	0	13.81782	0	0	**25.88185**
16	4.249672	0	7.436925	3.081012	2.97477	5.683936	5.205848	7.1182	7.224442	8.021255	10.99603	0	13.12086	4.14343	13.81143	7.543167	5.312089	**105.9231**
17	4.250329	0	7.438076	3.081489	2.97523	5.684815	5.206653	7.119301	7.225559	8.022496	10.99773	0	13.12289	4.144071	0	7.544334	5.312911	***92.12588***
18	4.25281	0	7.442418	3.083287	2.976967	5.688133	5.209692	7.123457	7.229777	8.027179	11.00415	0	13.13055	4.14649	0	7.548738	5.316013	**92.17966**
19	0	0	0	0	0	0	0	0	0	0	0	8.024744	0	0	0	0	0	**8.024744**
20	0	4.040656	0	0	0	0	0	0	0	0	0	8.028146	0	0	13.8233	0	0	**25.8921**
21	0	4.038712	0	0	0	0	0	0	0	0	0	8.024282	0	0	13.81664	0	0	**25.87964**
22	4.253098	4.040443	7.442921	3.083496	2.977168	5.688518	5.210045	7.123938	7.230266	8.027722	11.00489	0	13.13144	4.14677	13.82257	7.549248	5.316372	**110.0489**
23	4.252607	0	7.442063	3.08314	2.976825	5.687862	5.209444	7.123117	7.229433	8.026797	11.00362	0	13.12993	4.146292	0	7.548378	5.315759	***92.17527***
24	4.254498	0	7.445371	3.084511	2.978148	5.69039	5.21176	7.126283	7.232646	8.030364	11.00851	0	13.13576	4.148135	0	7.551733	5.318122	***92.21623***

temperature set-point form the preference value ($20°C$) to a colder environment *i.e.,* ($18°C$). Such that, devices during this time can store additional thermal energy. Second, the discharge time which is required to move the devices back to ($20°C$). This transition will enable the devices to dissipate the stored thermal energy as a load reduction. The third time is the no dispatch time which is specified by the aggregators to make sure that all devices under control return back to the preference set-point ($20°C$). In this case, both the charging time and the discharging time are specified as 3 hours for all aggreagtors, while a no dispatch time is assigned as one hour. It is also assumed that the load aggregators do not request costs to provide the load shifting services. Cost reduction can be used as a basis to estimate the incentives each aggregator deserve.

The effects of the load shifting programs on the net-load waveform for the three participation scenarios are shown in Fig. 7.11. Load aggregators are dispatched to increase the load (charge) during the low price periods while provide load reduction (discharge) during the high price periods. In general, such participation will help in decreasing the amount of the system' ramp requirement and decommitting the expensive units during the peak hours. The objective function value has been reduced from $617,974.462 in the base case scenario to $607,054.5142 for scenario 1, and to $604,689.8892 for scenario 2, and to $601,341.2109 scenario 3. Table. 7.4 shows the results of the

dispatch for all DR aggregators as in scenario 1. Positive signs indicate the load increase periods while the negative sign refer to the load decrease periods.

Due to the load shifting program restrictions.*i.e.,* devices are required to charge at a colder set-point to be able to provide the load reduction. It is most likely that the aggregators will be dispatched just before the peak-hours for the load increase and provide the the load reduction during the peak-hours as illustrated in Table.7.11.

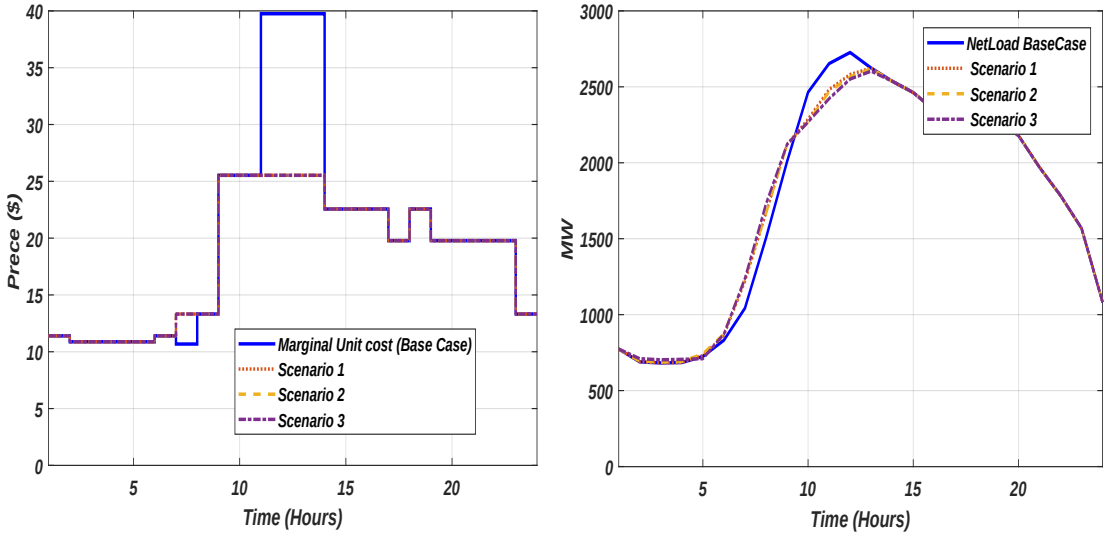

Figure 7.11 Aggregators Participating in the Load shifting program.

Table 7.4 Aggregators participation in the load shifting program (scenario 1)

time	*Agg 1*	*Agg 2*	*Agg 3*	*Agg 4*	*Agg 5*	*Agg 6*	*Agg 7*	*Agg 8*	*Agg 9*	*Agg 10*	*Agg 11*	*Agg 12*	*Agg 13*	*Agg 14*	*Agg 15*	*Agg 16*	*Agg 17*	*Total*
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
5	0	0	0	0	0	0	0	0	0	0	0	0	0	4.148149	0	0	0	**4.148149**
6	0	0	0	3.083131	2.976817	0	0	0	0	0	11.00359	0	0	4.14628	0	0	0	**21.20982**
7	4.253113	4.040458	7.442948	3.083507	2.977179	5.688539	5.210064	7.123965	7.230292	8.027751	11.00493	8.027751	13.13149	4.146785	13.82262	7.549276	5.316392	**118.0771**
8	4.251256	4.038693	7.439698	3.082161	2.975879	5.686055	5.207788	7.120854	7.227135	8.024246	11.00012	8.024246	13.12575	-4.14497	13.81658	7.545979	5.31407	**109.7355**
9	4.253333	4.040666	7.443333	-3.08367	-2.97733	5.688833	5.210333	7.124333	7.230666	8.028166	-11.0055	8.028166	13.13217	-4.147	13.82333	7.549666	5.316666	**75.65616**
10	-4.25152	-4.03894	-7.44015	-3.08235	-2.97606	-5.6864	-5.20811	-7.12129	-7.22758	-8.02474	-11.0008	-8.02474	-13.1266	-4.14523	-13.8174	-7.54644	-5.31439	**-118.033**
11	-4.25282	-4.04017	-7.44243	-3.08329	-2.97697	-5.68814	-5.2097	-7.12347	-7.22979	-8.02719	-11.0042	-8.02719	-13.1306	0	-13.8216	-7.54875	-5.31602	**-113.922**
12	-4.25031	-4.0378	-7.43805	0	0	-5.68479	-5.20663	-7.11927	-7.22553	-8.02246	0	-8.02246	-13.1228	0	-13.8135	-7.5443	-5.31289	**-96.8009**
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	***0***
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	***0***
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	***0***

7.9 Conclusion

The capability of the TCLs has been evaluated in the day-ahead market for providing three main demand response ancillary services; Namely, spinning reserves requirements, load reduction, and load shifting. Appropriate models are designed for each demand response program and integrated to the security constraint unit commitment problem for dispatching purposes. The models and its related constraints are designed based on the performance of the controller (model predictive control) and its capability to modify the aggregated power in each program. It is assumed that the customers' preference temperature set-point is $20°C$, and the assessment is made based on a maximum set-point deviations of $\pm 2°C$.

It has been shown that effective utilization of the TCL resources has great benefits in decommitting the expensive peaking units and thus reducing the market clearing price volatility. It also shown that the maximum benefits (cost reductions) can be obtained when all TCLs are utilized for the system spinning reserves. This is due to the fact that the spinning reserves program discharge the devices in much higher rates than in the load reduction and the load shifting programs. This feature gives the spinning reserves program the advantage since the TCLs can provide the maximum load reduction capability in only one hour then start the charging time for another dispatch order. Therefore, each aggregator is committed multiple times and can provide more energy reductions.

Load shifting program was able to move loads from the peak hours and decommit the peaking units. Thus, the Market clearing price converged to less prices than in the base case scenario. However, the reduction in cost of the objective function was limited to small amounts. The total amount of energy involved in the load shifting program is preserved compared to the load reduction or the reserves programs where the actual energy consumption is reduced. The amount of load is shifted to another period and supplied by less expensive units. Load shifting program shows its advantage in reducing the net-load ramp ramping rates compared to the load reduction program where an overall net-load energy efficiency is achieved.

CHAPTER 8. GENERAL CONCLUSIONS AND FUTURE WORK

The objective of this thesis is to quantify the capability of utilizing the Thermostatically controlled loads (TCLs) to support the power system flexibility in providing three main ancillary services; mainly, load reduction, load-shifting, and the emergency spinning reserves. Such ancillary services can have significant positive impacts on future grid flexibility by maintaining security as renewable penetration increases on both the bulk and distribution sides. Such reinforcement has become a priority for all power systems stakeholders.

The Major contributions of this thesis can be summarized as follows:

- *The Development of the Extended Markov Chain Modeling Approach:*

The work in this thesis provides a novel methodology to aggregate the TCLs. The Extended Markov Model (EMM) has been developed to aggregate the TCLs when the operating temperature set-point is modified. Due to this improvement to the existing Markov chain abstraction method, the TCLs can be seen as gird-scale storage facility. The set-point control is required to partially charge or discharge the accumulated thermal energy and thus modify the aggregated power demand. Large-scale implementations over ultimately millions of devices can provide substantial support to future grid operations while maintaining the customers comfort within acceptable ranges.

The EMM advantages manifested in two main merits. First, the flexibility to describe small and large set-point adjustments in both directions. Second, the fast construction, mainly online, since it is based on combining and restructuring to Markov chains which are already built during the offline stage at fixed set-points. Appropriate linear mapping are established based on set-point change magnitude and direction.

- *Sequential temperature Set-point Control*

Under set-point control, the TCLs aggregate power experience two main challenges before it converges to the new steady-state value, the abrupt load change and the power oscillations. Model Predictive Control (MPC) with direct ON/OFF switching capability is proposed in this thesis to overcome these challenges. This control input is used for two main purposes. First, determine the optimal sequential control law, where appropriate time delay is assigned to devices at the instant of the set-point change. Second, determine the ON/OFF switching actions required to minimize devices' synchronization and to curb possible oscillations. The advantage of this methodology, since it is based on the thermal energy, is that the ON/OFF switching actions are minimal and converge to zero over time. Adopting our approach will therefore not cause additional wear and tear to devices or reduce their life expectancies.

- *Heterogeneous Parameters Estimation and Clustering*

The performance of the EMM is prone to relatively high errors under devices heterogeneity, as buildings and devices have different characteristics. Limited amount of heterogeneity has been reported in literature with relatively narrow Gaussian distributions. The work in this thesis provides a more comprehensive definition to the heterogeneous parameters which covers a wide variety of possible building characteristics and power ratings. A Systematic approach is proposed based on the cycling equations derived from the Equivalent thermal parameter (ETP) model to estimate the new heterogeneous parameters. In order to minimize the EMM error, clustering of building-specific TCLs characteristics and power rating is performed using the K-mean clustering methodology. TCLs are divided into clusters based on heat charging/discharging characteristics and their power ratings. As a result, The EMM performance is highly improved and the error is reduced to almost 2%.

- *Demand Response Programs for Ancillary Services*

In the last part of this thesis, the MPC is used to quantify the capability of the TCLs in providing three main ancillary services; spinning reserves, load reduction, and load shifting. The

EMM is used as a predictive model when the temperature set-point is increased or decreased by $2°C$. Appropriate demand response models are designed for each ancillary service and integrated to the security constraint unit commitment problem. The models are used to investigate the economical benefits that each demand response program can provide to the the day-ahead market. It has been show that effective utilization of the TCLs can avoid committing the expensive peaking units, reduce the market clearing price, provide peak clipping and overall energy efficiency aspects, and finally contributes in shifting portion of the load while reducing the ramp rates associated with the net-load.

Over all, Utilizing the TCLs as demand response resources can provide potential and valuable support to the grid operation and control. Under large scale implementations, the TCLs can be treated as gird-scale storage facility and provide ancillary services with minimal impacts on customers' comfort and devices' integrity. Advanced modeling, monitoring, and control methodologies need to be implemented to safely extract the services. The advancements of smart grid appliances such as the advanced metering infrastructure and the intelligent thermostats can make this possible in our future grid operations.

The future work can be summarized as follows:

- *Quantify the capability of the Water heaters*

The TCLs considered in this thesis are mainly air-conditioning devices. However, to improve the capability of the DR aggregators it is required to exploit all the flexible resources. Adopting the existing Extended Markov Modeling approach and the model predictive control architecture to account for set-point control of water heater system is required. This results in the development of a control coordination between the air conditioners and water heaters models for the various ancillary services.

- *Consider the distribution Network Topology*

The clustering analysis performed in this thesis consider only the power ratings and the charging characteristics. However, another dimension can be added to account for devices location in the

distribution system. This feature will enable the aggregators to select the optimal control locations for both reliability and economical aspects.

- *Co-simulation platform*

investigate the effects of demand side control for both the air-conditioners and water heaters on the network operational parameters.

BIBLIOGRAPHY

[1] Martinot, Eric. "Grid Integration of Renewable Energy: Flexibility, Innovation, and Experience." Annual Review of Environment and Resources 41.1 (2016).

[2] Market and Infrastructure Policy, 2013 flexible capacity procurement requirement, Tech. Rep., March 2012. [Online]. Available: http://www.CAISO.com/.

[3] S. Meyn, M. Negrete-Pincetic, G. Wang, A. Kowli, and E. Shafieepoorfard, The value of volatile resources in electricity markets, in CDC2010, 2010, pp. 10291036, and submitted to IEEE TAC, 2012

[4] Y. Makarov, C. Loutan, J. Ma, and P. de Mello, Operational impacts of wind generation on california power systems, IEEE Transactions on Power Systems, vol. 24, no. 2, pp. 10391050, May 2009.

[5] J. Smith, M. Milligan, E. DeMeo, and B. Parsons, Utility wind integration and operating impact state of the art, IEEE Transactions on Power Systems, vol. 22, no. 3, pp. 900908, August 2007

[6] B. Kirby, Ancillary services: Technical and commercial insights, Tech. Rep., July 2007. [Online].
Available:http://www.consultkirby.com.

[7] Miller, N. W., et al. "Eastern frequency response study." Contract 303 (2013): 275-3000.

[8] Miller, N. W., et al. "California ISO (CAISO) - Frequency Response Study" General Electric International, Report (2011).

[9] Energy Information Administration EIA "International Energy Outlook, 2016 " U.S. Department of Energy, Washington, DC 20585. May 2015

[10] FERC Order 755 " Frequency Regulation Compensation in the Organized Wholesale Power Markets,October,2011.

[11] MISO, Frequency regulation compensation - FERC order no. 755, Tech. Rep., March 2013. [Online]. Available: https: //www.midwestiso.org

[12] National Renewable Energy Laboratory. "Grid integration of aggregated demand response, part 1: load availability profiles and constraints for the western interconnection." September 2013.

[13] QDR, Q. "Benefits of demand response in electricity markets and recommendations for achieving them." US Dept. Energy, Washington, DC, USA, Tech. Rep (2006).

[14] Storage participation in ERCOT (prepared by the Texas energy storage alliance), January 2010. [Online]. Available: http://www.ercot.com/

[15] Energy Information Administration, "Residential energy consumption survey," U.S. Dept. Energy, Washington, DC, Tech. Rep., 2001.

[16] Southern California Edison, Summer Discount Plan demand response program. Available: https://www.sce.com/wps/portal/home/residential/rebates-savings/summer-discount-plan/terms

[17] Commonwealth Edison Company (CoomEd), Peak time saving demand response program. Available: https://www.comed.com/WaysToSave/ForYourHome/Pages/PeakTimeSavings.aspx

[18] Potomac Electric Power Company (Maryland), Smart Grid Project. Available: https: //www.smartgrid.gov/project.html

[19] City of Ames, Efficient Air Conditioner Rebate. Available at: http://www.cityofames.org

[20] PGE, Smart AC program. [Online]. Available: http://www.pge.com

[21] FPL, On call savings program. [Online]. Available: http:// www.fpl.com

[22] Chong, C.Y.; Debs, A.S., "Statistical synthesis of power system functional load models," in Decision and Control including the Symposium on Adaptive Processes, 1979 18th IEEE Conference on , vol.2, no., pp.264-269, 12-14 Dec. 1979.

[23] Ihara, S.; Schweppe, F.C., "Physically Based Modeling of Cold Load Pickup," in Power Apparatus and Systems, IEEE Transactions on , vol.PAS-100, no.9, pp.4142-4150, Sept. 1981.

[24] M. M. Adibi and L. H. Fink, "Power system restoration planning," in IEEE Transactions on Power Systems, vol. 9, no. 1, pp. 22-28, Feb 1994.

[25] Malhame, R.; Chong, Chee-Yee, "Electric load model synthesis by diffusion approximation of a high-order hybrid-state stochastic system," in Automatic Control, IEEE Trans, vol.30, no.9, pp.854-860, Sep 1985.

[26] Malham, Roland, and Chee-Yee Chong. "On the statistical properties of a cyclic diffusion process arising in the modeling of thermostat-controlled electric power system loads." SIAM (1988): 465-480.

[27] Mortensen, R.E.; Haggerty, K.P., "A stochastic computer model for heating and cooling loads," in Power Systems, IEEE Transactions on , vol.3, no.3, pp.1213-1219, Aug. 1988.

[28] Mortensen, R. E., and K. P. Haggerty. "Dynamics of heating and cooling loads: models, simulation, and actual utility data." IEEE Transactions on Power Systems 5.1 (1990): 243-249.

[29] Callaway, Duncan S. "Tapping the energy storage potential in electric loads to deliver load following and regulation, with application to wind energy," Energy Conversion and Management 50.5 (2009): 1389-1400.

[30] Kara, Emre Can, Michaelangelo D. Tabone, Jason S. MacDonald, Duncan S. Callaway, and Sila Kiliccote. "Quantifying flexibility of residential thermostatically controlled loads for demand response: a data-driven approach." In Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings, pp. 140-147. ACM, 2014

[31] Lu, Ning, and David P. Chassin. "A state-queueing model of thermostatically controlled appliances." Power Systems, IEEE Transactions on 19.3 (2004): 1666-1673.

[32] Kundu, Soumya, et al. "Modeling and control of thermostatically controlled loads," arXiv preprint arXiv:1101.2157 (2011).

[33] Koch, Stephan, Johanna L. Mathieu, and Duncan S. Callaway. "Modeling and control of aggregated heterogeneous thermostatically controlled loads for ancillary services." In Proc. PSCC, pp. 1-7. 2011.

[34] J. L. Mathieu and D. S. Callaway, "State Estimation and Control of Heterogeneous Thermostatically Controlled Loads for Load Following," 2012 45th Hawaii International Conference on System Sciences, Maui, HI, 2012, pp. 2002-2011.

[35] J. L. Mathieu, M. Kamgarpour, J. Lygeros and D. S. Callaway, "Energy arbitrage with thermostatically controlled loads," 2013 European Control Conference (ECC), Zurich, 2013, pp. 2519-2526.

[36] J. L. Mathieu, S. Koch and D. S. Callaway, "State Estimation and Control of Electric Loads to Manage Real-Time Energy Imbalance," IEEE Trans. Power Syst., vol. 28, no. 1, pp. 430-440, Feb. 2013.

[37] S. Bashash and H. K. Fathy, "Modeling and control insights into demand-side energy management through setpoint control of thermostatic loads," Proceedings of the 2011 American Control Conference, San Francisco, CA, 2011, pp. 4546-4553.

[38] Bashash, S.; Fathy, H.K., "Modeling and Control of Aggregate Air Conditioning Loads for Robust Renewable Power Management," in Control Systems Technology, IEEE Transactions on , vol.21, no.4, pp.1318-1327, July 2013.

[39] W. Zhang, J. Lian, C. Y. Chang and K. Kalsi, "Aggregated Modeling and Control of Air Conditioning Loads for Demand Response," IEEE Trans. Power Syst., vol. 28, no. 4, pp. 4655-4664, Nov. 2013.

[40] Kamgarpour, M.; Ellen, C.; Soudjani, S.E.Z.; Gerwinn, S.; Mathieu, J.L.; Mullner, N.; Abate, A.; Callaway, D.S.; Franzle, M.; Lygeros, J., "Modeling options for demand side participation of thermostatically controlled loads," in Bulk Power System Dynamics and Control - IX Optimization, Security and Control of the Emerging Power Grid (IREP), 2013 IREP Symposium , vol., no., pp.1-15, 25-30 Aug. 2013.

[41] S. Esmaeil Zadeh Soudjani and A. Abate, "Aggregation of thermostatically controlled loads by formal abstractions," 2013 European Control Conference (ECC), Zurich, 2013, pp. 4232-4237.

[42] Soudjani, Sadegh Esmaeil Zadeh, Sebastian Gerwinn, Christian Ellen, Martin Frnzle, and Alessandro Abate. "Formal synthesis and validation of inhomogeneous thermostatically controlled loads." In International Conference on Quantitative Evaluation of Systems, pp. 57-73. Springer International Publishing, 2014.

[43] S. Esmaeil Soudjani and A. Abate, "Aggregation and Control of Populations of Thermostatically Controlled Loads by Formal Abstractions," IEEE Trans. Control Syst. Technol., vol. 23, no.3, pp.975-990, 2015.

[44] B. M. Sanandaji, H. Hao and K. Poolla, "Fast Regulation Service Provision via Aggregation of Thermostatically Controlled Loads," 2014 47th Hawaii International Conference on System Sciences, Waikoloa, HI, 2014, pp. 2388-2397.

[45] H. Hao, B. M. Sanandaji, K. Poolla and T. L. Vincent, "Aggregate Flexibility of Thermostatically Controlled Loads,"IEEE Trans. Power Syst., vol. 30, no. 1, pp. 189-198, Jan. 2015.

[46] Sanandaji, Borhan M., He Hao, Kameshwar Poolla, and Tyrone L. Vincent. "Improved battery models of an aggregation of thermostatically controlled loads for frequency regulation." In American Control Conference (ACC), 2014, pp. 38-45. IEEE, 2014.

[47] Hao, He, Borhan M. Sanandaji, Kameshwar Poolla, and Tyrone L. Vincent. "A generalized battery model of a collection of thermostatically controlled loads for providing ancillary ser-

vice." In Communication, Control, and Computing (Allerton), 2013 51st Annual Allerton Conference on, pp. 551-558. IEEE, 2013.

[48] Sinitsyn, Nikolai A., Soumya Kundu, and Scott Backhaus. "Safe protocols for generating power pulses with heterogeneous populations of thermostatically controlled loads." Energy Conversion and Management 67 (2013): 297-308.

[49] Mehta, Nishant, Nikolai A. Sinitsyn, Scott Backhaus, and Bernard C. Lesieutre. "Safe control of thermostatically controlled loads with installed timers for demand side management." Energy Conversion and Management 86 (2014): 784-791.

[50] J. Hu; J. Cao; M. Z. Chen; J. Yu; J. Yao; S. Yang; T. Yong, "Load Following of Multiple Heterogeneous TCL Aggregators by Centralized Control," in IEEE Transactions on Power Systems , vol.PP, no.99, pp.1-1 (has been accepted for publication)

[51] H. Hao, B. M. Sanandaji, K. Poolla, and T. L. Vincent, "A generalized battery model of a collection of thermostatically controlled loads for providing ancillary service," in Proc. 51st Annu. Allerton Conf. Commun., Control and Computing, 2013, pp. 551558.

[52] H. Hao, B. M. Sanandaji, K. Poolla and T. L. Vincent, "Aggregate Flexibility of Thermostatically Controlled Loads," IEEE Trans. Power Syst., vol. 30, no. 1, pp. 189-198, Jan. 2015.

[53] S. Vandael, B. Claessens, M. Hommelberg, T. Holvoet, and G. Deconinck, "A scalable three-step approach for demand side management of plug-in hybrid vehicles," IEEE Trans. Smart Grid, vol. 4, no. 2, pp. 720728, Jun. 2013.

[54] F. Ruelens, B. J. Claessens, S. Vandael, S. Iacovella, P. Vingerhoets and R. Belmans, "Demand response of a heterogeneous cluster of electric water heaters using batch reinforcement learning," 2014 Power Systems Computation Conference, Wroclaw, 2014, pp. 1-7.

[55] F. Ruelens; B. J. Claessens; S. Vandael; B. De Schutter; R. Babuka; R. Belmans, "Residential Demand Response of Thermostatically Controlled Loads Using Batch Reinforcement Learning," in IEEE Trans. on Smart Grid , vol.PP, no.99, pp.1-11.

[56] S. Iacovella, F. Ruelens, P. Vingerhoets, B. Claessens and G. Deconinck, "Cluster Control of Heterogeneous Thermostatically Controlled Loads Using Tracer Devices," in IEEE Trans. on Smart Grid, vol. 8, no. 2, pp. 528-536, March 2017.

[57] M. Liu and Y. Shi, "Distributed model predictive control of thermostatically controlled appliances for providing balancing service," in Proc. IEEE Conf. Decision and Control, Los Angeles, CA, USA, Dec. 2014, pp. 48504855.

[58] M. Liu, Y. Shi and X. Liu, "Distributed MPC of Aggregated Heterogeneous Thermostatically Controlled Loads in Smart Grid," in IEEE Trans. on Industrial Electronics, vol. 63, no. 2, pp. 1120-1129, Feb. 2016.

[59] K. Meng, D. Wang, Z. Y. Dong, X. Gao, Y. Zheng and K. P. Wong, "Distributed control of thermostatically controlled loads in distribution network with high penetration of solar PV," in CSEE Journal of Power and Energy Systems, vol. 3, no. 1, pp. 53-62, March 2017.

[60] M. Liu and Y. Shi, "Model Predictive Control of Aggregated Heterogeneous Second-Order Thermostatically Controlled Loads for Ancillary Services," IEEE Trans. Power Syst., vol. 31, no. 3, pp. 1963-1971, 2016.

[61] A. Radaideh and V. Ajjarapu, "Extracting expedient short-term services from Homogeneous Group of Thermostatically Controlled Loads," 2016 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT), Minneapolis, MN, USA, 2016, pp. 1-5.

[62] Ilic, S. M., C. W. Bullard, and P. S. Hrnjak," Effect of shorter compressor on/off cycle times on a/c system performance," Air Conditioning and Refrigeration Center. College of Engineering. University of Illinois at Urbana-Champaign., 2001.

[63] E. F. Camacho and C. Bordons, Model Predictive Control. Berlin, Germany: Springer, 1999.

[64] Naidu, D. Subbaram. Optimal control systems. CRC press, 2002.

[65] James, Gareth, et al. "An introduction to statistical learning,". Vol. 6. New York: springer, 2013.

[66] Reliability Test System Task Force, "The IEEE Reliability Test System-1996," IEEE Trans.Power Syst, vol.14, no.3, pp.1010,1020, Aug 1999

[67] H. Pinto, F. Magnago, S. Brignone, O. Alsa, B. Stott, Security Constrained Unit Commitment: Network Modeling and Solution Issues, Proc. of the 2006 IEEE PES Power Systems Conference and Exposition, Oct. 29 2006-Nov. 1 2006, pp. 1759 1766.

[68] R. Chhetri, B. Venkatesh, E. Hill, Security Constraints Unit Commitment for a Multi-Regional Electricity Market, Proc. of the 2006 Large Engineering Systems Conference on Power Engineering, July 2006, pp. 47 52.

[69] J. Guy, Security Constrained Unit Commitment, IEEE Transactions on Power Apparatus and Systems Vol. PAS-90, Issue 3, May 1971, pp. 1385-1390.

[70] B. Hobbs, M. Rothkopf, R. ONeill, and H. Chao, editors, The Next Generation of Electric Power Unit Commitment Models, Kluwer, 2001.