2020

Data Driven Synthetic Load Modeling for Smart City Energy Management Studies

Fernando Bereta dos Reis
South Dakota State University

DATA DRIVEN SYNTHETIC LOAD MODELING FOR SMART CITY ENERGY MANAGEMENT STUDIES

BY

FERNANDO BERETA DOS REIS

A dissertation submitted in partial fulfillment of the requirements for the

Doctor of Philosophy

Major in Electrical Engineering

South Dakota State University

2020

DISSERTATION ACCEPTANCE PAGE

Fernando Bereta dos Reis

This dissertation is approved as a creditable and independent investigation by a candidate for the Doctor of Philosophy degree and is acceptable for meeting the dissertation requirements for this degree. Acceptance of this does not imply that the conclusions reached by the candidate are necessarily the conclusions of the major department.

Timothy M. Hansen
Advisor Date

Siddharth Suryanarayanan
Department Head Date

Dean, Graduate School Date

I dedicate this work to my colleagues, advisor, and my family. None of this may have been possible without their love and support.

"The greatest illusion of this world is the illusion of separation. Things you think are separate and different are actually one and the same. We are all people, but we live as if divided."

Guru Pathik

ACKNOWLEDGEMENTS

I want to thank God for life and my family, Fernando Soares dos Reis (Father), Angela Maria Bereta dos Reis (Mother), Fernanda Bereta dos Reis (Sister) and Lívia Maria Bereta dos Reis (Sister). I had a lot of support from my whole big family, grandparents, uncles, aunts, etc., in the most difficult moments it was they who kept me going, encouraging me to go ahead.

To my advisor, Dr. Timothy M. Hansen, for all the dedication, attention, understanding, patience, experience, and collaboration so that the dissertation could be completed satisfactorily. I would like to acknowledge him for conceptualizing and funding this work. This work would not be possible without his support, motivation, and enthusiasm. Besides my advisor, I would like to thank my committee, Dr. Reinaldo Tonkoski, Dr. Robert Fourney, and graduate faculty representative Dr. Laura Hasselquist for their participation and interest in my work.

Additionally, I would thank Dr. Siddharth Suryanarayana, Dr. Howard Jay Siegel, Dr. Anthony A. Maciejewski, Swagata Sharma, Berk Celik from Colorado State University (CSU) and Hendy Syahril for their contributions to this work. I would like to thank all the graduate students who were present here during my courses for their help and support in providing constructive feedback on my work.

The research in this dissertation was supported by the South Dakota Board of Regants (SDBoR) PhD funds, National Science Foundation under grant numbers ECCS-1608722, MRI-1726964, OAC-1924302, and CNS-1726946, and the U.S. Department of Energy under Grant Number DE-SC0020281.

CONTENTS

ABBREVIATIONS

Acronyms

DR	demande response
HVAC	heating, ventilation, and air conditioning
LIN-S	linear-with-time-shift
LV	low voltage
DG	distributed generation

Nomenclature

T	probability distribution of inter-arrival times
X	probability distribution of service times
C	number of servers (represents the maximum load power that can be served)
K	queue capacity (maximum number of elements in the queue)
Z	serving policy (order in which the queue is served)
P	number of all the elements that can arrive in the queue
$T/X/C$	simplified Kendall notation (assumes $K = \infty$, Z is first come first served, and $P = \infty$)
$M_t/G/\infty$	queueing load model with a time-dependent Poisson process, general probability distribution of service times, and infinite capacity
t	time
$\lambda(t)$	time-varying appliance rate of arrival in the queue

D	duration of a set of customer appliances
P	power rating of a set of customer appliances
$l(t)$	expected aggregated household load
$C_L(t)$	openly available hourly load data from any distribution company
b_{min}	minimum expected residential load for a given time period
b_{max}	maximum expected residential load for a given time period
$\underline{T}$	simulation time lower bound
$\overline{T}$	simulation time upper bound
ψ	set of customer appliances
$\mathbb{E}[P]$	expected power of appliances in the set
$\mathbb{E}[D]$	expected duration of appliances in the set
$Arrival$	list of appliances with arrival time and other user-defined appliance attributes
Δt_i	inter-arrival time between appliances
i	index of the arrival of an appliance in $Arrival$ list
app	specific appliance
$M_t/G/C$	queueing load model with a time-dependent Poisson process, general probability distribution of service times, and finite capacity
k_C	user-defined gain
$P_h(t)$	aggregated power usage
t_{add}	time an arriving appliance starts to be served
δ	simulation time resolution
$M_t/G/C_t$	queueing load model with a time-dependent Poisson process, general probability distribution of service times, and time-dependent capacity

k	Gamma distribution shape
θ	Gamma distribution scale
μ	mean
σ	standard deviation
Z_p	constant active impedance
I_p	constant active current
P_p	constant active power
Z_q	constant reactive impedance
I_q	constant reactive current
P_q	constant reactive power
V	local voltage
V_0	nominal voltage
P_0	active power at the nominal voltage
Q_0	reactive power at the nominal voltage
SW_{start}	start of the appliance scheduling window
SW_{end}	end of the appliance scheduling window
$B(t)$	expected non-appliance load
$B_l(t)$	expected household appliance load
N	number of independent customers
$C_L^*(t)$	known aggregated load curve

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

DATA DRIVEN SYNTHETIC LOAD MODELING FOR SMART CITY ENERGY MANAGEMENT STUDIES

FERNANDO BERETA DOS REIS

2020

The primary aim of this dissertation is to provide synthetic residential load models with granular level information on the customers having information about the appliances that constitute each individual residential customer through time. The synthetic load model is capable of being widely utilized by the power system research community since only publicly available data is utilized for its generation. This gives researcher's access to how the synthetic load was made and also how accurate the model is in representing real power system regions. As the title of the dissertation suggests, the synthetic residential load models are intended for smart city energy management studies. Smart city energy management studies have the ability to control tens of thousands of electricity customers in a coordinated manner to enact system-wide electric load changes. Such load changes have the potential to reduce congestion (i.e. stress on power system components) and peak demand (i.e. the need for peaking generation), among other benefits. For smart city energy management studies to have the capability of evaluating how their strategies would impact the actual power system, datasets that accurately characterize the system load are required that also contain individual loads of all buildings in a given area. Currently, such data is publicly unavailable due to privacy concerns. This dissertation's synthetic residential load model combines a top down and bottom up approach for modeling

individual residential customers and their individual electric assets, each possessing their own characteristics, using time-varying queueing models. The aggregation of all customer loads created by the queueing models represents a known city-sized load curve to be used in smart city energy management studies. The dissertation presents three queueing residential load models that make use of only publicly available data to alleviate privacy concerns. The proposed approach is mainly driven by the aggregated distribution companies load. An open-source Python tool to allow researchers to generate residential load data for their studies is also provided. The simulation results comparing the three queueing synthetic load models consider the ComEd region (utility company from Chicago, IL) to demonstrate the model's characteristics, impact of the choice of model parameters, and scalability performance of the Python tool.

The developed residential synthetic queueing load models are utilized to create the Midwest 240-Node distribution test case system, which generates appliance-level synthetic residential load for 1,120 homes for the Iowa State distribution system test case with 193 load nodes over three feeders. The Midwest 240-Node is a real distribution system from the Midwest region of the U.S. with real one-year smart meter data at the hourly aggregated node level resolution for 2017 available in an OpenDSS model. The synthetic residential queueing load model generated for the Midwest 240-Node one-year date has a mean absolute percentage error of 2.5828% in relation to the real smart meter data. The Midwest 240-Node distribution system OpenDSS model was converted to GridLAB-D to enable smart grid and transactive energy studies. The percentage of maximum error observed on voltage magnitude from the OpenDSS to GridLAB-D model is below 0.0009%. The GridLAB-D model and the generated synthetic residential load is

made publicly available. The Midwest 240-Node real distribution system with the synthetic residential load that follows the real data from smart meters is intended to be a distributed energy active consumer test system network.

The focus of the developed synthetic residential load models is smart city energy management studies; however, they can be utilized in many power systems studies to evaluate economic and technical impacts of distributed energy resources. For example, this dissertation also presents the utilization of the synthetic models for a PV rich low voltage network.

The main component of the smart grid is demand response. Demand response, or energy management, utilizes commonly passive load in to active power system resources. Residential demand response, when aggregated, is capable of performing system-wide changes that enable its participation in the power system markets. This dissertation developed residential synthetic models to enable the standardization of approaches and allow different approaches to be compared under the same environment.

The key contributions of this dissertation are:

- the development of a data driven residential synthetic queueing load model for smart city energy management studies,
- the creation of a distribution test system with the synthetic load model based on real smart meter data, to the same real distribution network from the U.S. Midwest region,
- both the residential synthetic load models and the distribution test system utilized publicly available data and are also made publicly available.

CHAPTER 1 Introduction

1.1 Background

Power system studies possess a vast range of techniques from multiple fields of knowledge. In a broad overview summary, there are: modeling, transmission, distribution, markets, optimization, and forecasting. Modeling of the power system and its elements is, in my view, the most significant part of every power system study. The assumptions made in the models and the actual elements they are represented in different conditions and scenarios must be fully understood to properly interpret the simulations results. The commonly modeled elements in power system are generators (e.g. conventional, and renewable), protection elements (e.g. relays, breakers, fusses, and lightning protection system), transmission system (e.g. long high voltage three phase lines, three phase transformers, phase shifting transformer, inductive and capacitive reactance banks), and distribution system (e.g. short low voltage lines, transformers, capacitive reactance banks, and customer loads). Markets development effectively regulates power systems; and optimization techniques minimize the cost of energy while maintaining system security. Forecasting, attempting to predict power system uncertainties, is the focus of this Introduction.

The conventional power system structure had no wind and solar generation, which are non-dispatchable forms of renewable generation. A non-dispatchable source of electricity cannot control the amount of output power in order to meet societies fluctuating electricity needs. This contrasts with flexible dispatchable power supplies, which can change their output to meet power demand. Non-dispatchable power supplies are usually

highly intermittent; thus, cannot be continuously used due to uncontrollable factors like weather. In conventional power system structure the only source of uncertainty is the system demand (i.e. load consumption), but, to keep the power system operational, the amount of energy being generated must match the consumption plus system losses. Legacy dispatchable resources have the capability of controlling their output; however, they were much slower (e.g. multiple days to reach full power) than contemporary gas generation units (e.g. less than one hour). To accommodate changes in demand, the available fast technologies were conventional hydroelectric generation and pumped storage plants. Pumped storage plants for hydroelectric power in the Unites States were built primarily between 1960 and 1990 [1]. In 2018, the United States had 22.9 gigawatts (GW) of pumped storage hydroelectric generating capacity. Pumped storage behaves as the name would suggest, pumping water into a storage reservoir at an elevated location during times of relatively low electricity demand and low electricity prices, such as during the night. When electricity demand is high, water flows downhill from the reservoir through hydroelectric generators at a dam, behaving as a battery. Similar approaches are used today when utilizing batteries to mitigate the uncertainties of the power system with increasing non-dispatchable sources such as solar and wind.

A possible approach to address the uncertainties is the deployment of demand response (DR). The new Federal Energy Regulatory Commission (FERC) regulations enable participation of demand flexibility in the market [2] is under effect. With the load flexibility provided by DR, the balancing of supply and demand can control the supply, and, to some extent, the demand. The presence of renewable non-dispatchable energy sources is not the only characteristic of the power system of the future. The distribution

network was passive, only changing their demand, but, with the deployment of distribution level resources, the distribution network is becoming more and more active. Thus, the distribution system once treated as a passive component that only changes its consumption independently of what is happening in the power system is changing. Fig. 1.1 presents the environment for power system studies with a view in to the future, where the distribution system plays an increasingly active role.

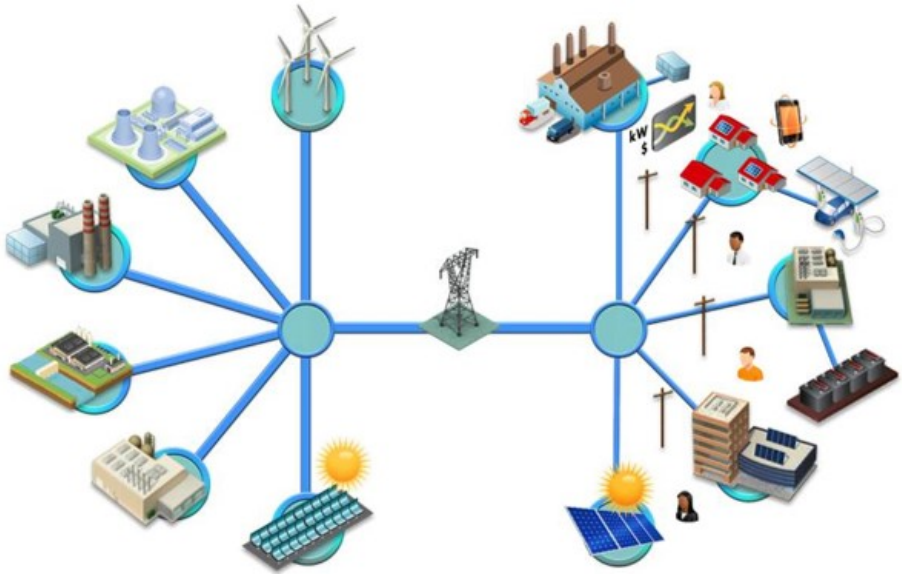

Figure 1.1. Power system studies environment and view for the future. Image adapted from [3].

Fig. 1.1 presents the high voltage power system on the left and the distribution system on the right. With the deployment of DR approaches, the distribution system is influenced by the price of energy—possibly having resources such as photovoltaic (PV), battery storage systems, electric vehicles, heat pumps, and/or other appliances. Thus, the power system is changing in order to include distribution level resources to power system operation and its markets.

The uncertainties of power systems create challenges; a relatable example to all

this year, 2020, is COVID-19. The pandemic health crises also impacted power systems [4]–[9]. Power systems are critical infrastructures for a national economy and security [5]. The quarantine and the closing down of business resulted demand reduction became another worldwide problem. Multiple nations are reporting a 3% to 10% demand reduction [6]. However, there is not just a demand reduction, the demand throughout the day has changed. In Australia, the load shape changed from a camel-like curve, having morning and evening peaks, to a much flatter curve. The demand in Britain is a mixture of weekend holidays with the presence of some businesses that are in operation remotely or not that would be closed for holidays [5]. The residential loads are now a priority load in Britain. It was made a priority since people in mandatory quarantine with no access to energy, and consequently the internet, will reduce quarantine obedience, which creates civil unrest. The California independent system operator has noticed a 13% March peak load reduction [8]. Due to California's large photovoltaic participation, the demand curve did not get flatter as in Australia. The peak-to-valley difference increased by 5%. To balance a larger demand difference, more flexible generation resources are needed. Such flexibility is not available though, so more photovoltaic generation must be dumped.

The articles [4]–[9] give an overview of the importance of the power system: grid modernization (i.e., Smart Grid), change in demand, resiliency (e.g., microgrids, flexible load, and demand response), classification of critical loads, and market impact for a short and long-term view on the participants. Grid modernization is not only necessary to effectively deploy smart strategies to improve the performance of the power system. It is also necessary to minimize the physical presence of essential personal to lower risk [5]. Southern California Edison, for example, to minimize risk to personal, has eight thousand

remote workers and five thousand essential workers [4].

In [7], the significance of resilience is discussed given that most ventilators are electric and are responsible for reducing the chances of death from COVID-19. The change in fuel mix and the resulting technical challenges are also presented in [6]. [9] urges development of power systems protocols to be able to better operate the power systems to treat COVID-19 patients and others.

1.2 State of the Art Power System Research

This section presents the background (i.e. state of the art) of four areas of interest for this dissertation. Section 1.2.1 presents the state of the art of residential demand response. Section 1.2.2 presents the state of the art for residential load modeling for DR research. Section 1.2.3 presents the state of the art for photovoltaic generation overvoltage challenge in distribution systems research. Section 1.2.4 presents the state of the art for overvoltage prevention with DR in PV rich distribution systems research.

1.2.1 Residential Demand Response Research

Residential loads represent approximately 38% of total energy consumption in the U.S. [10]. Residential demand response can provide major benefits in the electricity market: (a) participant financial benefits; (b) market-wide financial benefits; (c) reliability benefits; and (d) market performance benefits [11], [12]. Industrial customers have been utilizing DR programs developed for them given their significant demand magnitude modification [13]. To include the residential load flexibility that would have a significant impact on the system, multiple customers must participate. Residential DR makes system-wide changes that require tens of thousands of buildings, each with many

individual electric energy devices to be controlled [14].

A novel microgrid energy management system with model predictive control is presented in [15]. This management systems is capable of simultaneously considering the unity commitment constraints (i.e. in summary verifies that the generator scheduled to supply the demand does not violate physical constraints), power flow, DR, and energy storage system(s). Further, it can consider the residential load flexibility for scheduling the generator to supply the load. In [13], residential loads are separated in a multi-class queueing system. The class contains similar appliances to be optimized in a similar strategy by scheduling the flexible load with the day-ahead power system market information. In [16],automatic infrastructure is assumed to perform changes in the residential load in real-time. Thus, differently than [13] in which scheduled appliances were given the clear day-ahead energy market, [16] utilizes real-time. [17] presents a game designed aggregate game approach intending to seek the price (i.e. Nash equilibrium) of energy considering the necessary residential reward of performance changes in demand.

1.2.2 Residential Load Modeling for DR Research

Residential DR makes system-wide changes that require tens of thousands of buildings, each with many individual electric energy devices, to be controlled [14]. Evaluating the impact of residential DR strategies on electric power system operation and markets requires large-scale residential load data for use in simulation studies. The input parameters to such simulations should include the unique characteristics of each individual residential customer, along with their individual electric energy assets. The aggregate of

all such customer load data should behave as a typical city or region. Typically, large-scale customer residential data is either unavailable or proprietary due to privacy concerns [15], [18], [19]. For example, a load model that makes use of a large proprietary database that includes measurements of appliances and household loads is presented in [18]. In a second study, the interaction of DR and unit commitment of a microgrid is described [15]. The controllable smart loads are modeled with a neural network that uses measured and simulated data from an actual energy hub management system for supervised training. In the case of [19], the loading of a distribution transformer is used to generate load curves. Smart meter data has also been utilized for modeling individual home appliances in [20], where a Hidden Markov Model with differential observations is employed to attempt to identify individual appliance usage from the customer load.

Other techniques used are the static customer behavior method [16], [17], [21] (i.e. load characteristics are acquired from static parameters); statistical methods [13], [14], [22] (i.e. commonly requiring customer surveys which may not be available for different regions, customers, or times); and physical methods [23]. The table method used in [16], [21] possesses a set of schedulable appliances that is static every day for every customer. Thus, all the homes are assumed to have the same occupancy habits, not representing changes in behavior through time. In [17], the static customer behavior method does not contain schedulable appliances, but rather a reference load with upper and lower limits. With the flexibility knowledge of market participants, the convergence of customer interaction on the system is analyzed with game theory. In [13], [14], aggregated residential DR is performed by scheduling appliances, and [13] considers heating, ventilation, and air conditioning (HVAC). A probabilistic method is proposed in [14]

where 18 schedulable appliances are used to statistically generate residential loads to reflect the total energy in an average household. The appliances have a specific percentage penetration, power rating, and start time with mean and standard deviations; however, the aggregate load of all customers does not change each day nor consider regional variations. The appliances in [13] are modeled according to [18] with the same limitations for such studies. A Markov Chain Monte Carlo approach is developed in [22] that has the chain represent the state of the resident (e.g. presence of inhabitants), which affects energy consumption. The statistical model is fitted using Netherlands public data surveys. The physical method proposed in [23] develops a load simulator for residential customers considering the physical characteristics of a home. The model considers some home configurations, HVAC, and characteristics of other loads (e.g., washer, dishwasher, water heater). The method is intended to model an individual home and is not suitable to be scaled for a city-sized study.

1.2.3 Photovoltaic Generation Overvoltage Challenge in Distribution Systems Research

[1]The installation of photovoltaic (PV) generation in residential systems is rapidly increasing due to environmental concerns, decreasing costs of PV modules, and government incentives [25]. Distributed PV systems are connected to low/medium voltage (LV/MV) distribution systems in the form of distributed generation (DG), but increased PV installation has led to operational issues [26]–[32]. Traditionally, utilities use voltage compensation techniques based on line voltage drop from the substation considering

[1]This work was performed jointly with the full list of co-authors available in [24]. This work is supported by the National Science Foundation (NSF) under grant number ECCS-1608722, U.S. Department of Energy (DOE) Grant Number DE-SC0020281, and the SDSU Joint Research, Scholarship and Creative Activity Challenge Fund.

unidirectional power flow to the end-user. However, with the increasing installation of PV, power flow is not always unidirectional and can lead to voltage-rise in distribution feeders [33].

Different approaches to solve overvoltage issues due to high PV penetration in distribution systems have been discussed in the literature. Traditional voltage regulating devices, such as line voltage regulators [34], switched capacitor banks [35], and on-load tap changing transformers do not act in a sufficiently short time interval and result in poor regulation [36]. Even if these approaches did limit the voltage fluctuations, the large number of switching operations would shorten their operational life. As an alternative, utilities can increase the conductor size (decrease conductor resistance) of distribution lines to reduce voltage-rise, but upgrading the distribution system is not always economically viable [37].

PV inverter control methods for preventing overvoltage in LV distribution feeders are widely studied. Popular approaches in network independent PV inverter overvoltage control are active power curtailment (APC) based on voltage deviations [38], [39], reactive power absorption based on linear Volt/var droop [40], [41], and combined active-reactive power management using limited communication [42].

1.2.4 Overvoltage Prevention With DR in PV Rich Distribution Systems Research

The possible solutions to address voltage regulation in the presence of distributed generation are presented, e.g. curtailment, demand response (DR), and static synchronous compensator. A review of distributed and decentralized voltage control of smart distribution networks is presented in [43],where DR is presented as a strategy that should

be further explored for voltage support. The impact of DR on the distribution system voltage profile in the presence of renewable energy resources is presented in [44]. Similarly, the impact of DR on load, losses, and load factor is presented in [45]. Distribution system overvoltage due to renewable energy resources, such as photovoltaic (PV), are more likely in periods with valley demand and peak generation. The relationship between self-consumption of renewable energy resources and the required curtailment is presented in [46]. The distribution system impact of load changes to mitigate overvoltage in the presence of renewable energy resources is presented in [43], [44], [47]–[52].

A multi-agent, multi-objective renewable energy management scheme with hierarchical control is presented in [47]to balance 3 objectives: minimizing electricity bills, reducing power purchased from the main grid, and optimizing the power quality. In [48], a distributed algorithm is implemented with a multi-agent structure. The network is partitioned into zones where each zone-coordinator dispatches the active and reactive power of various distributed energy resources and DR using a gradient descent method. In [49], a distributed algorithm to control active and reactive power from PV's is presented to consider optimization in two-time scales, i.e. legacy conventional voltage control devices and fast PV inverters and DR resources. In [50], strategies to mitigate overvoltage problems in the distribution grid are discussed by presenting the change in the load having 4 setpoints based in a real-time voltage signal in a specific system. A centralized direct control optimization with receding time horizon to mitigate uncertainties is presented in [51]where the water heater of multiple customers performs the change in demand. In [52], the cost to curtail PV generation and perform load shifts is estimated with the distribution system elements models and their impact on each other (distribution network

Jacobian matrix).

1.3 Objectives

Smart grid (i.e. smart city) research with grid modernization has been the center of many power system research studies. The deployment of non-dispatchable resources increases interest in the flexibility demand DR programs provide, so much so that, according to [53], the main component of the smart grid is DR. Given the availability of data (open access) the core objectives of this dissertation are:

1. develop synthetic scalable residential load model that possess granular level information for customers (i.e. load by customer and appliances that constitute that load) and the actual behavior of the power system demand for smart cities studies,
2. deploy the generated synthetic load model to a large real distribution test system to standardize smart cities studies,
3. utilize the developed synthetic scalable residential load model for DR.

1.4 Contributions

The following contributions from this work are aimed at improving existing state-of-the-art in power system research:

1. provide the power system research community with a synthetic scalable residential load model that aggregates to the behavior of power system; thus, the impact of residential DR approaches in the power system can be identified,
2. provide a real distribution test system for the deployment of residential demand response considering the characteristics of low voltage distribution systems; thus,

having the ability to explore new smart grid initiatives impact on a real distribution system.

1.5 Dissertation Outline

In Chapter 2, the development of the synthetic scalable residential load model with queueing theory is presented. The developed synthetic scalable residential queueing load model to a real system and its validation is presented in 3. Chapter 4 presents the models utilized for a residential DR approach in a PV rich distribution network. In Chapter 5, the possible expansion of the synthetic residential load model to consider HVAC is presented. Also, a simplified approach for selecting the queueing load mode is presented. Finally, in Chapter 6, the conclusion and future work is presented.

CHAPTER 2 Synthetic Residential Load Models for Smart City Energy Management Simulations

2.1 Overview

The ability to control tens of thousands of residential electricity customers in a coordinated manner has the potential to enact system-wide electric load changes, such as reduce congestion and peak demand, among other benefits. To quantify the potential benefits of demand side management and other power system simulation studies (e.g., home energy management, large-scale residential demand response), synthetic load datasets that accurately characterize the system load are required. This chapter designs a combined top-down and bottom-up approach for modeling individual residential customers and their individual electric assets, each possessing their own characteristics, using time-varying queueing models. The aggregation of all customer loads created by the queueing models represents a known city-sized load curve to be used in simulation studies. The three presented residential queueing load models use only publicly available data. An open-source Python tool to allow researchers to generate residential load data for their studies is also provided. The simulation results presented consider the ComEd region (utility company from Chicago, IL) and demonstrate the characteristics of the three proposed residential queueing load models, impact of the choice of model parameters, and scalability performance of the Python tool.

This work was performed jointly with the full list of co-authors available in [54]. This work was supported by the National Science Foundation under grant numbers ECCS-1608722 and CNS-1726946.

2.2 Introduction

Residential loads represent approximately 38% of the total energy consumption in the U.S. [10]. Residential demand response (DR) can provide major benefits in the electricity market: (a) participant financial benefits; (b) market-wide financial benefits; (c) reliability benefits; and (d) market performance benefits [11], [12]. Residential DR makes system-wide changes that require tens of thousands of buildings, each with many individual electric energy devices, to be controlled [14]. To evaluate the impact of residential DR strategies on electric power system operation and markets requires large-scale residential load data for use in simulation studies. The input parameters to such simulations should include the unique characteristics of each individual residential customer, along with their individual electric energy assets. The aggregate of all such customer load data should behave as a typical city or region.

Typically, large-scale customer residential data is either unavailable or proprietary due to privacy concerns [15], [18], [19]. For example, a load model that makes use of a large proprietary database that includes measurements of appliances and household loads is presented in [18]. In a second study, the interaction of DR and unit commitment of a microgrid is described [15]. The controllable smart loads are modeled with a neural network that uses measured and simulated data from an actual energy hub management system for supervised training. In the case of [19], the loading of a distribution transformer is used to generate load curves. Because the data in these studies is not publicly available, it is not possible to replicate the simulation results, compare new DR and other demand side management methods to others, or generalize the results to other

customers/regions. Smart meter data has also been utilized for modeling individual home appliances in [20], where a Hidden Markov Model with differential observations is employed to attempt to identify individual appliance usage from the customer load.

Other techniques used are the static customer behavior method [16], [17], [21] (i.e., load characteristics are acquired from static parameters); statistical methods [13], [14], [22] (i.e., commonly requiring customer surveys which may not be available for different regions, customers, or times); and physical methods [23]. The table method used in [16], [21] possesses a set of schedulable appliances that is static every day for every customer. Thus, all the homes are assumed to have the same occupancy habits, not representing changes in behavior through time. In [17], the static customer behavior method does not contain schedulable appliances, but rather a reference load with upper and lower limits. With the flexibility knowledge of market participants, the convergence of customer interaction on the system is analyzed with game theory. In [13], [14], aggregated residential DR is performed by scheduling appliances, and [13] considers heating, ventilation, and air conditioning (HVAC). A probabilistic method is proposed in [14], where 18 schedulable appliances are used to statistically generate residential loads to reflect the total energy in an average household. The appliances have a specific percentage penetration, power rating, and start time with mean and standard deviation, however the aggregate load of all customers does not change each day nor consider regional variations. The appliances in [13] are modeled according to [18], with the same limitations for such studies.

A Markov Chain Monte Carlo approach is developed in [22], having the chain represent the state of the resident (e.g., presence of inhabitants) which affects energy

consumption. The statistical model is fit using Netherlands public data surveys. The physical method proposed in [23] develops a load simulator for residential customers considering the physical characteristics of a home. The model considers some home configurations, HVAC, and characteristics of other loads (e.g., washer, dishwasher, water heater). The method is intended to model an individual home, and is not suitable to be scaled for a city-sized study.

The literature that presents residential customer load models can be roughly divided into two categories — top-down and bottom-up [22] — which may or may not make use of proprietary data. Top-down models use aggregated load to generate individual load curves [15], [18], [19]. The study in [15] has knowledge of the aggregated load and smart appliances, while [19] only has knowledge of the aggregated distribution transformer. The bottom-up approach uses the given characteristics of appliances and statistical behavior of customers to generate the load profile [13], [14], [16], [17], [21]–[23]. Table 2.1 summarizes the methods in literature and demonstrates the need for the proposed synthetic residential load models. The table is divided into the required inputs for the literature methods, and the generated outputs. The numerous load models present in literature (i) are dependent on data that is not publicly available nor applicable to all regions of study, (ii) do not aggregate to the system load curve, or (iii) maintain the same daily customer behavior throughout the simulation. There is a need for residential load data that is openly available (i.e., results can be replicated and compared), aggregates to a known system curve (i.e., large-scale studies that represent the expected electric energy behavior of a city), varies through time (e.g., hourly, daily, and seasonal variation), and does not require extensive customer surveys as they may not be available for every

region.

Table 2.1. Literature methods classification

Customers demand input data	**Synthetic customer demand output**
Statistical or survey derived data [13], [14], [16]–[18], [21]–[23]	Distinct day customer load variability (does not aggregate to system load curve) [13], [15], [18], [19], [22]
Individual customer measurement [18], [20]	Low load variability (might or might not aggregate to system load curve) [16], [17], [21], [23]
Aggregation of small regions [15], [19]	
Aggregated load information by utility None	Aggregates to system load curve None

This chapter addresses the needs of residential large-scale load data by using flexible time-varying queueing models to generate synthetic residential load data to allow simulation studies to be replicated and compared by the research community to new state-of-the-art methods. The proposed top-down and bottom-up approach addresses the challenge of unavailable and proprietary customer data by utilizing available aggregate load data for a region as an input to generate individual load profiles comprised of individual residential electric assets. The aggregate of the individual synthetic customer load data generated by the queueing models properly represents a known system load curve and contains the time-varying characteristics of an actual power system region.

In this chapter we expand on the $M_t/G/\infty$ queueing load model from [55], which incorporates the arrival of appliances that comprise aggregate individual residential customer load with time-varying behavior. In that work, however, the physical limitations of homes were not considered, the appliance model was limited to active power

consumption only, and scalability for city-size synthetic load datasets is not addressed. Thus, the primary contributions in this work are:

- Design of two new synthetic time-varying residential queueing load models that are capable of incorporating the physical limitations of residential customers without loss of generality;
- Creation of a general residential appliance model that possesses many attributes (e.g., power, duration, schedulability, ZIP load parameters), and is extensible according to the necessities of the specific study being performed; and
- Development of a scalable Python tool that generates the synthetic residential load models in parallel using high-performance computing.

The remainder of this chapter is organized as follows: Section 2.3 presents the overview of queueing theory and the residential queueing load models. The appliance model and its variations to incorporate voltage dependencies, scheduling windows, and the consideration of non-arriving loads (e.g., HVAC) are presented in Section 2.4. Section 2.5 presents the necessary inputs for the proposed synthetic residential load models. The behavior of the proposed models is presented and validated in Section 2.6. Concluding remarks on the models are discussed Section 2.8.

2.3 Queueing Load Models

2.3.1 Overview

Queueing theory models the behavior of a queue, i.e., waiting in line, and was initially employed in the communication field (e.g., phone operators) to evaluate the

performance of the system and determine how to operate the system more efficiently. Queueing theory can model applications in many disciplines and provides useful insight into distinct systems. In [56], an emergency hospital modeled the random arrival of admitted patients as a queue and the condition of patients as the priority. With the model, the flow of patients can be analyzed (e.g., queue waiting time). The authors in [57] make use of queueing models for the load profile of plugin electric vehicles at charging stations, similar to the application in this study.

Queueing models are defined by the probability distribution of inter-arrival times, probability distribution of service times, number of servers, queue capacity, size of the population, and a service discipline. Furthermore, the characteristics can be constant or time-dependent (e.g., inter-arrival times as a function of time). Fig. 2.1 presents an overview of the behavior of a queue, having the probability distribution of inter-arrival times T. An element arrives in the queue at a time and possesses its own characteristics (e.g., priority of arriving jobs in a server). The probability distribution of service times is defined by X, i.e., distribution of time to serve an arriving element. The number of servers, C, is a physical constraint of the system (e.g., the maximum capacity of the servers to serve the arriving jobs). The characteristics of the queue are its capacity, K, and serving policy, Z (i.e., the maximum number of elements in the queue and the order in which they will be served). The size of the population P, is the number of all the elements that can arrive in the queue.

For all the queueing models presented in this chapter, the following three assumptions have been made: the queue length is infinite (i.e., $K = \infty$, no loss of appliances arriving to the system); the population is infinite (i.e., $P = \infty$, arrival process is

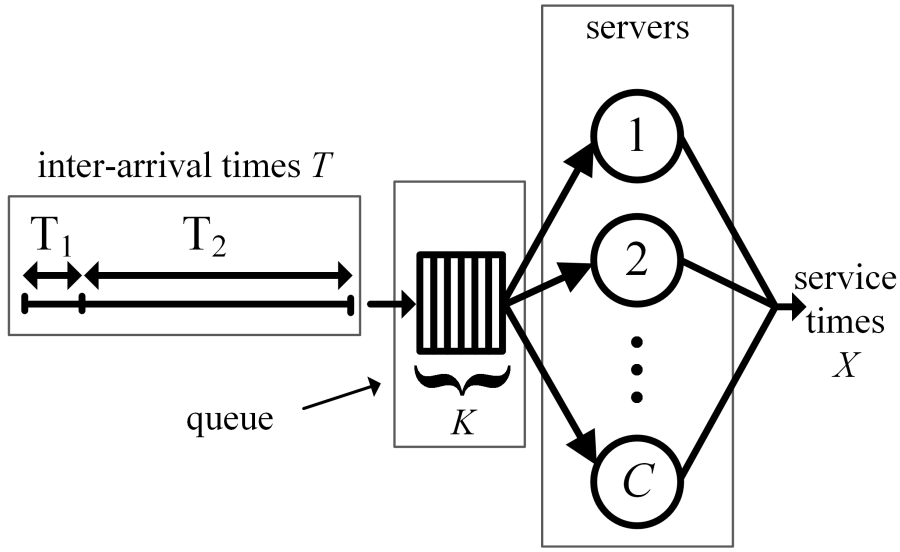

Figure 2.1. Overview of the queueing model. Random arrival of elements in the queue (defined by inter-arrival times T), size (K) and serving policy of the queue, number of available servers (C), and the service time (X).

not dependent on the appliances currently present in the system); and the service policy is first come first served. Given the queueing model assumptions, the simplified Kendall notation is used to described the queue behavior, i.e., $T/X/C$. The main advantages of using queueing theory for generating synthetic residential loads is its relation to load:

- Residential customers use their appliances according to their individual behavior, thus from the view of an outside observer the start times of appliances are random. Customers are distinct and assumed to be unable to influence the usage of other customers. Nevertheless the aggregation of a considerable number of customers is known (i.e., top-down);

- Every arriving appliance possesses its own different characteristics as generic elements, e.g., electric, temporal, and schedulability; and

- Non-homogeneous Poisson process has an average rate of arrivals that varies

through time. Thus, the time-varying characteristics of hourly, daily, weekly, and seasonal behavior of load is naturally considered.

The queueing studies presented in this chapter are specifically interested in the probability distribution of inter-arrival times and usage of servers as it applies to residential electric load. Fig. 2.2 illustrates the overall behavior of the queueing load models proposed in this chapter for a single home. The arriving elements are appliances with a time-varying exponential distribution. Thus, during peak load there is a larger probability of small inter-arrival times (i.e., more electric energy arriving into the system) and the opposite for valleys. The output of the queueing process in this work is the utilization of "servers" which corresponds to the active power consumption of a residential home. When representing electric load, the serving capacity need not be represented by an integer (e.g., an arrival process could utilize 50.239 servers at a given time as active power is in the real number set).

Sections 2.3.2–2.3.4 present the three proposed queueing load models. Fig. 2.3 presents the overall procedure the load models follow. The load generation for each customer is independent, thus the overall queueing procedure is the same for all customers. Each model possesses unique characteristics, but the overall procedure is maintained for all.

2.3.2 $M_t/G/\infty$ Queueing Load Model

The $M_t/G/\infty$ queueing load model presented here builds on the previous work in [55]. The queueing load model represents inter-arrival times as a time-dependent Poisson process ($T := M_t$), the probability distribution of service times is general

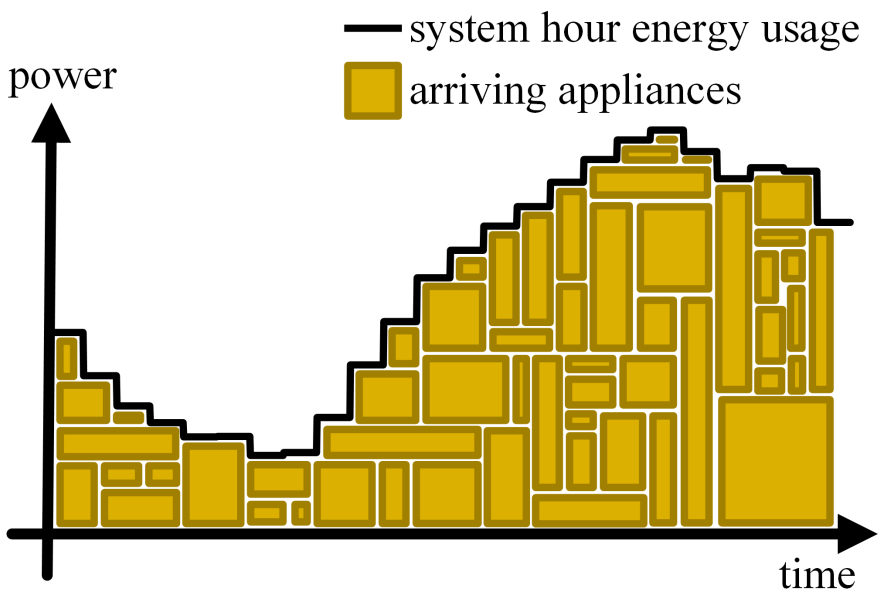

Figure 2.2. Queueing load model output. The utilization of servers at a given time represents the aggregation of appliances (yellow boxes where the height is active power consumption width is time duration, and area is energy consumption) being utilized, thus resulting in a load curve (block line).

($X := G$), and the power capacity in the home is infinite ($C := \infty$). The arrival of appliances in the $M_t/G/\infty$ queueing model is time-dependent to capture the temporal behavior of customers. Furthermore, because there is an infinite capacity, the arriving appliances are served as soon as they arrive in the queue.

At time t, let $\lambda(t)$ be the time-varying appliance rate into the system, D and P be the random variables describing the duration and power rating of the set of customer appliances, respectively, and $l(t)$ be the expected aggregated household load. The time-varying appliance rate into the system with a Poisson process is described as

$$\lambda(t) = \frac{l(t + \mathbb{E}[D])}{\mathbb{E}[P]\,\mathbb{E}[D]}. \tag{2.1}$$

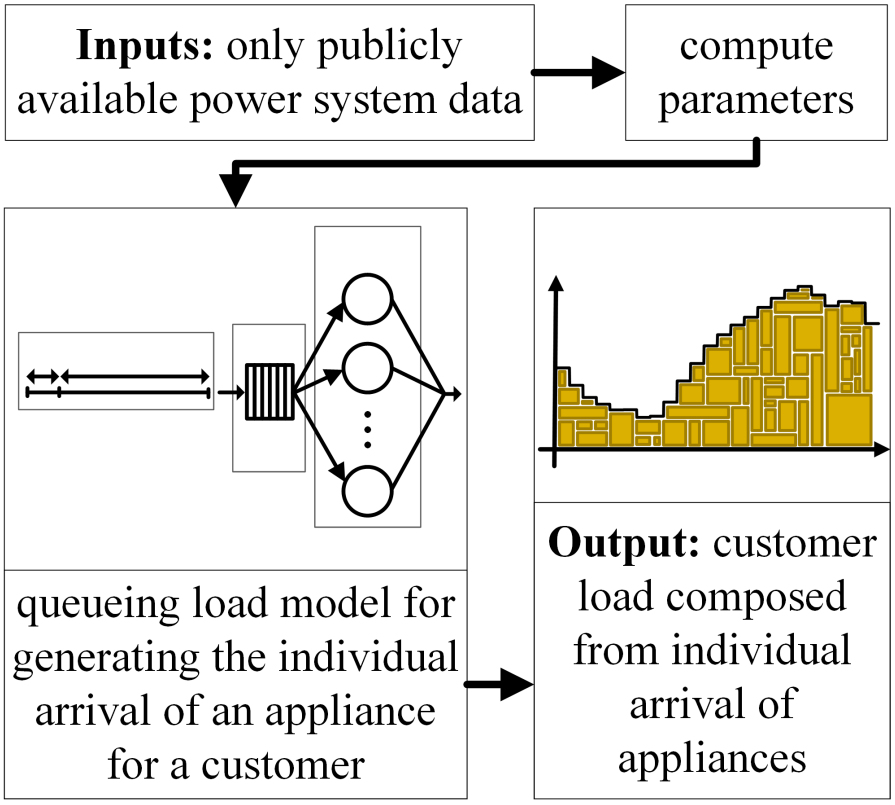

Figure 2.3. Schematic overview of the queueing load model generation procedure for a single residential customer.

The mathematical derivation of (2.1) is presented in [55]. The derivation makes use of the linear-with-time-shift (LIN-S) approximation from [58] and assumes no causality as the data is to be used in simulation studies.

The expected home load $l(t)$ is generated with the openly available hourly load data $C_L(t)$ from any distribution company. $C_L(t)$ naturally describes the aggregated behavior of customers in a given region containing its geographic characteristics, e.g., climate and customer preferences. The load data is scaled to generate the expected individual home load (2.2), where b_{min} and b_{max} are the minimum and maximum expected residential load for a given time period.

$$l(t) = b_{min} + \frac{C_L(t) - min(C_L)}{max(C_L) - min(C_L)} \cdot (b_{max} - b_{min}) \tag{2.2}$$

A flow chart for generating the synthetic residential load model $M_t/G/\infty$ is presented in Fig. 4. The user-defined input parameters are the load scaling factors b_{min} and b_{max}, and the range of the simulation time from $\underline{T}$ to $\overline{T}$. The data input is the aggregate load curve $C_L(t)$ and the set of appliances ψ, which will be discussed in detail in Section 2.4.1. The expected power $\mathbb{E}[P]$ and expected duration $\mathbb{E}[D]$ used in (2.1) are computed for the customer set of appliances ψ. The process starts by initializing the variables where *Arrival* is a list of appliances with arrival time and other user-defined appliance attributes (e.g., schedulability, ZIP load parameters). This general appliance model allows the researcher to extend the output to match the study of interest (e.g., home energy management). The variable Δt_i is the inter-arrival time between appliances (i.e., time for next arrival).

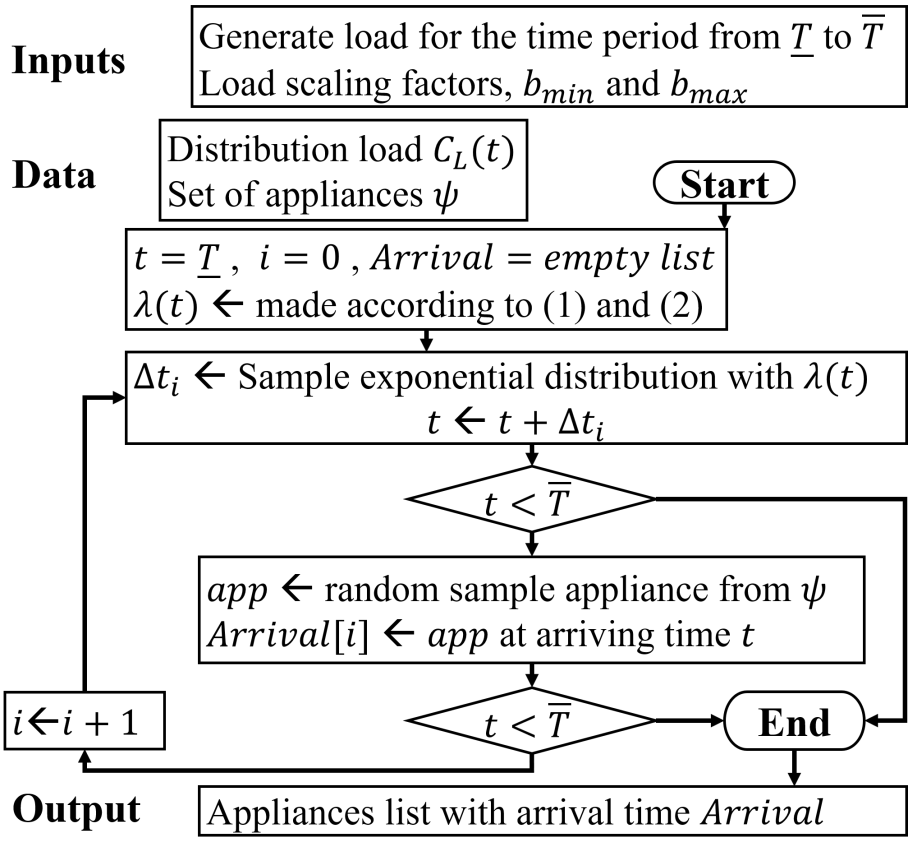

Figure 2.4. Synthetic residential queueing load model $M_t/G/\infty$ for a single customer.

The $M_t/G/\infty$ queueing model serves the appliances as soon as they arrive in the system. Making use of (2.1), the generated customer load when aggregated approximates the known load of the distribution company, $C_L(t)$. The number of servers being infinite is justifiable because as soon as an appliance is turned ON, it instantaneously starts operating, i.e., no waiting in line to consume power. Due to the intrinsic random behavior of the queueing model, however, the generated load peaks could surpass an individual residential building peak load consumption.

The unrealistic load peaks are a result of multiple appliances arriving in a short amount of time, commonly referred to as burstiness [59]. For the same arrival rate $\lambda(t)$, different burstiness levels can occur. As the elements that form the distribution test case and residential building have physical limitations, to be able to utilize the load generated by the queueing model in distribution system test cases the residential peak load may need a limit. To address this characteristic, the assumption of an infinite number of servers can be changed which will be addressed in the two new proposed queueing load models in the following sections.

2.3.3 $M_t/G/C$ Queueing Load Model

To address the issue with unrealistic peak load from the $M_t/G/\infty$ model, the serving capacity of the queue can be limited. In the $M_t/G/C$ queueing model, the system is unable to serve an infinite amount of arriving appliances. In this model, when an appliance arrives in the queue it may no longer be immediately served, but it will instead depend on the available capacity. This addresses the issue of unfeasible peak load consumption of a residential home given the physical limitations of the system. In the

$M_t/G/C$ queueing load model, the power capacity/maximum load consumption that can be served must be defined based on the physical limitations of the power system. One such method for setting the limit while maintaining the independent nature of each customer queue is by defining a gain based on the customer's expected home load. For a given customer, let C be the residential home capacity for the queue model, and k_C be the user-defined gain. We define, for a given customer, the residential home capacity as:

$$C = max\,(l(t))\,k_C. \tag{2.3}$$

The power capacity C can be defined from (2.3) or be explicitly chosen by the user. In this work, we assumed all customers have the same C, k_C, and scaled $l(t)$ according to (2) and (3), but this can be scaled based on home sizes of the particular study of interest with no loss of generality.

With the limitation of the residential home capacity in the $M_t/G/C$ queueing load model, unfeasible peak load consumption given the physical limitations of a system are no longer created. However, in the $M_t/G/C$ queue there may still be unrealistic peaks for low values of $l(t)$. Thus, even though the load does not surpass the *physical* limitations of the system, unrealistic load peaks based on customer *behavior* may still be generated by the queueing model as illustrated in Fig. 2.5. Depending on the analyses being performed by the user, the unrealistic peaks for low values of $l(t)$ may or may not be relevant.

The flow chart for generating the synthetic residential load model using the $M_t/G/C$ queue is presented in Fig. 6. The simulation time t is not necessarily the time an appliance will be served. The aggregated power usage from the appliances *actual* run time

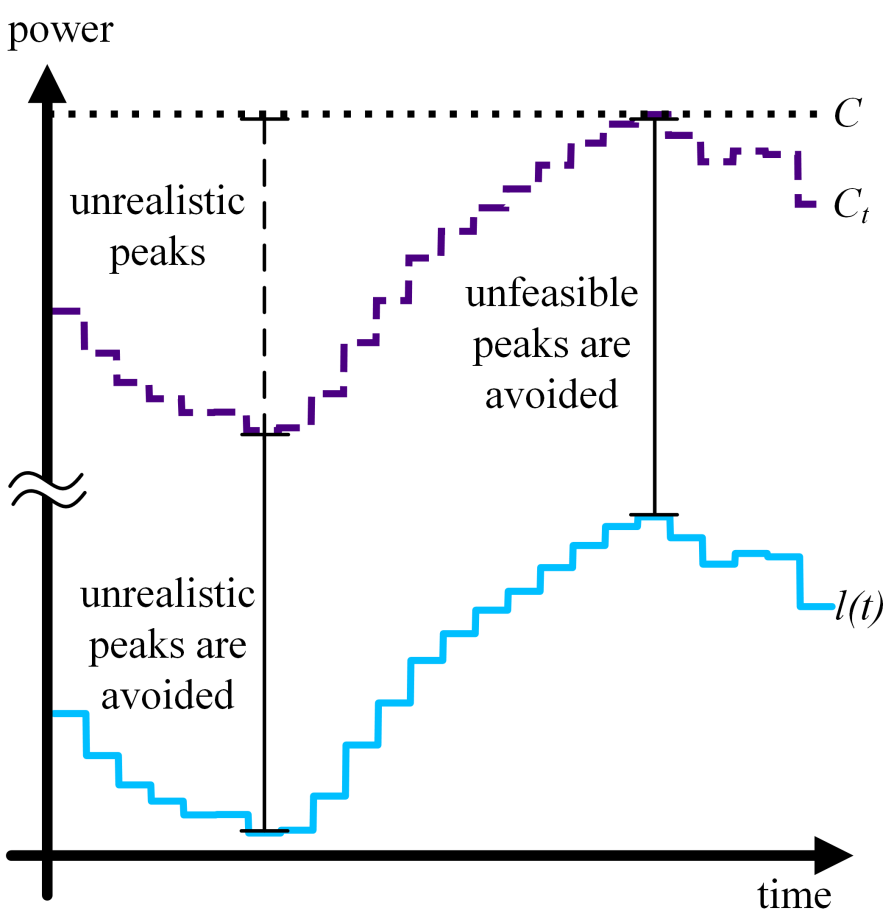

Figure 2.5. $M_t/G/C$ and $M_t/G/C_t$ queueing load models power capacity, C and C_t, respectively. The $M_t/G/\infty$ queueing load model is unbounded.

(i.e., when the appliance actually runs, not when it arrives in the queue), given by $P_h(t)$, is necessary because there is a power capacity and the appliance may need to wait in the queue, with a service policy of first come first served. The appliance, after being sampled from the set of appliances, will be served as soon as possible given the limitation of the power capacity. The time an appliance will be served t_{add} is searched in the internal loop where δ is the simulation time resolution, hence the load will never be greater than C.

2.3.4 $M_t/G/C_t$ Queueing Load Model

The $M_t/G/C$ addressed the issue of unfeasible peak load consumption given the physical limitations of a system, but it may have unrealistic peaks for low values of $l(t)$ given the expected customer behavior, as illustrated in Fig. 2.5. The $M_t/G/C_t$ queueing load model addresses the issue of unrealistic peaks for low values of $l(t)$ by replacing the constant power capacity C with a time-varying power capacity C_t. The time-varying power capacity, C_t, can be defined from (2.4) or any user-defined time-varying curve. In this work, we assumed all customers have the same C_t, k_C, and scaled $l(t)$ according to (2) and (3), but this can be generalized to a time-varying k_C. The time-varying C_t is calculated as:

$$C_t = l(t)k_C. \tag{2.4}$$

With the limitation of the power capacity $M_t/G/C_t$, unfeasible and unrealistic peak load consumption given the physical limitations of a system and the expected customer behavior are no longer created, Fig. 2.5. The procedure for the $M_t/G/C_t$ queueing load model is the same as the flow chart in Fig. 6, except C is replaced with C_t, thus the power

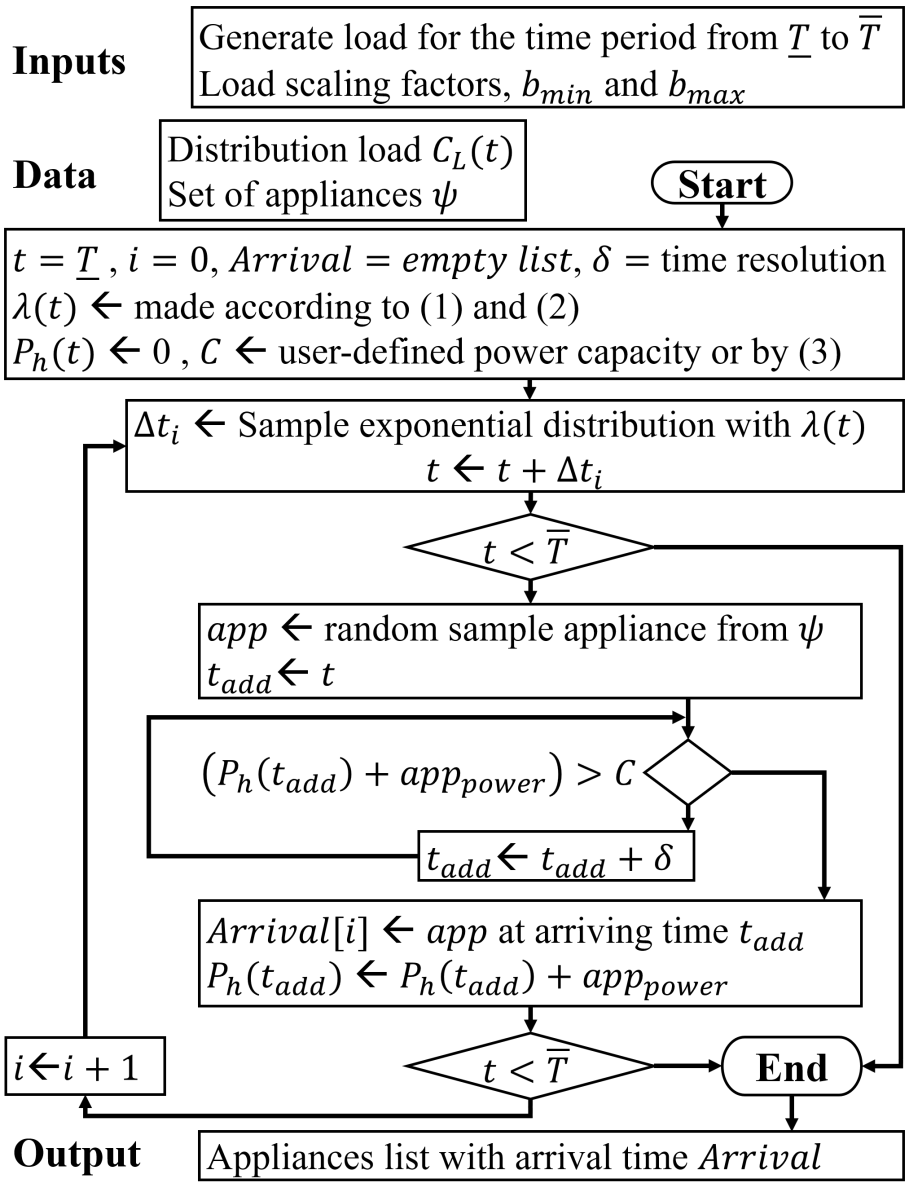

Figure 2.6. Synthetic residential queueing load model $M_t/G/C$ for a single customer.

capacity is computed with (4) and the internal loop condition is replaced by:

$$(P_h(t_{add}) + app_{power}) > C_t(t_{add}). \tag{2.5}$$

2.4 Appliance Model

2.4.1 Generic Appliance Model

The queueing model used to generate the synthetic load data is comprised of arriving appliances, therefore it is necessary to consider the assumptions of the appliances. In Sections 2.3.2–2.3.4, the synthetic queueing load models randomly sample a set of appliances ψ. Appliances are studied in [60] presenting the electric power consumption of household appliances, and the data is available online. The appliance model in Fig. 2.7a presents a generic appliance as a block of energy, having constant-power draw over a given time duration. In Fig. 2.7b, the output of the generic constant-power draw model is compared to a washing machine with time-varying power consumption from [60] with one-minute resolution. The generic appliance model consumes the same amount of energy as the actual appliance, just at different rates throughout the appliance duration. To validate this approximation, we compare the energy consumption of the two models through the appliance duration, Fig. 2.7c. Although the two models rarely consume the same power at any given time, the total energy consumed is the same and is relatively close throughout the duration. Therefore, for energy management simulation purposes, the generic constant load model is considered adequate and will be used in this chapter. Additionally, the output of the $M_t/G/\infty$ queueing load model was shown to work with the real appliance power in [55] (i.e., due to G in the queueing models). The models presented

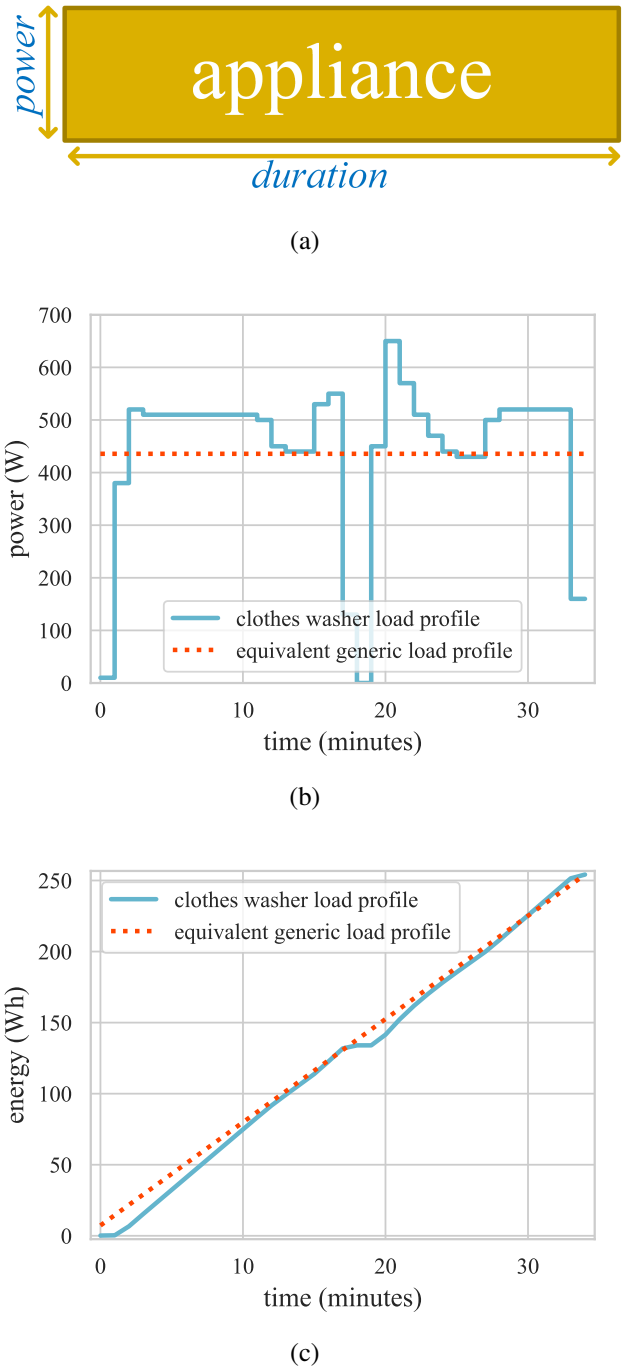

Figure 2.7. Generic appliance model characteristics, assumptions, and justification. (a) Illustrates the simplifying assumption of the appliance model of a constant power draw and defined time duration; (b)the load profile of the washer from [60] (blue line) versus the equivalent constant average load profile (red dotted line) at a one-minute resolution; and (c) the energy consumption of the real (blue line) versus equivalent model (red dotted line) through time.

in this chapter will work with any set of appliances.

To generate the appliance power and duration, two gamma distributions are randomly sampled. Gamma distributions are continuous probability distributions in the positive real number set (a useful characteristic given the appliance power and duration must be positive) defined by two parameters (i.e., shape k and scale θ). The mean of a gamma distribution is $\mathbb{E}[X] = k\theta$, and the variance is $Var(X) = k\theta^2$. Thus, by defining the mean μ and standard deviation σ, the gamma parameters k and θ are computed with $k = \mu^2/\sigma^2$ and $\theta = \sigma^2/\mu$. Fig. 2.8 illustrates the two gamma distributions sampled that determine the power and duration of the appliances. The gamma distributions are made from the expected mean and standard deviation of power in W and the mean and standard deviation of the duration in time.

The *Arrival* list (i.e., the output) contains the characteristics of all appliance power ratings in W and the time duration in hours, for each and every appliance that has arrived in the queue. Each element in *Arrival* represents a single appliance. Every row in *Arrival* represents an appliance i with its characteristics. Thus, *Arrival* contains all the arriving appliances for a residential customer for the generated simulation period (i.e., from $\underline{T}$ to $\overline{T}$). The appliance model can possess more characteristics depending on the user-defined study of interest without significant changes to the model, further discussed in Sections 2.4.3 and 2.4.4.

2.4.2 Overview of Appliance Model Variations

The synthetic residential queueing load models presented are considerably flexible. By making small changes to the appliance and load inputs, the ability of the

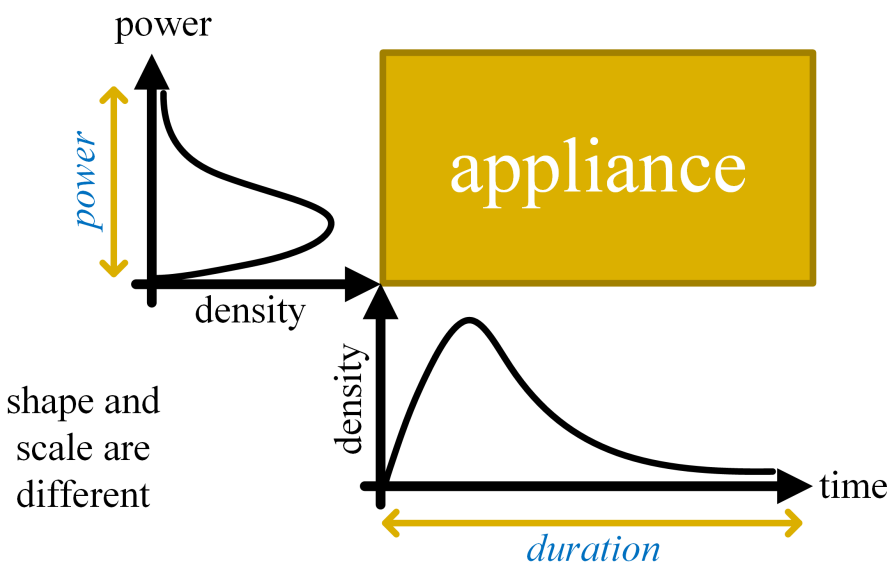

Figure 2.8. Random sampling of two distinct gamma distributions to define appliance power and duration. The shape and scale of the gamma distribution are defined based on mean and standard deviation of actual appliances.

models to address a wide range of researcher-specific projects can be achieved.

Section 2.4.3 incorporates ZIP polynomial appliance characteristics and reactive power to the generic appliance model (i.e., appliance consumption has active and reactive power that are dependent on the local voltage of the customer) to be used in distribution voltage control studies. Scheduling characteristics of appliances are provided in Section 2.4.4 to be used in energy management studies. In Section 2.4.5, it is demonstrated how the reference curve $l(t)$ can be altered so that a defined portion of the customer load is non-arriving, allowing other residential energy devices to be modeled that do not behave as the generic appliance model (e.g., HVAC, batteries, electric vehicles).

2.4.3 ZIP Appliance Load Model

Loads in a low voltage (LV) residential distribution network have a dependency on voltage [61]–[63]. Distributed generation (DG) presents technical challenges in LV

distribution networks. The voltage must be maintained within a predefined range, thus the power consumption of appliances and generation from DG in LV networks interact indirectly through voltage. Studies that analyze DG in LV distribution system networks or microgrids could make use of the presented queueing load models (e.g., [26], [38], [64], [65]). The queueing load models generate distinct load consumption patterns for every residential customer, allowing impact assessment of the change in load on the change in local voltage.

Considerable changes in appliances have occurred in recent years due to advances in power electronics, leading to a change in load characteristics. ZIP load models have been used to characterize the load dynamics with respect to voltage [62], [63]. In New York City, a study was conducted to characterize the effects of voltage variations in load consumption with field validation [62] with the intention of energy conservation using Volt/var control at the substation level [61].

ZIP load models are flexible, the parameters are easily changed to better represent load dynamics, and reduce to other load models (e.g., constant active power, constant active and reactive power, constant resistance, constant impedance). ZIP is a static representation of load models, and assumes that the static characteristics of the active power of a load can be defined by three components: constant impedance (Z_p), constant current (I_p), and constant power (P_p), represented by (2.6). Similarly, reactive power dynamics can be obtained by (2.7) using the parameters Z_q, I_q, and P_q. In (2.6) and (2.7), V is the local voltage, V_0 is the nominal voltage, P_0 is the reference active power at the nominal voltage, and Q_0 is the reference reactive power at the nominal voltage. Load models that behave as constant resistance and/or impedance have a quadratic dependence

on voltage change, where load models that behave as constant current have a linear dependence on any change in voltage.

$$P = P_0 \left[Z_p \left(\frac{V}{V_0} \right)^2 + I_p \left(\frac{V}{V_0} \right) + P_p \right] \tag{2.6}$$

$$Q = Q_0 \left[Z_q \left(\frac{V}{V_0} \right)^2 + I_q \left(\frac{V}{V_0} \right) + P_q \right] \tag{2.7}$$

The ZIP coefficients have the following two constraints:

$$Z_p + I_p + P_p = 1 \tag{2.8}$$

$$Z_q + I_q + P_q = 1 \tag{2.9}$$

2.4.4 Appliance Scheduling Characteristics

Energy management benchmarks as in [13], [14], [16], [21], [66] schedule appliances in a time window to achieve a goal (e.g., reduce cost of energy). In a similar method to the ZIP characteristics, scheduling parameters can be added for the arriving appliances from the queueing models to allow the synthetic load model presented in this chapter to generate inputs for the studies in [13], [14], [16], [21], [66] and more energy management studies.

To create the scheduling characteristics for each appliance, it needs to be determined which appliances are schedulable, and the scheduling constraints (e.g., start and end time of the scheduling window). Therefore, the appliances generated by the

queueing models require more inputs to define these characteristics. One method for determining if each appliance is schedulable is to generate a random sample and compare it to a user-defined threshold, which determines the percentage of appliances that should be schedulable. To specify the scheduling constraints, two gamma distributions are created and sampled to specify the start of the scheduling window SW_{start}, and the end of the scheduling window SW_{end} (i.e., the time after the arrival plus the duration of the appliance), as illustrated in Fig. 2.9. The scheduling characteristics could be further extended depending on the user's study of interest.

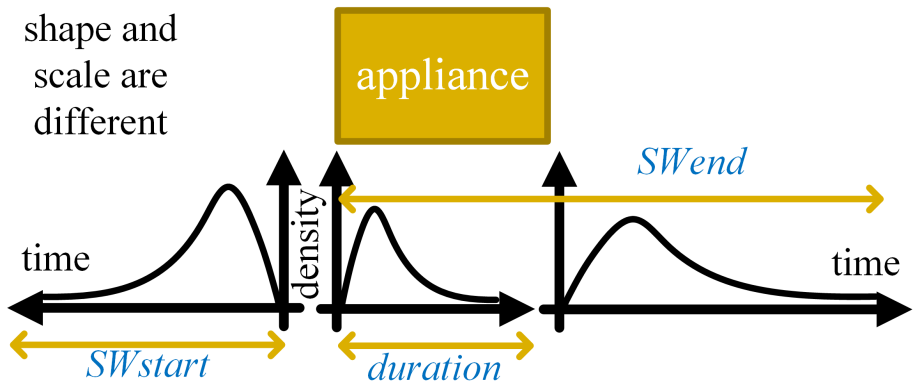

Figure 2.9. Random sampling two distinct gamma distributions to define the appliance scheduling window start and end. The shape and scale of the gamma distribution are defined based on mean and standard deviation of the scheduling window.

2.4.5 Non-Arriving Loads

Residential customers have more electric energy devices than just appliances modeled by the synthetic queueing load models. Portions of the customer load profile possess climate dependencies, such as HVAC and electric water heaters. In the proposed queueing models, these are modeled as a conjunction of non-schedulable appliances rather than containing their climate dependencies. As the energy consumption of such thermal

loads changes based on use and climate, the energy is not able to be directly shifted to a more opportune time (i.e., preheating or cooling a home does not imply that the same amount of energy would be used at a later time). Most studies that consider thermal loads take into account comfort to not violate the comfort of the inhabitants [13], [16], thus proper thermal models must be used and need to have their energy separated from conventional appliances and the presented queueing models.

Furthermore, other electric energy devices (e.g., electric vehicles) have energy requirements dependent on usage, and are also not always available at the residence. The use of the large battery from electric vehicles in DR is appealing, but the characteristics of such a resource requires different considerations than the appliance model.

The presented queueing load models in Section 2.3 can be used with a small adjustment to consider electric energy devices with different dependencies, characteristics, and behavior. Non-arriving appliances (e.g., HVAC) should be removed from $l(t)$ prior to use in the queueing process. At time t, let $B(t)$ be the expected non-appliance load, and $B_l(t)$ be the expected household appliance load. If modeling non-appliance load, $B_l(t)$ is used in place of $l(t)$ for the queueing load models, and is defined as:

$$B_l(t) = l(t) - B(t). \tag{2.10}$$

Fig. 2.10 demonstrates the removal of the non-arriving appliances from $l(t)$, thus maintaining the aggregated behavior of the customers. The sum of the generated appliances from $B_l(t)$ in the queueing load models plus $B(t)$ from all customers has the same proportions to the sum of the original load $l(t)$ from all the customers. This allows

researchers the ability to model any electric energy device in conjunction with the synthetic queueing load models without losing the aggregation property to the known input load curve.

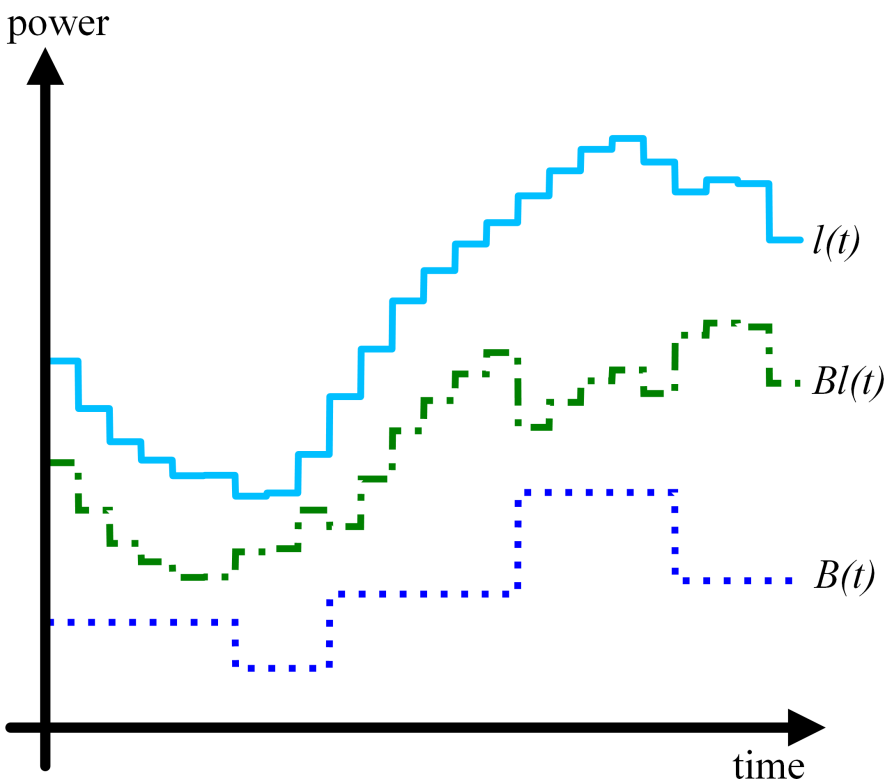

Figure 2.10. Theoretical illustration of removing non-arriving appliance loads $B(t)$ (e.g., HVAC) from the aggregated household load $l(t)$, thus generating the new $B_l(t)$ to be used in the queueing models. The sum of the two will still approximate the known input load curve, $l(t)$.

2.5 Synthetic Queueing Load Models Inputs

2.5.1 Inputs to the Queueing Load Models

There are two input groups for all the queueing models: the publicly available aggregate distribution system load data and the appliance parameters. The historic data obtained from distribution companies is input as time-series $C_L(t)$ from (2.2). The

residential building load curve $l(t)$ is generated with the user-defined choice of b_{min} and b_{max} (i.e. minimum and maximum expected residential load characteristics).

The appliances arriving into the queue are created as presented in Section 2.4.1, thus generating the input set of appliances ψ. To generate ψ, the following parameters must be chosen by the user:

- Number of appliances (i.e., size of the set ψ);
- Standard deviation and mean power of appliances (i.e., inputs to the gamma distribution). The selection has a direct impact in the expected power of the set $\mathbb{E}[P]$ (y-axis of Fig. 2.7a);
- Standard deviation and mean duration of appliances (i.e., inputs to the duration gamma distribution). The selection has a direct impact in the expected duration of the set $\mathbb{E}[D]$ (x-axis of Fig. 2.7a);

Thus, the selection of the gamma distributions to generate the appliances will impact the arrival rate of appliances because (2.1) is dependent on $\mathbb{E}[P]$ and $\mathbb{E}[D]$. Notice that the chosen mean of the gamma distribution is not used in (2.1), but rather $\mathbb{E}[P]$ and $\mathbb{E}[D]$ from the set of generated appliances ψ. The larger the number of appliances in ψ, the closer these values will approximate the gamma distribution.

In this chapter, the validation of the models presented in Section 2.6 makes use of the same input parameters unless stated otherwise. Historical data from ComEd [67] (the utility company for Chicago, IL) is utilized as the $C_L(t)$ input for the $M_t/G/\infty$, $M_t/G/C$, and $M_t/G/C_t$ queueing load models. The period utilized from ComEd is the year of 2014,

having $b_{min} = 500$, $b_{max} = 5000$, and $k_C = 2$. The appliance set ψ is generated as illustrated in Fig. 2.7, having the gamma distribution parameters as power (W) $\mu = 500$ and $\sigma = 100$, and appliance duration (hour) $\mu = 0.5$ and $\sigma = 0.25$.

2.5.2 ZIP Appliance Model Input

The ZIP coefficients for residential, commercial, and industrial loads can be estimated using field and/or experimental data as shown in [62], [63]. In [63], 29 appliances have their ZIP parameters modeled; Table 2.2 presents three example appliances. Furthermore, the appliances also present the number of tested equipment, cutoff voltage, nominal voltage, and active and reactive power at nominal voltage. Thus, one of the inputs to generate the appliances that incorporates the ZIP polynomial parameters is the complete set of appliances from [63].

Table 2.2. Example of three ZIP appliances coefficients [63].

Equipment / component	Z_p	I_p	P_p	Z_q	I_q	P_q
Air Conditioner	1.17	-1.83	1.66	15.68	-27.15	12.47
Vacuum Cleaner	1.18	-0.38	0.2	4.1	-5.87	2.77
Television	0.11	-0.17	1.06	1.58	-1.72	1.14

In Table 2.2, the representation of ZIP polynomial parameters does not contain the contribution of each appliance. To randomly generate a set of customer appliances, as explained in Section 2.4.1, that maintain the aggregated behavior of the system the contributions of each appliance must be maintained. In [63], the contribution of each appliance is characterized. As the use of appliances changes throughout the year, Table 2.3 presents the summary of appliance contributions normalized by season. From [63], fall and spring are expected to possess the same behavior. The complete table with

the contribution of appliances for residential customer is presented in [63].

Table 2.3. Typical residential load composition and seasonality [63].

Equipment	Total No.	Rated Power (kW)	Duty Cycle	Season Factor			Total Power (kW)		
				Spring	Summer	Winter	Spring	Summer	Winter
TV	1	0.208	1	0.2	1	0.2	0.8	0.6	0.8
PC	1	0.119	1	0.2	0.2	0.2	0.6	0.6	0.6
Laptop Ch.	1	0.036	0.6	0.6	0.4	0.6	0.3	0.2	0.3
Minibar	1	0.091	0.5	1	1	1	1.1	1.1	1.1
Incandescent	1	0.087	1	0.2	0.2	0.2	0.4	0.4	0.5
CFL Bulb	1	0.026	1	0.4	0.3	0.6	0.3	0.2	0.3
Fan	1	0.163	1	0.1	0.1	0	0.4	0.6	0
Air Cond.	1	0.496	1	0.1	0.3	0	1.2	3.9	0
Total							5.02	7.46	3.57
Reported average peak power (weekdays)							5.05	7.40	3.56

The creation of the set of appliances follows the same process as presented in Section 2.4.1, but more characteristics than the appliance power (i.e., P_0 in the ZIP model) and time duration are needed. When an appliance arrives into the queue, another sample is made to define Q_0 and the ZIP polynomial coefficients. A weighted sample is made on the set of appliances given the overall contribution in the particular season being generated (i.e., from Table 2.3). From the random sample, the ZIP polynomial coefficients are used directly, and Q_0 of the appliance is defined to maintain its power factor. As the contribution from each appliance changes with the season, the queueing load models must change the set of arriving appliances ψ at every change of season. The results in this chapter use the data from [63], but this can easily be altered by the user for the study in

question with no loss of generality.

2.6 Validation of the Proposed Synthetic Queueing Model Behavior

2.6.1 Comparing the Three Synthetic Queueing Load Models

The three queueing load models are compared in detail for two days containing the minimum and peak hour from the ComEd region in 2014 (May 25, 2014, and July 22, 2014, respectively) in Fig. 2.11. Each queueing model was used to generate 1,000 customers. The top row of Fig. 2.11 presents the minimum load day, while the bottom row presents the peak day. The three columns from Fig. 2.11 represent the $M_t/G/\infty$, $M_t/G/C$, and $M_t/G/C_t$, respectively. In each plot, the dashed purple line is the user-defined expected load curve of a single customer $l(t)$, from (2.2) with $b_{min} = 500$ and $b_{max} = 5000$. The first to third quartiles are represented by the dark shaded blue area, and the minimum and maximum of the 1,000 customers are represented by the light shaded blue area. The first quartile of the 1,000 customer loads splits the lowest 25% of the customer load data from the highest 75%. Similarly, the third quartile splits the highest 25% of the load data, hence the first to third quartile range represents the active power region where 50% of the customers are located (please note that a specific customer may move in and out of this region from one time period to the next). Fig. 2.11 presents the mean load value of the 1,000 customers from the output of the queueing models as the solid black line, which follows the input reference curve $l(t)$, showing that independently generated loads for each customer have distinct behavior that on average follows the known reference load.

For the $M_t/G/C$ and $M_t/G/C_t$ queueing load models, the red dotted line in Fig. 2.11 represents the power capacity of a single home (i.e., C and C_t, respectively). The

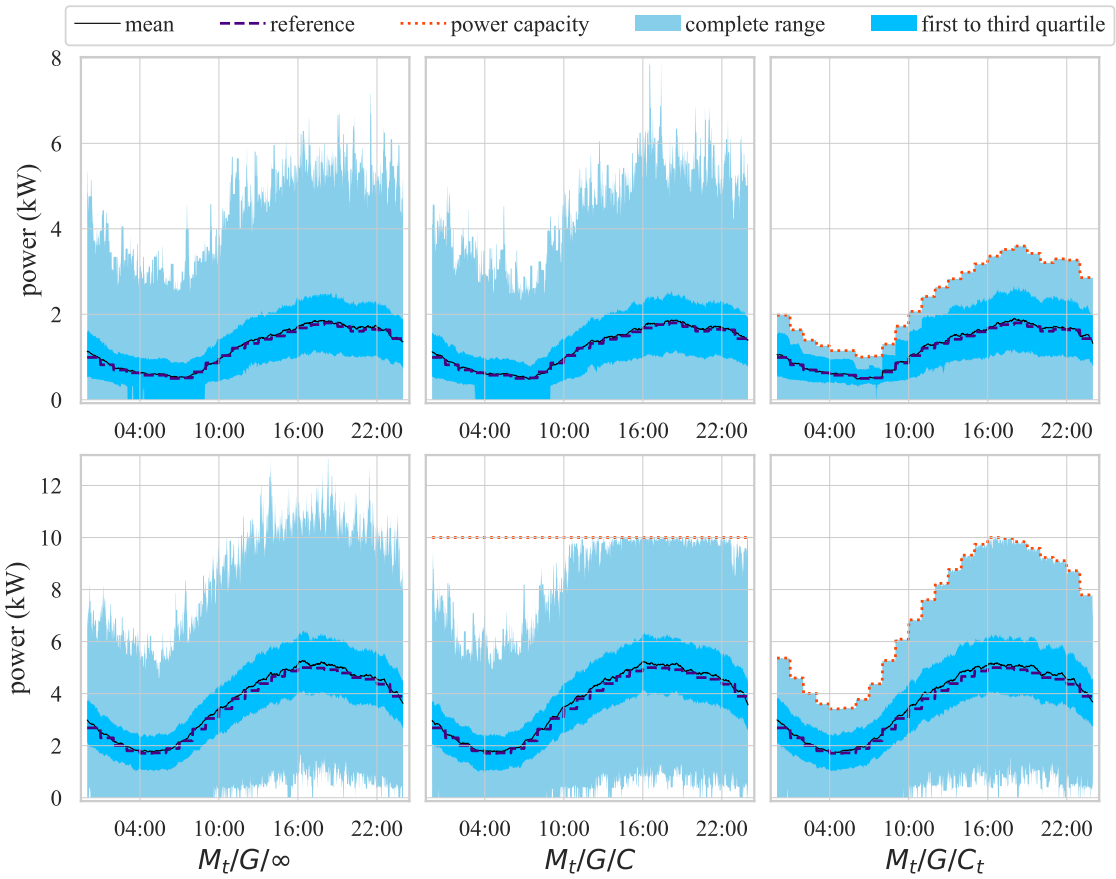

Figure 2.11. The output of the three synthetic queueing load models $M_t/G/\infty$, $M_t/G/C$, and $M_t/G/C_t$ — from left to right, respectively. The top row presents the day that contains the valley hour (i.e., May 25, 2014) and the bottom row presents the day with the peak hour (i.e., July 22, 2014). In each plot, the dashed purple line is the user-defined expected load curve of a single customer $l(t)$, and the red dotted line is the power capacity of the home (i.e., C and C_t from the queueing models — there is no limit in the $M_t/G/\infty$ queueing model). For each plot, 1,000 customers are synthetically created using the proposed queueing load models. The mean of the 1,000 customers is the solid black line, the first to third quartiles are represented by the dark shaded blue area, and the minimum and maximum are represented by the light shaded blue area. Each of the three models for both simulated days average to the expected load curve, $l(t)$, and hence will aggregate to the known system load curve $C_L(t)$.

power capacity C and C_t were computed with (3) and (4), respectively, having $k_C = 2$. The power capacity for the $M_t/G/C$ queue does not appear in the minimum power day because $C =$10 kW is out of the y-axis range. The $M_t/G/\infty$ is unbounded, thus not having a power capacity. In Fig. 2.11, it can be observed that only $M_t/G/C_t$ is affected by the power capacity for the minimum load day (i.e., the top row). Because the $M_t/G/C_t$ active power range for the minimum load day is considerably reduced, there is a low likelihood for the arrival of appliances to be served during the low-capacity time period. This characteristic is elaborated in Section 2.6.2.

The mean of the 1,000 customers follows the reference load $l(t)$, showing that the average behavior of the independently generated customer loads is known, as presented in Fig. 2.11. This characteristic can be extrapolated to generate any given number of independent customers N to a known aggregated load curve $C_L^*(t)$ with,

$$C_L^*(t) = \sum_{n=0}^{N-1} l_n(t). \tag{2.11}$$

For a large number of N, $C_L^*(t)$ can be used for large-scale smart city sized energy management studies. In Fig. 2.12, the aggregate behavior of all 1,000 customers is shown. The output of each of the three models is compared to the sum of the reference curves, $C_L^*(t)$. As each of the three models closely approximates $C_L^*(t)$, the proposed models are validated for use in large-scale smart city sized energy management studies. This demonstrates that even with the differences in the individual output of the queueing models for every customer — i.e., random process and independently generated using (2.11) — (i) the average behavior of a customer is known, (ii) the aggregate

behavior of all customers represents a known system curve, and (iii) each customer is independent with a unique load curve.

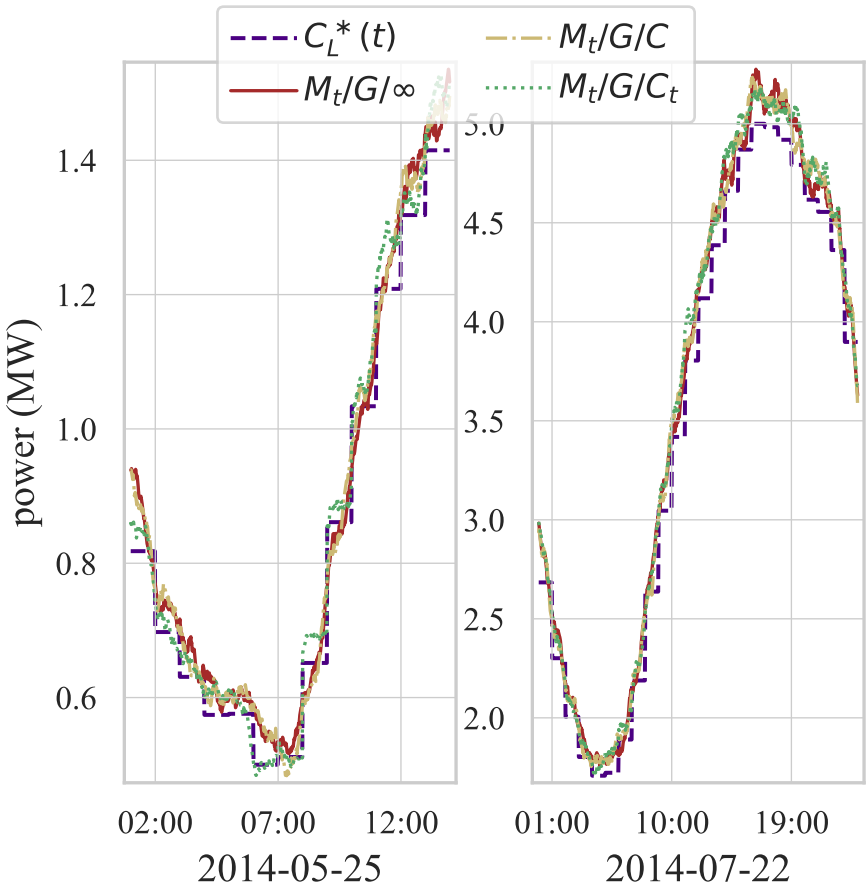

Figure 2.12. The known aggregated load curve $C_L^*(t)$ compared to the summation of 1,000 customer outputs of the three synthetic queueing load models. The independently generated customer output of each of the three queueing load models are similar to the behavior of the known aggregated load curve, validating the methods.

2.6.2 Impact of Queueing Model Parameter Choice

The choice of input parameters impacts the output of the queueing load models. Section 2.3.2 demonstrates that $\mathbb{E}[P]$ and $\mathbb{E}[D]$ have a direct impact on $\lambda(t)$ (i.e., the arrival rate of appliances) according to (2.1). Thus, the gamma distribution for generating the appliances, as presented in Section 2.4.1, has an impact on $\lambda(t)$. Additionally, the

selection of appliance parameters impacts the size of the output list of appliances generated by the queueing models. A small value of $\mathbb{E}[P]$ and $\mathbb{E}[D]$ requires many more appliances to arrive for the same reference values, and vice versa.

For the same input parameters used in Section 2.6.1, if b_{min} is changed to 100 W, the queueing load model behaves differently for the minimum hour day as shown in Fig. 2.13. For the $M_t/G/\infty$ queueing model, the choice of b_{min} impacts the arrival rate of appliances. Based on (2.1) and (2.2), with an $l(t) = 100$ W at the minimum time and $\mathbb{E}[P] \approx 500$ W (depending on the sampling of the gamma distribution for ψ), the arrival rate $\lambda(t)$ significantly reduces (i.e., the period between arrivals is increased), as shown in the left plot of Fig. 2.13.

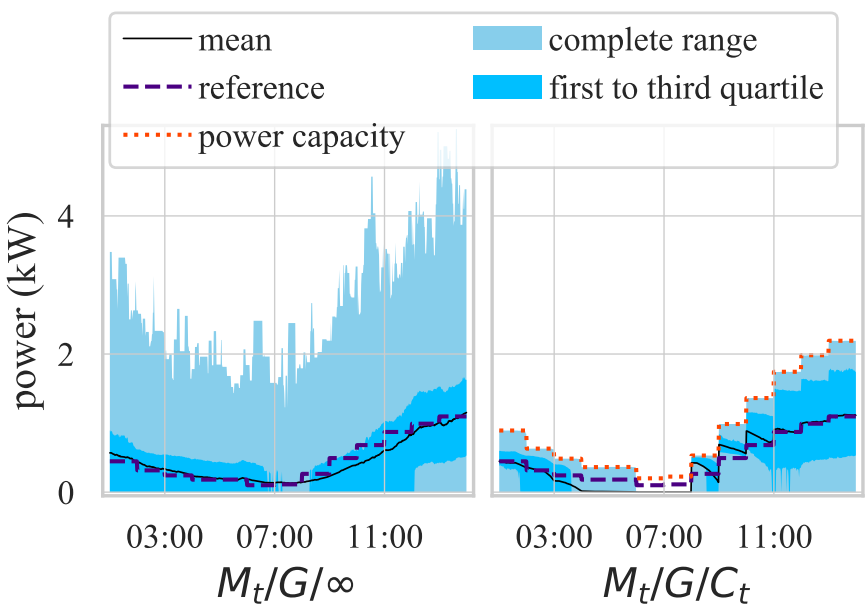

Figure 2.13. The impact of lowering b_{min} to 100 W on the $M_t/G/\infty$ and $M_t/G/C_t$ queueing load models. Each queueing model generated 1,000 customers with: $b_{min} = 100$, $b_{max} = 5000$, $k_C = 2$, appliance power (W) $\mu = 500$ and $\sigma = 100$, and appliance duration (hour) $\mu = 0.5$ and $\sigma = 0.25$. Due to the input parameters chosen, the output behavior of the models may be unstable (e.g., 7:00 in the $M_t/G/C_t$ queueing model).

For the $M_t/G/C_t$ queueing model, because $\mathbb{E}[P]$ of the appliance set is ~ 5 times larger than the reference curve at hour 7:00, the probability of a customer having an appliance smaller or equal to 100 W given the gamma distribution parameters are $\mu = 500$ and $\sigma = 100$ is approximately zero. Thus, considering the probability of such appliances arriving at a time in which they could be scheduled is negligible. The first to third quartile range becomes non existent for a small period of time. The $M_t/G/C_t$ possesses a power capacity of 200 W for the hour, thus for all 1,000 customers no appliances are present that can be served, making the load equal to zero. Because the service policy is first come first served, the arrived appliances are started as soon as possible in the subsequent hours, the mean value tends towards the power capacity and it takes a few hours to return to steady-state behavior.

Maintaining the same parameters used in Section 2.6.1 except with b_{min} still equal to 100 W, we significantly reduced $\mathbb{E}[P]$ and $\mathbb{E}[D]$ (i.e., appliance power (W) $\mu = 10$ and $\sigma = 2$, and appliance duration (hour) $\mu = 0.2$ and $\sigma = 0.1$) to study the impact of small appliances in the queueing models, shown in Fig. 2.14. The mean, complete range, and first to third quartile range behave as expected for $M_t/G/\infty$ and $M_t/G/C_t$, however the generated ranges are considerably smaller, i.e., significantly reducing the difference between the individual customers.

In summary, for the output of the queueing models to be stable, it is desirable that multiple appliances can arrive and be served at any moment (i.e., Fig. 2.13 at 7:00). At the same time, if a large number of small appliances arrive and are served at the same time, the differences between the customers is greatly reduced as in Fig. 2.14. The choice of the gamma distribution parameters for generating the set of appliances ψ, size of the set of

appliances $|\psi|$, b_{min}, b_{max}, and queueing model power capacity has an impact on the behavior of the queueing load models, and should be considered by the user when choosing the input parameters for future studies using these models. These choices will vary depending on the specific user problem of interest (e.g., HEMS, geographical area, and other relevant input for a specific study).

The input variables b_{min} and b_{max} appear to be independent, but if it is desirable for the aggregated behavior of the customers to approximate $C_L(t)$, b_{min} and b_{max} are not independent. For the aggregated load of customers to behave as $C_L(t)$, the ratio of $max(C_L)$ and $min(C_L)$ needs to be maintained in b_{max} and b_{min}, and b_{max} must be an integer multiple of $max(C_L)$ as the number of customers is an integer. Thus, with a known number of customers and b_{max}, it is known that $b_{min} = {}^{min(C_L)}/_{max(C_L)} b_{max}$.

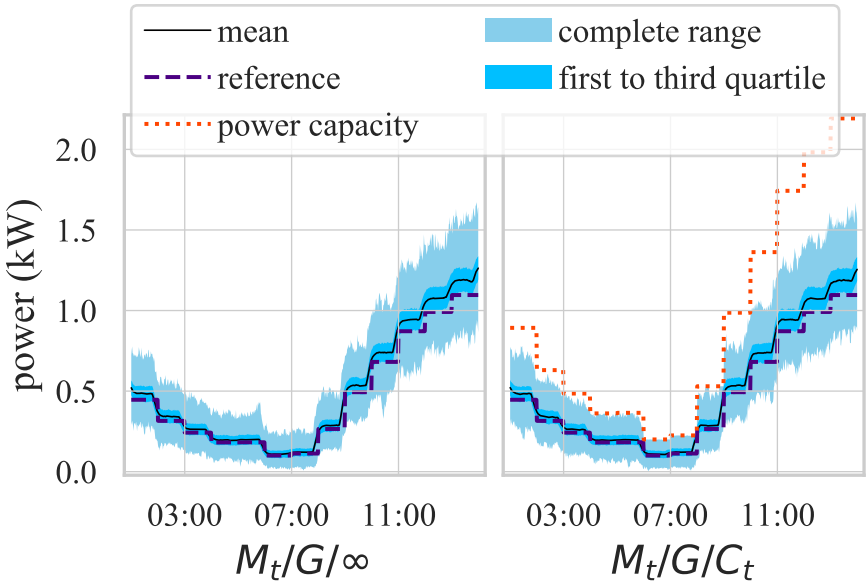

Figure 2.14. Impact of small appliances on the $M_t/G/\infty$ and $M_t/G/C_t$ queueing load models. Each queueing model generated 1,000 customers with: $b_{min} = 100$, $b_{max} = 5000$, $k_C = 2$, appliance power (W) $\mu = 10$ and $\sigma = 2$, and appliance duration (hour) $\mu = 0.2$ and $\sigma = 0.1$. The impact of smaller appliances reduces the range of generated customer load curves.

2.7 Queueing Model Discussion

This subsection presents a comparative discussion of the results presented in Subsections 5.1 and 5.2, summarizing the behavior of the queueing models and their interaction with the appliance models. Residential customers naturally follow a predictable behavior when aggregated, thus the average load approximates to the reference (i.e., scaled system load curve). However, an individual residential customers is expected to possess large variability through time (i.e., a single load in a house greatly impacts its power consumption, but the same load is quite small compared to the system curve). The ranges of generated values of the thousand residential customers is used to demonstrate the behavior of the generated loads on the valley- and peak-hour days in Table 2.4. The mean of the generated range from the first to the third quartile and the complete range for the generated synthetic residential load models are presented, showing the impact of the load model and choice of parameters in the model output. Cases 1, 2, and 3 in Table 2.4 refer to the presented cases in Subsections 5.1 and 5.2. Case 1 is presented in Subsection 5.1 has load scaling parameters $k_C = 2$, $b_{min} = 500$ (W), and $b_{max} = 5$ (kW); appliance duration $\mu = 0.5$ (hour), $\sigma = 0.25$ (hour); and appliance power $\mu = 500$ (W), and $\sigma = 100$ (W). Case 2 presented in Subsection 5.2 has load scaling parameters $k_C = 2$, $b_{max} = 5$ (kW), and $b_{min} = 100$ (W); and the same appliance parameters as Case 1. Finally, Case 3 presented in Subsection 5.2 has the same load scaling parameters as Case 2; appliance duration $\mu = 0.2$ (hour), $\sigma = 0.1$ (hour); and appliance power $\mu = 10$ (W), and $\sigma = 2$ (W). In Table 2.4, the queueing load model time-varying server availability and appliance parameters have a similar impact on the

complete ranges in Case 1, however not on the first to third quartile. Reducing the appliance size will impact the complete and first to third quartile ranges in Case 3. Note we only present the valley-hour day for Cases 2 and 3 because these were used to highlight the impact of appliance size and scale parameters of the reference load on the model output in Subsection 5.2, hence there is no data presented for the peak-hour day.

Table 2.4. Impact of parameters on range of generated queueing model outputs.

Case	Model	Mean range (kW)			
		Valley hour day		Peak hour day	
		First to third quartile	Complete	First to third quartile	Complete
1	$M_t/G/\infty$	1.09	4.44	1.83	8.37
	$M_t/G/C$	1.09	4.50	1.83	8.11
	$M_t/G/C_t$	1.03	2.34	1.86	6.64
2	$M_t/G/\infty$	0.62	2.74		
	$M_t/G/C_t$	0.45	0.89		
3	$M_t/G/\infty$	0.09	0.44		
	$M_t/G/C_t$	0.09	0.43		

2.7.1 ZIP Appliances with $M_t/G/C_t$

As seen in Section 2.4.3, the appliance load consumption is affected by the local voltage. A common appliance model that considers this behavior is the ZIP load model. ZIP load model characteristics were added to the appliances as shown in Section 2.5.2 with parameters from [63]. To demonstrate the behavior of the appliances with the ZIP parameters, 50 homes were generated making use of the $M_t/G/C_t$ queueing load model with the same parameters from Section 2.6.1. Appliances arrive into the queue with their individual characteristics (i.e., d_i, P_0, Q_0, Z_p, I_p, P_p, Z_q, I_q, P_q). Thus, to visualize the behavior of the home, the appliances operating simultaneously are aggregated as

$$\begin{cases} A_{P_0} = \sum_i P_{0_i} \\ \\ A_{(.)_p} = \sum_i \frac{P_{0_i}}{A_{P_0}} (.)_{p_i} \end{cases} \qquad \begin{cases} A_{Q_0} = \sum_i Q_{0_i} \\ \\ A_{(.)_q} = \sum_i \frac{Q_{0_i}}{A_{Q_0}} (.)_{q_i} \end{cases} \tag{2.12}$$

where $(.) = \{Z, I, P\}$. Although 50 homes were visualized and analyzed, the output of each home was similar, so a single home was chosen to illustrate the average characteristics of the ZIP parameters for the queueing load models. Fig. 2.15 presents the active and reactive power of the selected home for July 22, 2014, using the $M_t/G/C_t$ queueing model assuming a variation in voltage at the point of connection to the electric power system. The top plot in Fig. 2.15 presents the reference curve for the queueing model (black dashed line), active power of the home at nominal voltage (solid black line), queueing model power capacity (red dotted line), and the range the active power (green shaded area) assuming a voltage range from 0.95 p.u. to 1.05 p.u. [68]. The bottom plot in Fig. 2.15 presents the reactive power at nominal voltage (solid black line) and the range of reactive power for the same variation in voltage (blue shaded area). The reactive power of appliances are dependent on the appliance model (i.e., user-defined ZIP characteristics and power factor), not on the queueing model which only governs the arrival of appliances based on the active power (hence the active power reference curve and a power capacity for $M_t/G/C_t$).

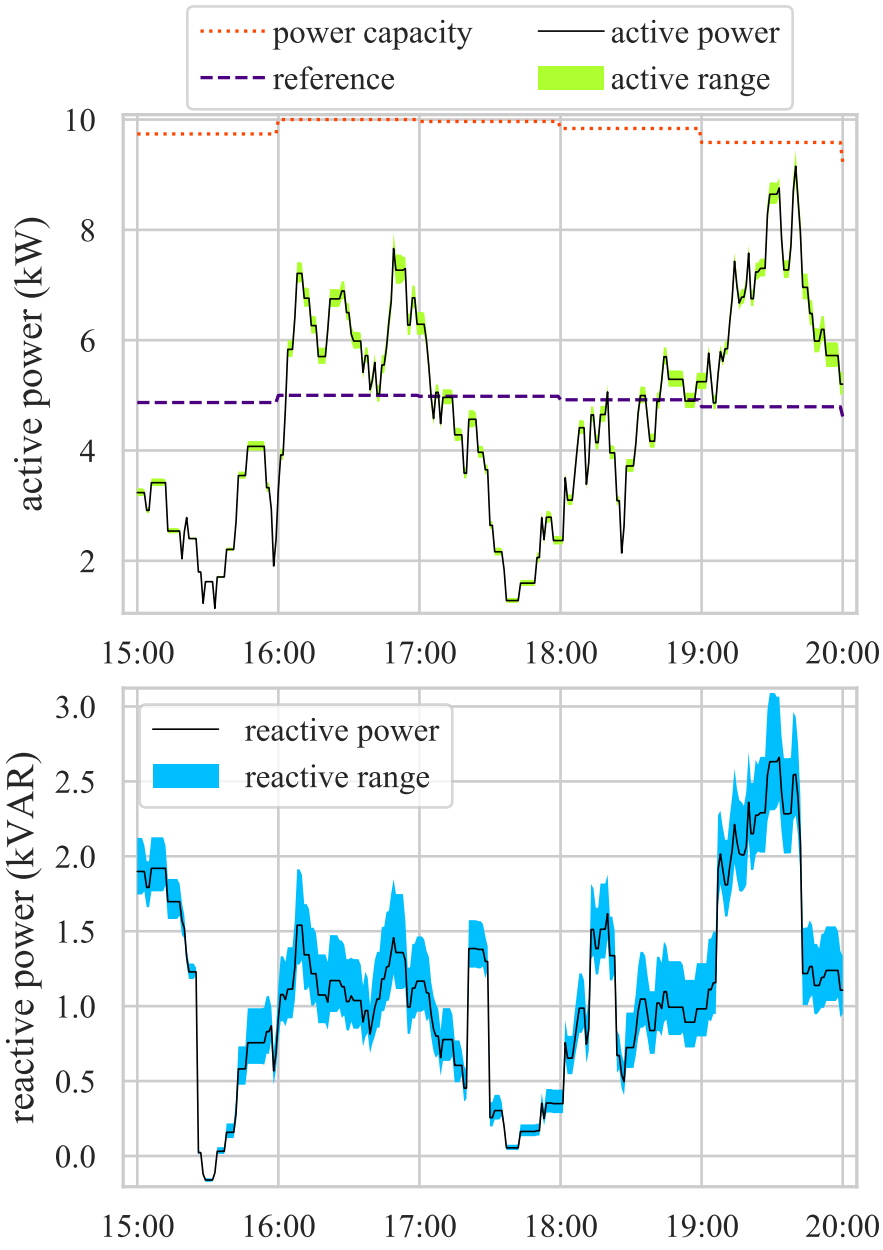

Figure 2.15. Active and reactive power for a single home using the $M_t/G/C_t$ queueing load model with ZIP appliances. The areas for active and reactive power consider a voltage range from 0.95 p.u. to 1.05 p.u. Note that there is no capacity or reference input for the reactive power curve, rather these come directly from the ZIP characteristics and power factor of the appliance set.

2.7.2 Computational Performance of the Synthetic Queueing Load Models

The synthetic queueing load models are intended for power system studies for large-scale smart city-sized assessment, which contain thousands to millions of electric customers. Thus, any synthetic load generation approach for residential city-size studies must be computationally efficient. As shown in the previous sections, the presented synthetic queueing load models are capable of being independently generated for each customer, thus they can be created in parallel with minimal interprocess communication. The South Dakota State University Roaring Thunder High-Performance Computing Cluster was used to generate customer loads in parallel to measure the scalability of the proposed methods. Roaring Thunder possesses 56 compute nodes, each with dual socket Intel Skylake 6148 CPUs, 40 CPU cores (20 cores per socket), 192 GB RAM, and 240 GB SSD local storage. The code was developed in Python [69] and makes use of the `SCOOP` (Scalable COncurrent Operations in Python) package to spread the customer load generation work across the available compute resources in single and/or multiple compute nodes. `SCOOP` maintains a master processor to manage and monitor the work in a worker-pool model, thus when a process finishes generating and saving the data for one customer it is assigned another customer from the work pool until all customer loads have been generated. The data is saved using HDF5 file format and written using the cluster parallel file system that enables concurrent read/write to disk.

The scalability of the queueing load models are presented in Fig. 2.16 for generating 100 annual customer loads for the year 2014 using the different queueing load models averaged over four trials. The y-axis presents the speedup normalized to the

runtime for each algorithm using 26 processing elements and compared to the ideal speedup (i.e., linear speedup with each additional processing element). The relative performance of the methods differs due to the internal loop of $M_t/G/C$ and $M_t/G/C_t$ in Fig. 6 for shifting the arriving appliances depending on the available power capacity. The scalability of the three queueing models behave similarly through 76 processors, but the three methods deviate in performance with 101 processors (i.e., 1 processor per customer and 1 master process). Because of the internal loop, if each process only generates one customer load, the single slowest customer load sets the entire runtime and negatively impacts performance compared to $M_t/G/\infty$. The time requirements of the internal loop of $M_t/G/C_t$ are more demanding (i.e., stricter power capacity), thus having a larger impact on relative performance and scalability.

The absolute times required to generate the data are presented in Table 2.5, which are averaged over four trials for each case. The individual customer time is computed by multiplying the total time by the number of processing elements and dividing the result by the number of generated customers. Notice that the average time for an individual customer has little variation with the same queueing load model across differing numbers of processing elements. As the time to create a one-year dataset per customer is relatively low and the data is able to be generated independently in parallel with near-linear speedup, the synthetic load models are promising to generate Smart City-sized datasets that aggregate to a known system load curve.

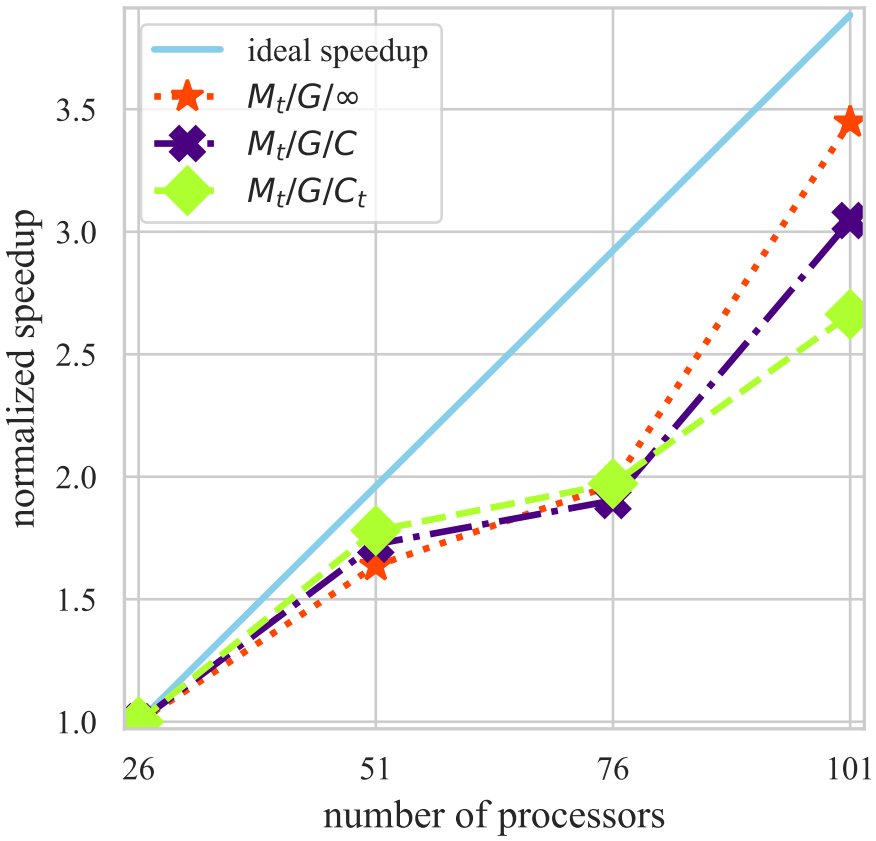

Figure 2.16. Normalized speedup to 26 processors of the synthetic queueing load models compared to the ideal parallel speedup using the Roaring Thunder Cluster. The queueing load models were used to generate 100 customers for the entire year of 2014 averaged over four trials.

Table 2.5. Absolute time and average time per customer for the cases presented in Fig. 2.16.

Number of Processing Elements		26	51	76	101
$M_t/G/\infty$	total	37.53	22.93	19.08	10.89
(minutes)	individual	9.76	11.69	14.50	11.00
$M_t/G/C$	total	74.70	43.31	39.25	24.53
(minutes)	individual	19.42	22.09	29.83	24.77
$M_t/G/C_t$	total	88.64	49.79	44.95	33.29
(minutes)	individual	23.05	25.39	34.16	33.62

2.8 Conclusions

This chapter proposed an approach for modeling individual residential customers and their individual electric assets using time-varying queueing models. The queueing load models presented in this chapter address the challenges of unavailability and proprietary customer data by using only public available aggregated load data for a region, allowing researchers to replicate results in many studies and compare their methods to the state-of-the-art. In addition, by aggregating to a known system load curve, the economic and technical impacts of new research methods can be better evaluated. The model assumes that the aggregated distribution system behavior is known while including the stochastic nature of individual customers and their electric assets (i.e., combined top-down bottom-up modeling). The models are general enough to incorporate other characteristics, such as non-arriving portions of customer loads (e.g., HVAC), voltage dependencies (e.g., ZIP polynomial coefficients), scheduling characteristics, and more depending on the needs of the individual researcher. The models were validated by visualizing the differences in output between a thousand customers and by their aggregated load that characterizes and

follows a known system curve. As the proposed models were shown to scale in a near-linear fashion and individual customer loads can be independently generated, the methods can be used in large-scale demand side management studies (e.g., Smart City demand response) with individual customer load data that maintains the time-varying characteristics of an actual power system region. The future work is to expand the appliance models to include HVAC and other characteristics such as frequency dependencies.

CHAPTER 3 A Real Distribution System Test Case with One-Year Appliance-Level Load Data Derived from Utility Smart Meters for Transactive Energy Studies

3.1 Overview

The Iowa State distribution system test case is a real 240-node distribution system from the Midwest region of the U.S. in OpenDSS with one-year smart meter node-level load data for 2017. This article derives a synthetic appliance-level residential load using the queueing load model for 1,120 homes on the Iowa State distribution system test case for use in distributed energy management studies. The expanded Midwest 240-Node test case provides granular-level information for all homes in the distribution system (i.e., individual appliances that constitute the home load), and the aggregate of all customer load emulates the real smart meter data. The one-year synthetic appliance data has a mean absolute percentage error of 2.58% compared to the smart meter data. The Midwest 240-Node test case is validated and provided in open-source OpenDSS and GridLAB-D models to enable transactive energy studies with active electric end-users.

3.2 Introduction

Given the increased variability in generation from the increased participation of non-dispatchable resources (e.g., wind and solar), there is a need for increased operational flexibility for the future environmentally friendly, economical, and secure power system [70]–[72]. Demand response (DR) is one such source of flexibility, and according to [53], is a main component of the smart grid. DR encourages consumers to change their

demand concerning power system conditions — a generalized form of DR that considers or coordinates both supply and demand is commonly referred to as transactive energy [73]–[75]. Residential loads represent approximately 38% of the total energy consumption in the U.S. [10]. Residential DR can provide the needed operational flexibility, and the major resulting benefits are: (a) participant financial benefits; (b) market-wide financial benefits; (c) reliability benefits; and (d) market performance benefits [11], [12]. Residential DR makes system-wide changes that require tens of thousands of buildings, each with many individual electric energy devices, to be controlled [14].

There is a missing link in the research community between the availability of aggregate power system demand, the individual customer demand that composes it, and the location of such demand on distribution system networks. Multiple home energy management system studies neglect their impact on the power system [76]–[79]. In [76], residential DR optimization models for scheduling individual customer appliances is presented. The paper makes use of actual real-time pricing information from an Illinois power company. The work in [77] is similar to [76], but it aggregates residential customers in a residential community to perform the optimization. The aggregation of customers in [78] is also similar to [77], but differs by focusing on multi-objective optimization tradeoffs between customer financial benefits and customer discomfort. In contrast to the multi-objective optimization from [78], a hierarchical controller framework bidding strategy for demand reduction events considering the consumer preferences is presented in [79]. Differently than [76]–[78], a considerable effort is presented to evaluate the change in locational marginal price given the change in demand [79]. However, the

studies presented in [76]–[79] do not have a residential customer demand that aggregates to a known region of the power system. A combined power system and home energy management system test case must emulate the behavior of a real distribution system with individual customer loads that aggregate to the known system load. Existing distribution system test cases do not have real time-series load data, except for the IEEE European LV [80], but this is only for a single day. By linking individual customer loads to the system load, calculation and analysis of system-level impacts of residential DR is enabled. Such analysis allows studies to more accurately demonstrate the flexibility and impacts of DR on power system operations (e.g, electricity markets, reduction in renewable energy curtailment).

The availability of a power system test case that accurately represents the behavior of a real system is considered an enabling development for the design and analysis of new and scalable approaches for the integration of distributed energy resources in [81]. The lack of U.S. electric distribution system test cases led the authors from [82] to create synthetic distribution test systems using street maps, equipment catalogs, and building expected behavior. The distribution test systems are intended for testing algorithms with considerable distributed resources present in distribution systems. The paper [81] continues the work of [83]–[87] for the creation and validation of synthetic transmission systems and [88]–[91] for the creation and validation of synthetic distribution systems with access to utility data. The work presented in [81] focuses on the U.S. distribution systems, creating and validating distribution synthetic systems of up to 10 million electric nodes. The authors from [81] are researchers at National Renewable Energy Laboratory (NREL) and have access to real utility data, which is considered Critical Energy/Electric

Infrastructure Information (CEII) and is not available to the general research community. Their approach enables the validation of flexible distribution synthetic systems, but the methods still require unavailable CEII. Additionally, the work in [81] validates three large scale synthetic test systems, with statistical quantification to infer how realistic the *networks* are compared to real data, where the work in this chapter is to create time-series synthetic load *data* at the granular-level from aggregated smart meter data on a real distribution network.

The Iowa State distribution system test in [92] made available a real distribution network from the U.S. Midwest region with one-year smart meter node-level load data for 2017. This unique test case combines a real utility distribution system network model with corresponding field measurements that are publicly available. To maintain individual consumer privacy, the available data is aggregated to node-level and is provided in an hourly resolution. In this chapter, the nodal load data is first divided into 1,120 homes across 193 load nodes over three feeders. The home data is further divided into appliance-level data using the queuing load model from [54]. The one-year mean absolute percentage error between the real smart meter data and granular-level synthetic data generated for the Midwest 240-Node is 2.58%. The main contributions of this chapter are

(a) the generation of synthetic granular-level residential load data from aggregated nodal smart meter data; and

(b) the development and validation of the open-source Midwest 240-Node transactive energy test case.

The open-source Midwest 240-Node transactive energy test case is provided in

both OpenDSS and GridLAB-D, and was validated with a maximum voltage magnitude error less than $10^{-3}\%$. This test case will enable researchers to perform granular-level smart grid and transactive energy studies, and measure the system-level impacts.

The remainder of this chapter is organized as follows: Section 3.3 is an overview of the publicly available Iowa State distribution system test case. The queueing load model for generating the granular-level synthetic load data from the node-level smart meter data is described in Section 3.4. In Section 3.5, the Midwest 240-Node system is validated in regards to both the load mismatch (real nodal smart meter data vs. synthetic granular-level data) and power flow impact (OpenDSS vs. GridLAB-D). Section 3.6 presents the main conclusions of this study. Finally, Appendix B details the conditioning of missing smart meter data from the published Iowa State distribution system test case for use in the Midwest 240-Node test case.

3.3 Describing the test system

Power system test cases, including distribution test systems, are derived from the general characteristics of real networks. Dr. Zhaoyu Wang from Iowa State University received permission from a utility partner to make publicly available a real distribution network from the Midwest U.S. [92] in OpenDSS format. The test system has 240 primary network nodes and 23 miles of primary feeder conductor. The real distribution network will be referred to as the Iowa State distribution system test case. In addition to the real network data, one-year smart meter measurements at the node-level were also provided.

The Iowa State distribution system test case is presented in Fig. 3.1 as a radial distribution system consisting of three feeders [92]. The feeders are labeled as S, M, and L

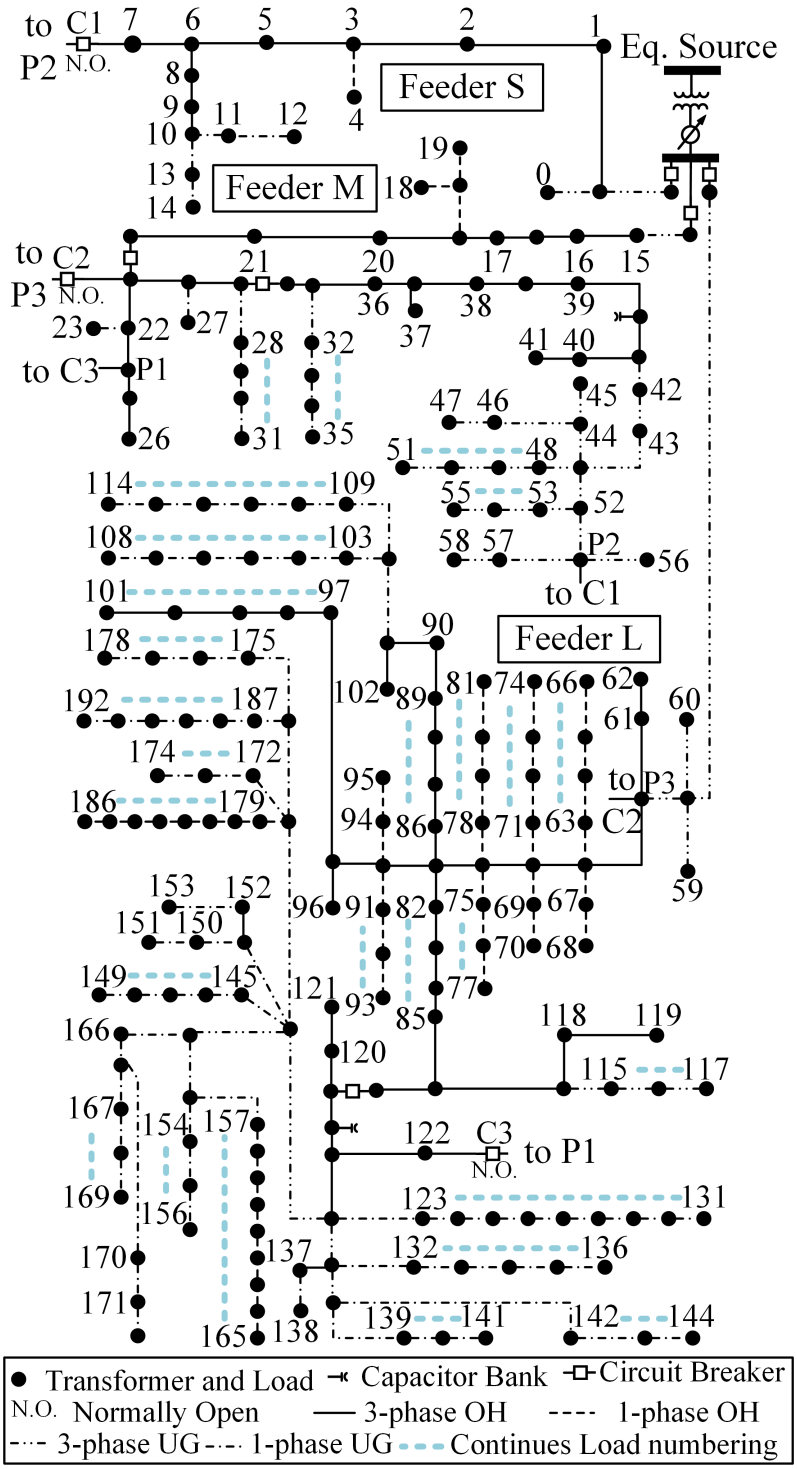

Figure 3.1. One line diagram of the test system. Adapted from [92].

referring to the relative size of the feeders as small, medium, and large, respectively. A 10 MVA delta-wye step-down 69/13.8 kV substation transformer supplies power for the three feeders. The substation transformer has a tap-changer mechanism that consists of three independent single-phase tap changers. Feeders M and L have shunt capacitor banks for voltage regulation. The utility has a strategy to switch on capacitor banks in normal operation to provide reactive power support. Iowa State distribution system test case has nine circuit breakers at the illustrated locations in Fig. 3.1 that are used for protection and reconfiguration. Six of the circuit breakers are normally closed, and three are normally open. All standard electric components in the Iowa State distribution system test case are modeled, such as overhead lines, underground cables, substation transformers with load tap changers, line switches, capacitor banks, and secondary distribution transformers.

The Iowa State distribution system test case has 1,120 homes, each with an installed smart meter [92]. There are 193 system load nodes with 15 on Feeder S, 44 on Feeder M, and 134 on Feeder L, each with a unique numeric number from 0 to 192. The assigned number for the load node follows the order from the provided files in [93] and is read in the following order: S, M, and L. The homes are connected to the primary network nodes via secondary distribution transformers, demonstrated in Fig. 3.1. The load data is measured using smart meters for the year 2017 in an hourly resolution (in kWh) by approximating the hourly energy consumption under the assumption that the customer demand is constant in each one-hour time interval [92]. To model reactive power for the load nodes, a power factor is randomly selected in the range of 0.9–0.95 [93]. The power factor and reactive power of each customer is calculated and aggregated for the customers in the same load node.

Although 1,120 homes are known to be on the network, the provided load data in [92] is aggregated at the node level to protect the privacy of individual customers. Additionally, it is unknown if any customers have distributed generation, such as solar photovoltaic.

3.4 Generating Granular-Level Synthetic Load Data

3.4.1 Overview

The real customer demand from smart meter measurements [93] are the aggregation of customers at a given load node with hourly resolution. Power system studies such as home energy management systems, distributed energy management, DR, and transactive energy require high-resolution individual customer load (i.e., the knowledge of appliances that compose the demand of each customer/home). To utilize the data provided in the Iowa State distribution system test case for such studies and taking advantage of the real customer demand data, in this section

(a) the provided nodal load data is analyzed, and time periods with erroneous smart meter data are statistically replaced (more information in Appendix B);

(b) the 1,120 homes are divided to the load nodes based on energy consumption;

(c) the appliance model parameters are described; and

(d) granular-level synthetic load data is generated using the queueing load model from [54] to create the Midwest 240-Node test case.

3.4.2 Evaluating the Nodal Load

As mentioned in the previous section, the Iowa State distribution system test case provides one-year nodal load data based on smart meter measurements for the 1,120 homes. After analyzing the provided data from [93], small portions (i.e., less than 0.21%) were suspected of being erroneous. Specifically,

(a) from hours 3,500–3,800 at load nodes 41, 154, 158, 162, and 163; and

(b) from hours 6,400–6,700 at load nodes 134, 140, 142, 149, 152, 180, and 183.

The data at these load nodes during these time periods were replaced with a statistical representation using a generalized linear model. The complete analysis of the data and explanation of the generalized linear model are presented in the Appendix B.

3.4.3 Parameters for the Appliance Model

The granular-level load data per home is assumed to be composed of individual appliances, shown as the yellow squares in Fig. 3.2. Each home has a set of appliances that is defined by average power rating and duration of each appliance. It was shown in [54], [55] that appliances with any time-varying power draw can be used (i.e., a researcher can use real appliance datasets if available), but in this work for generality it is assumed that the appliances are randomly generated with a constant power draw over a fixed duration. The appliance set for each home is generated by sampling two distinct gamma distributions, one for the power rating of the appliances and the other for the duration of the appliances. Gamma distributions are continuous probability distributions in the positive real number set defined by two parameters (i.e., shape k and scale θ). The

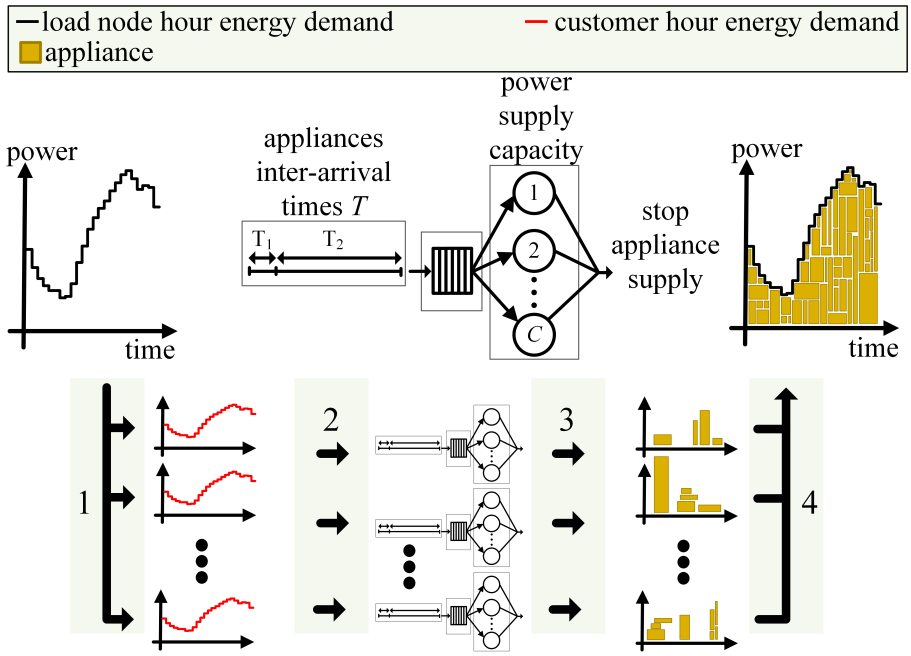

Figure 3.2. Summary of the synthetic queueing load model used to generate the granular-level data for each home on the Midwest 240-Node test case . At each load node, the node-level load is split into a per-home load reference curve, denoted by "1." Each home independently generates the granular-level appliance data using the synthetic queueing load model (2 and 3). Lastly, denoted by "4," the aggregated load from appliances from all homes on the load node will statistically represent the node-level reference curve.

mean of a gamma distribution is $\mathbb{E}[X] = k\theta$, and the variance is $Var(X) = k\theta^2$. Thus, by defining the mean μ and standard deviation σ, the gamma parameters k and θ are computed with $k = \mu^2/\sigma^2$ and $\theta = \sigma^2/\mu$. For the granular-level data, the appliance set ψ is generated with gamma distribution parameters as power (W) $\mu = 500$ and $\sigma = 100$, and appliance duration (hour) $\mu = 0.5$ and $\sigma = 0.25$, as utilized in [54].

3.4.4 Synthetic Queueing Load Model

Queueing models are defined by the probability distribution of inter-arrival times T (i.e., appliance inter-arrival times), probability distribution of service times X, number of servers C (i.e., power supply capacity), queue capacity, size of the population, and a service discipline. Furthermore, the characteristics can be constant or time-dependent (e.g., inter-arrival times as a function of time, as illustrated in Fig. 3.2). The queueing load models in [54] make three assumptions: the queue length is infinite (i.e., no loss of appliances arriving at the system); the population is infinite (i.e., arrival process is not dependent on the appliances currently present in the system); and the service policy is first come first served. Given the assumptions, the queueing load models are described with the simplified Kendall notation, i.e., $T/X/C$.

The synthetic queueing load model combines a top-down bottom-up approach. Having the expected load of a customer ($l(t)$) as the input for computing statistical time varying arrival rate of appliances for a customer. The appliance are modeled as generic blocks of energy as in [54]. Fig. 3.2 presents a summary of the process for generating the synthetic queueing load for one of the load nodes. Thus, as the output having the appliances that constitute the demand for each one of the customers in each load node for

the Midwest 240-Node test case . The numbers from 1 to 4 with the arrows on a light blue background are the steps for generating the load. Number 1 is making the reference curve for all the customers that constitute the total load node demand. Number 2 running the queueing load model, which are independent from each other. Number 3 the output of the arrived appliances in the queue for each customer for the generated period. Number 4 the aggregated arriving appliance load for all the customers approximates the original load node demand.

Three queueing load models are presented in [54], i.e., the $M_t/G/\infty$, $M_t/G/C$, and $M_t/G/C_t$. The models have a time-varying probability distribution of inter-arrival times (M_t) and the probability distribution of service times is general (G). However, each queueing load model has a distinct power supply capacity, being infinite (∞), constant (C), and time-varying (C_t) respectively. Given the natural random characteristic of the queueing models with the probability distribution of inter-arrival times it is expected that the larger the number of customers being generated the smaller the deviation from the reference curve for a given load node. The formulation for the queueing load models and further explanation are presented in [54]. Loads that want to be treated separately from the queueing arriving appliances can be simply subtracted from $l(t)$, as shown in [54].

3.4.5 Parameters for the Queueing Load Model

According to U.S. Energy Information Administration 2015 Residential Energy Consumption Survey [94], homes from the Midwest region have an average yearly consumption of 9,567 kWh. Assuming the yearly consumption divided by the expected yearly energy consumption the number of residential customers is 1,367. Being

considerably different from the 1,120 homes. By excluding the two regions of strange behavior and divided by the adjusted yearly consumption (i.e., average yearly consumption is linearly reduced by the reduction in the period of the year being considered) the number of residential customers is 1,378. It is known that the consumption is climate dependent and it is present in the data, as demonstrated in [92]. Using the month of May for selecting the number of homes without considering the periods of strange behavior (i.e., from May 1 to May 25) results in the number of homes of 1,187.

In Section 3.4.4 the queue load model is presented and with the demonstration from [54] better approximate the desired load curve with height reference energy values and large aggregation of customers. The height reference energy values increase the probability of multiple appliances being served, which is desirable. The large aggregation of customers enables the deviation from the reference energy values of individual customers to be minimized. The desirable number of customers is 1,120 allowing some control over the reference energy curve.

Fig. 3.3 presents the algorithm to remove the 67 extra homes. The algorithm requires the information of the node load $load_n$, number of homes by node NH_n, and average yearly consumption φ, excluding the periods of strange behavior for $load_n$ and φ. The subscript represents the individual node index n. The algorithm removes one home at a time giving priority to homes with low energy if two conditions are satisfied. First, the node in question must have more than one home. Second, the resulting homes energy consumption at the node will not surpass 1.5 times the φ. The value of 1.5 was chosen to avoid deviating too much from the average yearly consumption. The resulting number of homes by load node is presented in Fig. 3.4. Presenting the number of homes before and

Algorithm 1: Removes the extra homes.

```
load_n ← load at every node n
NH_n ← number of homes per node
φ ← average yearly home energy consumption
while Σ_{n=0}^{192} NH_n > 1,120 do
    υ_n = Σ load_n / NH_n ← home energy
    i ← index order from smaller to larger υ_n
    for (j = 0; j ≤ 192; j = j + 1) do
        if NH_{i[j]} > 1 & Σ load_{i[j]} / (NH_{i[j]} − 1) < 1.5 × φ
         then
            NH_{i[j]} ← NH_{i[j]} − 1
            break
```

Figure 3.3. Algorithm that removes the extra homes from Midwest 240-Node distribution system test case.

after running the Algorithm that removes the extra homes.

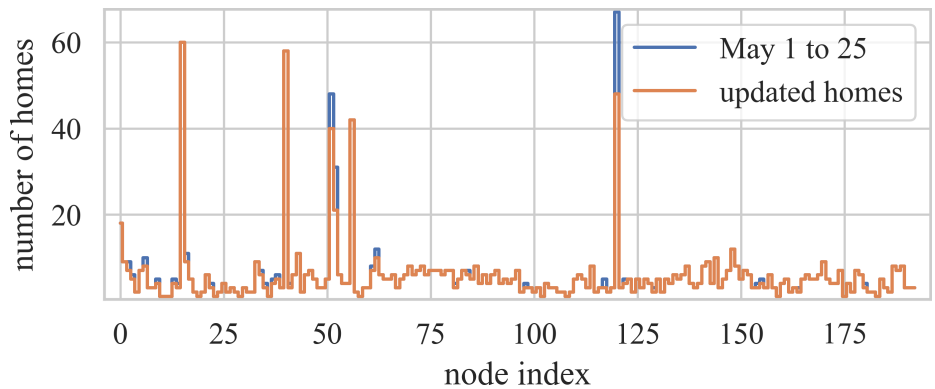

Figure 3.4. Number of homes by load index considering the period of May 1 to May 25 and the updated number of homes (i.e., the output from the algorithm).

The synthetic queueing load model $M_t/G/C_t$ is chosen to generate data for presenting lower deviation from the smart meter data than the other queueing models. The time-varying power supply capacity (C_t) is never allowed to be smaller than 1,500 W and made by giving a gain of 2 to the customer expected load. Thus, always having the possibility of serving multiple appliances, and suppressing unrealistic and unfeasible load peaks [54].

3.5 Validation

3.5.1 Overview

This section presents the validation of the proposed approach for the generated granular-level synthetic load compared to the original node-level smart meter data. The synthetic load generated with the queueing load model utilizes the nodal load smart meter real data for the year of 2017 is divided by the expected number of homes of that node, making the expected load of a customer ($l(t)$). With that, the queuing load model is run generating the arrival of appliances for each customer. Thus, having the load by customer

and the appliances that compose the customer load.

3.5.2 Smart Meter vs. Synthetic Load

The queueing model is a random process of arrival of appliances and as such the generated load will be different from the smart meter data. In this section the periods of strange smart meter behavior presented in Section 3.4.2 and further explored in Appendix B are not considered. A metric utilized for evaluating the distance or error from the smart meter load and synthetic load is the mean absolute percentage error (MAPE),

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|. \tag{3.1}$$

Where A_t is the smart meter data and F_t is the generated synthetic load. The subscript t is discreet time in an hour resolution. Thus, the synthetic load is converted to hourly consumption to enable the comparison.

From [54] it is known that one of the characteristics to reduce the difference from the reference curve is the number of aggregated customers. Thus, the larger the number of customers being aggregated to a node load a smaller error is expected. Fig. 3.5 demonstrates this characteristic. The nodes and their respective MAPE over the year are plotted having the load nodes ordered from small to large in respect to their number of customers. As it can be observed there is a negative correlation in between the number of homes and MAPE.

Fig. 3.5 last three node loads present a peculiar behavior. Given that they are the nodes with most houses but appear to have an increasing MAPE. Table 3.1 presents more variables to assist in explaining this behavior, i.e., load node numeric identifier (Node),

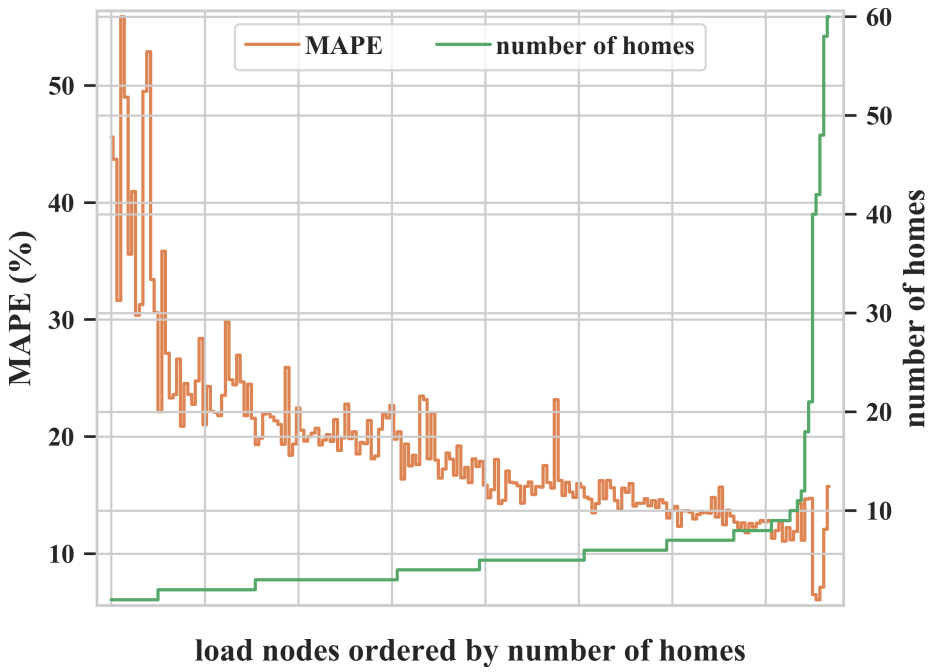

Figure 3.5. Year MAPE by load node in relation to the number of homes.

number of homes (*NH*), minimum customer load (min), median customer load (median), customer energy (energy), and load node MAPE. The table presents the last six load nodes from Fig. 3.5, having a line to separate the load Nodes 52, 51, and 56 (i.e., more *NH* lower MAPE) from the load Nodes 120, 40, and 15 (i.e., more *NH* larger MAPE). Other characteristics that increase the difference from synthetic to a reference are the low values of the reference curve being summarized in Table 3.1 with minimum, median, and energy. Low values of the reference curve are problematic given that some periods are not likely to have appliances arriving and/or large inter-arrival periods [54]. Node 120 is like Node 51 in minimum load and energy; however, the low median increased the MAPE. Node 40 is like Node 52 in energy; however, the higher minimum and median lowered the MAPE. Node 15 is like Node 56 in energy; however, the lower minimum and median increased the MAPE.

Table 3.1. Explaining the number of homes deviation from smaller MAPE.

Node	*NH*	min (W)	median (W)	energy (MWh)	MAPE (%)
52	21	156.19	681.90	12.99	14.72
51	40	238.05	1567.32	13.09	6.50
56	42	433.98	1063.80	9.26	6.07
120	48	214.58	836.79	13.24	7.12
40	58	413.79	1368.27	11.39	12.09
15	60	357.00	677.66	8.67	15.73

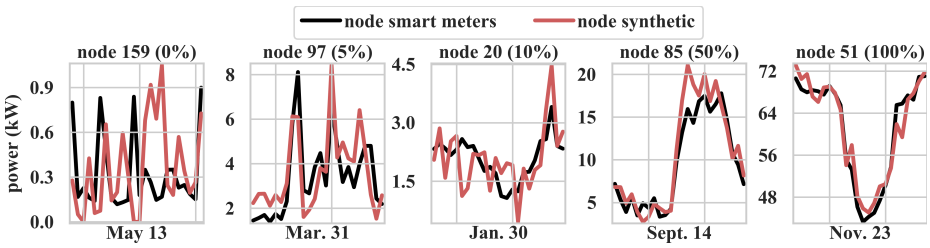

Figure 3.6. Day load node visual comparison of know load and generated load. The daily load of every load node are ordered according to the MAPE. Thus, 0% is the worst and 100% the best, i.e., 5%, and 10% are values in between.

The knowledge of the MAPE by explored node for the year in Fig. 3.5 is further explored graphically in Fig. 3.6. Where the MAPE is computed for a day for all the days in all the nodes. Furthermore, the resulting day and node MAPE is ordered in a decreasing order. Fig. 3.6 graphically presents five days giving an insight into the daily behavior. Each of the five plots has a title containing the load node number followed by a percentage wise position of the day on the ordered day and node MAPE. Thus, 0% is the worst and 100% the best, i.e., 5%, 10%, and 50% are values in between. The worst day node plot behaves as the name suggests, having the synthetic load barely following the actual nodal load. The load Node 159 has a single customer and the maximum load for that day is 0.9 kW. Thus, this situation has both the lack of customer aggregation and low reference values. However, the load Node 159 for May 13 is the worse day node load. The second

plot presents the behavior of the load Node 97, two customers, for March, 31. Presenting a day near the worse day node load, meaning that 95% of the synthetic load performs as good or better at following the smart meter reference curve. The third plot shows the load Node 20, two customers, for January 30. Where 90% of the synthetic load performs as good or better at following the smart meter reference curve. The fourth plot presents the load Node 85, five customers, for September 14. Representing the median load node day. Thus, 50% of load node days performer ether better, equal, or worse. The fifth and final plot shows the load Node 51, 40 customers, for November, 23. This particular node also appears in Table 3.1 for further information on it and is the best node load day for the synthetic generated queuing load model.

As expected and demonstrated in Fig. 3.5 and Fig. 3.6, aggregating customers reduces the differences in between the smart meter load and the synthetic load. In a similar fashion to the previously described Fig. 3.7 presents the worse, median, and best load days in four levels of aggregation. The four levels of aggregation are each row and are the tree feeders of the system and the total distribution system, i.e., Feeder S, Feeder M, Feeder L, and system (i.e., entire Midwest 240-Node test case). Every plot in Fig. 3.7 has a title containing that particular level of aggregation day MAPE. The smallest level of aggregation is Feeder S, that for the worse day is 7.92% MAPE, i.e., no other day in any other presented level of aggregation will perform worse. The Feeders S, M, and L have a clear reduction of MAPE from smaller to larger.

The levels of aggregation are ordered from smaller to larger in Fig. 3.7. However, Feeder L has lower MAPE for the worse and best day than the complete distribution system. This occurred given that we are presenting the MAPE for the day being presented.

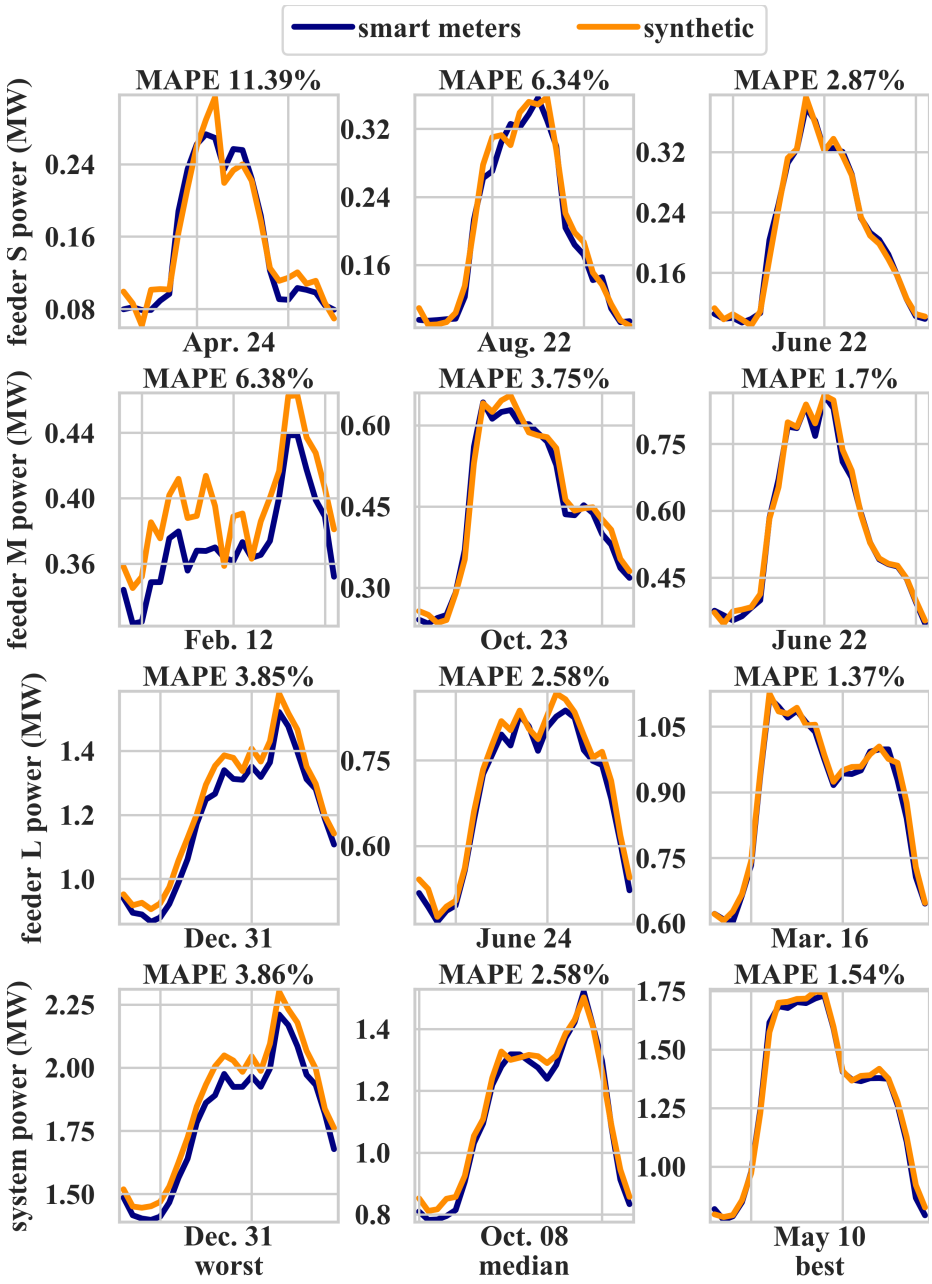

Figure 3.7. The worse, median, and best load days are presented in four levels of aggregation, i.e., Feeder S, Feeder M, Feeder L, and system (i.e., entire Midwest 240-Node test case). On the top of every plot is the day MAPE.

Table 3.2 presents the levels of aggregation in relation to the number of customers and the yearly MAPE. Demonstrating that increasing the aggregation will reduce the MAPE.

Table 3.2. MAPE for the year of 2017 in four levels of aggregation, i.e., Feeder S, Feeder M, Feeder L, and system.

	NH	MAPE (%)
Feeder S	76	6.4617
Feeder M	370	3.8090
Feeder L	674	2.5864
System	1,120	2.5828

3.5.3 Power Flow Comparison

The different input loads impact for the Midwest 240-Node test case power flow are presented in this section. Section 3.5.2 demonstrated that smart meter and synthetic load data are not exactly the same. However, the synthetic load follows the behavior of the smart meter load data. The impact on the power flow is presented by demonstrating the voltage behavior in four points of Midwest 240-Node test case . The comparison makes use of a violin plot showing the annual distribution of voltage magnitudes in p.u. located in four points of the Midwest 240-Node network for phases A, B, and C are presented in Fig. 3.8 (8,760 voltage magnitude samples for every half violin plot, thus, a total of 210,240 voltage magnitude samples for the four nodes and two load types using one-hour time resolution over one-year). This type of plot is like a box plot, but with the (rotated) kernel density plot on each side. The thickness (or density) represents how often each voltage magnitude occurred.

The node being presented in Fig. 3.8 are labeled 1 ST, 2 FS, 3 FM, and 4 FL, referred to the substation transformer primary side, node 10 from Feeder S, capacitor node Feeder M, and capacitor node Feeder L respectively. The location of the presented nodes

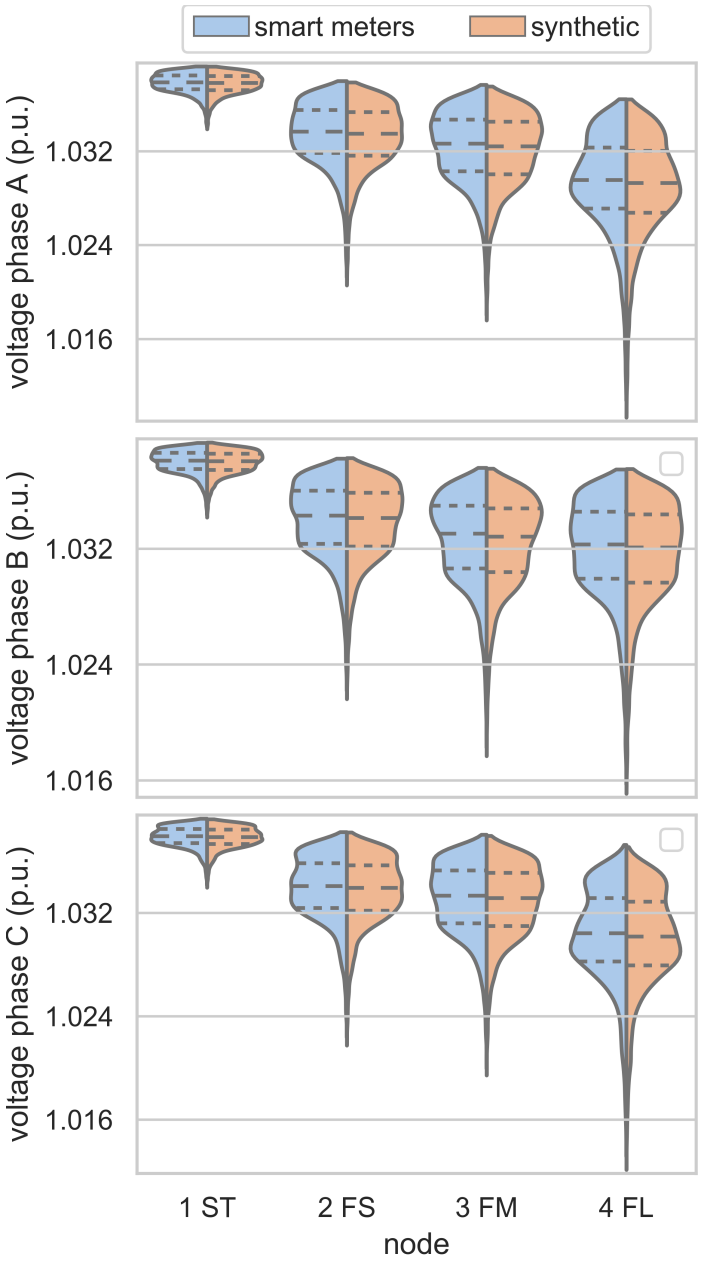

Figure 3.8. Distribution of voltage magnitudes for smart meter and synthetic load data for Midwest 240-Node distribution system test case for one-year. The labels 1 ST, 2 FS, 3 FM, and 4 FL, referrer to the substation transformer primary side, node 10 from Feeder S, capacitor node Feeder M, and capacitor node Feeder M respectively.

are identifiable in Fig. 3.1. The three phase nodes were chosen empirically with the intent of demonstrating the voltage at the substation and within each of the feeders. The violin plots in Fig. 3.8 are split in half having, smart meter on the left side, and the synthetic on the right. The split violin plots have similar shape to their other half and the median and quartile are near each other. Thus, the power flow studies with the generated synthetic load approximate the behavior of the smart meter load.

3.5.4 Created GridLAB-D Model

The GridLAB-D simulation software for distribution systems. The core of GridLAB-D has an advanced algorithm that simultaneously coordinates the state of millions of independent devices, each of which is described by multiple differential equations. GridLAB-D examines in detail the interplay of every part of a distribution system with every other. Incorporates an extensive suite of tools to build and manage studies and analyze results, e.g., agent-based and information-based modeling tools that allow users to create detailed models of how new end-use technologies, distributed energy resources, distribution automation, and retail markets interact and evolve over time. Thus, being of interest for multiple power system studies especially for smart grids, smart cities, demand response, and home energy management systems [95].

The OpenDSS model available from [93] is converted to GridLAB-D making use of the python packages DiTTo [96] and glm [97]. The package DiTTo makes an initial conversion file for GridLAB-D, however, without considering the split phase structure and the impedance to which the distribution system is connected to the main grid. The package “glm” is used to addresses the split phase structure and some unity mismatches.

The impedance from the swing node to the distribution system is computed as presented in [98]. Table 3.3 presents the comparison of voltage and current from OpenDSS and GridLAB-D, for a single power flow solution. The single power flow solution is the original available from [93]. The voltage magnitude comparison is performed for all the nodes, Table 3.3 presents only the worse for each of the phases. The percentage wise maximum error observed on voltage magnitude is below 0.0009%. The current magnitude comparison is performed for the lines and transformers primary, however, currents below 0.1 A on OpenDSS are not considered, Table 3.3 presents only the worse for each of the phases. The percentage wise maximum error observed on the considered current magnitude is below 0.04%. The GridLAB-D model with the synthetic load data is made publicly available at [99] and [100].

Table 3.3. Comparison of voltage and current from OpenDSS and GridLAB-D, for a single power flow solution.

Maximum error observed in all nodes	Phase		
	A	B	C
Voltage (mV)	0.7983	0.8102	0.7394
Maximum error observed in lines and transformers primary (currents below 0.1 A on OpenDSS are not consider)	Phase		
	A	B	C
Current (mA)	3.8504	0.6626	0.9765

3.6 Conclusions

This chapter developed synthetic load data inspired on real time-varying smart meter data for the Iowa State distribution system test case . The smart meter data is from a real distribution system in the U.S. Midwest region. The available smart meter data has an hour resolution and customers in the same distribution node are aggregated to preserve

their privacy. The generated synthetic queueing load data used only the publicly available data that approximate the aggregated behavior of the smart meter data. The generated synthetic load data models individual residential customers and their individual electric assets. The granular-level load is individually known for all the 1,120 homes to create the Midwest 240-Node test case . The appliances that compose every residential home is also known. The procedure presented in this chapter for the generation of the synthetic load data that aggregated to the complete power system region demand is applicable to other test systems. The procedure consists of analyzing the available demand, address possible challenges, assuming the nodal load is not available the demand would have to be segregated to nodal level, segregate the nodal demand to the customer level, and generate the load with the synthetic queueing load method. Assuming portions of the demand are desired to be treated differently it is only required to remove that demand from the reference given to the queueing load method. The studies of this test system with the synthetic load data are intended mainly for smart grid technologies. For this reason, the Iowa State distribution system test case OpenDSS model is converted to GridLAB-D and validated in this chapter. GridLAB-D is an agent-based approach for simulating smart grids, e.g., market design, building control system design, and integration of new technologies. The GridLAB-D model with the synthetic load data is made publicly available, allowing researchers to validate their methods.

CHAPTER 4 Combining HEMS with PV Overvoltage Mitigation in Low Voltage PV Rich Distribution Networks

4.1 Overview

The utilization of the developed synthetic queueing load model presented in Chapter 2 for energy management systems is demonstrated in this chapter. This creates a framework for testing home energy management systems in low voltage photovoltaic (PV) rich networks. Low voltage PV rich networks face the challenge of overvoltage, which limits the amount of energy generated by the PV arrays to be injected into the system. Thus, reducing the revenue of customers who invested in the PV systems (i.e. increasing even further the already long payback period of PVs systems). The developed framework presented here is capable of evaluating home energy management systems in low voltage systems with PV local inverter controllers. To the best of the author's knowledge, this has not been done before as no mention appears in the literature of the field. The framework offers the capability to make a local PV generation forecast. Furthermore, one of the home energy management systems utilizes a partially observable Markov decision process in an attempt to consider the uncertainty in the price of energy. As discussed in Chapter 1, uncertainties are important challenges in power system studies.

4.2 Introduction

The impact of distributed generation on the distribution system protection and voltage control is presented in [101], who discusses the challenges of protection coordination with distributed energy resources given their different characteristics and

bidirectional power flow on unidirectionally designed distribution networks. It also considers how renewable energy resource generation uncertainties impact legacy voltage control devices. The possible solutions to address the voltage regulation in the presence of distributed generation are presented: curtailment, demand response (DR), and static synchronous compensator. A review of distributed and decentralized voltage control of smart distribution networks is presented in [43] where DR is presented as a strategy that should be further explored for voltage support.

The impact of DR on the distribution system voltage profile in the presence of renewable energy resources is presented in [44]. The authors suggest for future work the inclusion of reactive power support from PV inverters. Similarly, the impact of DR on load, losses, and load factor is presented in [45]. The distribution system impact of load changes to mitigate overvoltage in the presence of renewable energy resources is then presented in [43], [44], [47]–[52], [101]–[103]. Changing the load, however, can also be utilized to mitigate undervoltage as discussed in [102] where the demand is reduced to mitigate undervoltage.

Distribution system overvoltage due to renewable energy resources, such as photovoltaic (PV), are more likely in periods with valley demand and peak generation. The relationship between self-consumption of renewable energy resources and the required curtailment is presented in [46]. However, the need to curtail PV is due to transformer limits, not overvoltage. Approaches that reduce the local mismatch of demand and generation by increasing the local consumption with the intent of mitigating overvoltage are presented in [49], [51], [52], [103]. Additionally, in [49], a distributed algorithm to control active and reactive power from PVs is presented. Considering

optimization in two-time scales, i.e. legacy conventional voltage control devices and fast PV inverters and DR resources, a centralized direct control optimization with receding time horizon to mitigate uncertainties is presented in [51] where the water heater of multiple customers performs the change in demand. In [52], the cost to curtail PV generation and perform load shifts is estimated with the distribution network Jacobian matrix. A multi-agent transactive energy management system is proposed in [103]; here the agents perform their heuristic optimization in series with updated price forecast given the actions of previous agents. Thus, only [49] considers reactive power support, but, the work assumes that every load node has some capability to control its power factor and the simulation aggregates the low voltage network.

Voltage support with DR is approached in [47], [48], [50] without directly attempting to increase self-consumption. Similar to work presented in [49], the authors in [48] make use of a distributed algorithm implemented in a multi-agent structure. The network is partitioned into zones where each zone-coordinator dispatches the active and reactive power of various DER and DR using a gradient descent method. In [50], mitigating overvoltage problems in the distribution grid are discussed. Changing the load to having 4 setpoints based in a real-time voltage signal in a specific system is presented. Thus, for [48], [50], DR only participates given the system voltage. In contrast, a multi-agent with a hierarchically controlled and multi-objective renewable energy management scheme is presented in [47]. The 3 objectives are lowering electricity bills, minimizing power purchased from the main grid, and optimizing the power quality. Thus, the hierarchical structure provides coordination for balancing the three objectives; however, reactive power support from renewable energy resources is not considered.

An increased self-consumption naturally assists in preventing overvoltage since overvoltage is more likely in periods with valley demand and peak generation. Overvoltage mitigation strategies that utilize DR will, at some level, attempt to increase self-consumption. For a PV rich distribution network, days during the summer with clear skies present a higher generation. Assuming the customers invested in automation to perform demand shifts, the resources should be utilized throughout the year (i.e., not only during one period of the year). From the literature review, no DR strategy in low voltage PV rich network considers the utilization of PV inverters local controllers for voltage support (mitigating overvoltage). This chapter develops a home energy management system DR strategy considering PV inverters local controllers.

The remainder of this chapter is organized as follows: Section 4.3 presents the system model. The developed scheduling appliances strategies are presented in Section 4.4. Section 4.5 describes the simulation. The simulation results are presented in Section 4.6. The discussion on the developed approach is available in Section 4.7.

4.3 System Model

4.3.1 Overview

The models being utilized in the simulation are discussed in this section. Section 4.3.2 reminds the reader of the synthetic queueing load model extensively discussed in Chapter 2. In Section 4.3.3, two local PV inverter controllers are presented. The implementation of the PV inverter controllers in a quasi-steady-state simulation environment is presented in Section 4.3.4. Section 4.3.5 presents the local statistical model for the solar irradiance forecast. An overview of the billing structure of the

distribution company of Chicago, IL, U.S., ComEd appears in Section 4.3.6. Finally, a discussion on the uncertainty of price is presented in Section 4.3.7.

4.3.2 Queueing load model

The synthetic queueing load model combines a top-down, bottom-up approach with the expected load of a customer ($l(t)$) as the input for computing statistical time varying arrival rate of appliances for a customer. The appliances are modeled as generic blocks of energy as in [54]. A detailed discussion on the synthetic queueing load model is presented in Chapters 2 and 3. The queueing load model utilized in this chapter is only a portion of the ones from Chapter 3.

4.3.3 Photovoltaic Inverter Controllers

The PV inverter controllers presented in this section are droop-based controllers from [24]. The active power curtailment droop-based approach gives a gain m (kW/V) to the difference between the measured voltage V and the critical voltage V_{cri}. The difference and the gain describes how much active power will be curtailed from the total available in the PV-array P_{MPPT}, resulting in the power actually injected by the PV inverter P_{inv} as described by [38]:

$$P_{inv} = \begin{cases} P_{MPPT} - m(V - V_{cri}), & \text{if } V \geq V_{cri} \\ P_{MPPT}, & \text{if } V < V_{cri} \end{cases} \tag{4.1}$$

Fig. 4.1 graphically illustrates the curve behavior of the P_{inv}, active power, assuming the P_{MPPT} is kept constant.

Active power presents a larger impact on the system voltage given that low voltage

distribution systems are much more resistive than reactive ($R >> X$) [27]. However, reactive power is capable of assisting in providing voltage support. An active-reactive droop PV inverter controller, in addition to the active curtailment describe by (4.1), presents reactive power support. The reactive power support is described by:

$$Q_{inv} = \begin{cases} 0 & V \leq V_{kick} \\ \frac{-Q_{max}}{V_{cri}-V_{kick}}(V - V_{kick}) & V_{kick} < V < V_{cri} \\ -Q_{max} & V \geq V_{cri} \end{cases} \tag{4.2}$$

Where V_{kick} is the voltage, the inverter starts absorbing reactive power; and Q_{max} is the maximum reactive power the inverter can absorb, having the totality of the reactive power available for utilization before performing active power curtailment. Fig. 4.1 presents the behavior of the PV inverter Q_{inv} and P_{inv} assuming the P_{MPPT} is kept constant.

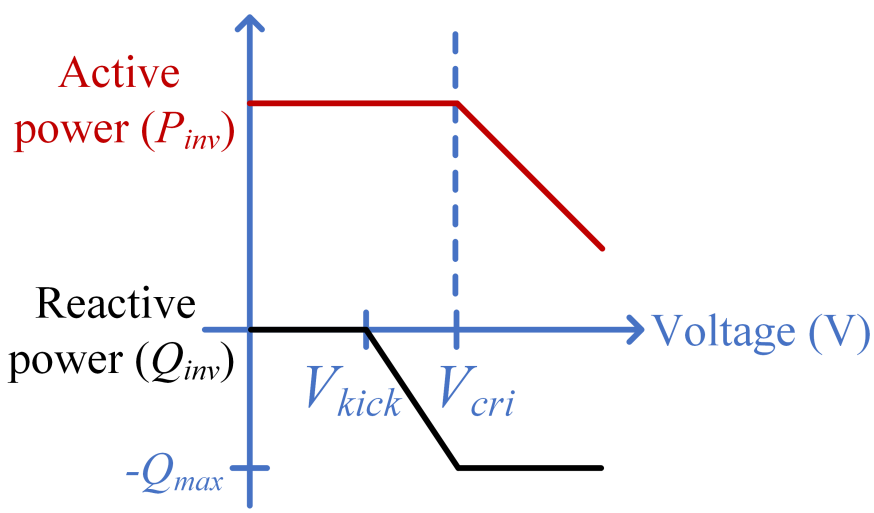

Figure 4.1. Droop-based PV inverter controllers. The active power curve from the inverter is presented in red. The reactive power curve from the inverter is presented in black.

4.3.4 Implementing the PV Inverter Controllers in Power Flow Simulations

The PV inverter controllers presented in Section 4.3.3 and other droop-based controllers operate as intended in dynamic simulations and are popular methods for preventing overvoltage in low voltage PV rich networks. To deploy the droop-based controllers in a quasi-steady-state simulation to reduce the computational burden of dynamic simulations, an interactive approach can be utilized to prevent numerical oscillation from occurring , as presented in [104].

For example, the implementation of the PV inverters droop-based active power curtailment is performed by the linear gain in the difference of the measured voltage V and the critical voltage V_{cri}. Developing a controller in quasi-steady-state simulation would cause large steps in the voltage since the controller is providing voltage support, and, at the same time, utilizes the voltage to define how much voltage support is necessary. Thus, presenting numerical oscillations of diploid directly, an interactive method is a possible solution to address the numerical oscillations. Instead of making large changes in the active power being injected at every customer, the changes are performed over iterations and the size of the step is dependent on the system characteristics. The voltage sensitivity matrix of the network ($\mathbf{S_V}$) contains the partial derivatives for changes in voltage in relation to the active and reactive power of its nodes. The $\Delta P, \Delta Q$ is the voltage change due to reactive power at each node and the $\Delta V, \Delta\delta$ the voltage change due to active power at each node. The network contains N-nodes. The voltage sensitivity matrix is written as,

$$\mathbf{S_V} = \begin{bmatrix} (\frac{\Delta\delta}{\Delta P}) & (\frac{\Delta\delta}{\Delta Q}) \\ (\frac{\Delta V}{\Delta P}) & (\frac{\Delta V}{\Delta Q}) \end{bmatrix}_{2N\times 2N} \tag{4.3}$$

The voltage sensitivity matrix is the base for implementing all types of droop-base PV inverter controllers in a quasi-steady-state simulation. The complete explanation is presented in [104].

4.3.5 Solar Irradiance Forecasting

Section 4.2 presented the challenges of increased integration of renewables at low voltage and the importance of self-consumption in order to mitigate overvoltage. However, it is not necessary to increase self-consumption for low generation periods, which enables the customer to schedule their consumption according to the price. In order to attempt to identify periods of large PV generation, a solar irradiance forecast is required. With the knowledge of the forecast irradiance, the necessity to increase self-consumption can be identified.

The solar irradiance forecast utilized is presented in [105], which was developed based on [106]. The solar irradiance presented consists of a statistical model that requires historical solar irradiance and locally measured irradiance. The strategy is developed with remote microgrids in mind (i.e. would have only access to local irradiance measurements). The solar irradiance forecast makes use of the Markov switching model. Markov switching model are models that combine two or more models to estimate or forecast a variable depending on an unknown state [107]. The Markov switching model was proposed given the empirical evidence suggesting that a time series behavior presents

different patterns through time. The unknown state is considered in the Markov process. Markov models have multiple states and the probability of being in those states alters the forecast or estimation. Commutation Markov models have been utilized in multiple fields of study because they excel at predictions or estimations of a non-linear nature. The commutation Markov model involves multiple structures that represent different behaviors during non-linear time series. The model offers the possibility to alter states in a probabilistic manner as it is capable of tracking a complex dynamic time series. Commonly, the hidden variable controls the change in between states of the Markov switching model. The hidden variable is commonly assumed to be a first order Markov chain.

The probability transition model for the Markov switching model for solar irradiance forecasting is presented in Fig. 4.2. The solar irradiance Markov switching model has 3 states, referred to as low, medium, and high. As the names suggest, the states are referred to as the solar irradiance states. The p_{num} refers to the probability p of the change in between states num.

In [105], the evaluation of the state based on the irradiance at the previous hour as well as other approaches is presented. However, the approach that makes use of the previously observed irradiance to discover the optimal state presented the best performance. Fig. 4.3 presented the forecast of the Markov switching model utilizing the first four hours of solar irradiance to discover the Markov state and illustrates the inspiration for the work in [105], which updates the forecast based on the previous hour. The past hour Pearson correlation coefficient from 2000 to 2011 were never below 0.966 for Brookings, SD, USA.

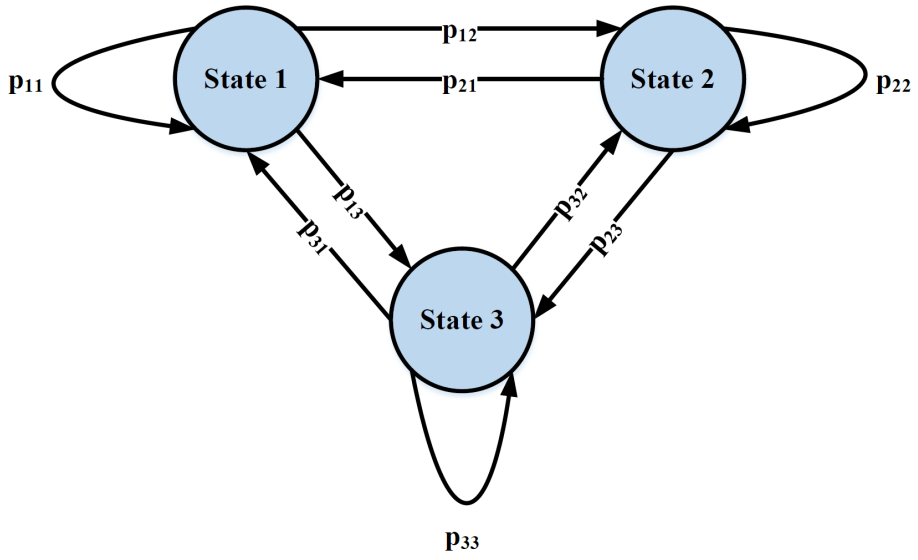

Figure 4.2. Probability transition for the Markov switching model. Image from [107].

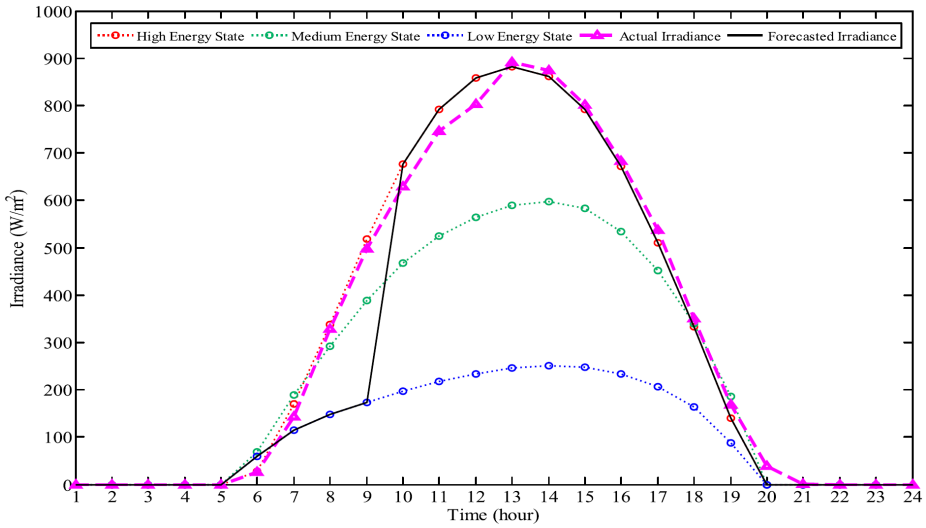

Figure 4.3. Irradiance variation for July 24, 2012. Image from [106].

The Markov switching model for solar irradiance is fitted in R with the package depmixS4 from [108]. The possible states forecast for leap and non-leap years are exported to file in which the forecast can now occur in any programming language.

4.3.6 ComEd Real-Time Pricing

The distribution company of Chicago, IL, U.S., ComEd presents Real-Time pricing (RTP). The RTP signal has a resolution of five minutes available at [109]. The electric bill of customers on RTP is comprised of a supply charge, delivery charge, capacity charge, taxes, and fees. The five minute RTP signal assists customer in understanding the RTP; however, the customer is billed according to the hourly price from Pennsylvania, New Jersey, and Maryland (PJM) interconnection. The PJM interconnection is a regional transmission organization that coordinates the movement of wholesale electricity in all or parts of 13 states and the District of Columbia. ComEd simply passes along the hourly market prices with no mark-up. The PJM real-time hourly price is the average of the previous 12 five minute interval signals. All real-time hourly market prices are subject to a 24-hour settlement period where the final price the customer will be billed is settled [109].

The capacity charge calculation is dependent on the customers previous year consumption during the system coincidental peak (PJM interconnection), the five hours of the summer when ComEd System demand was highest, and adjusting PJM factors [110]. These factors will be used to compute customer capacity obligation and the individual Capacity Charge. Thus, the larger the customer coincidental peak with ComEd and PJM for the previous year, the larger the Capacity Charge. A sample ComEd residential bill is

available in [111], presenting all the charges.

As explored in Chapter 1, a challenge of the power system is the uncertainty. Given the uncertainties and the required planning to have sufficient resources to supply energy to customers once the time arrives, there is a day-ahead market that is financially binding, which is different from the real-time market that is financially and generation binding. This means that the closer the assumptions of the power systems are to what actually happens, the closer the day-ahead cleared market price will be to the real-time market. The day-ahead market is cleared at approximately 4:30 p.m., providing cleared prices for the next day 24 hours in advance at an hourly resolution. Since the day-ahead cleared price has a tendency to approximate the real-time price, it will be referred to as the *forecast price*. The timing characteristic of the RTP and the forecast is given in Fig. 4.4.

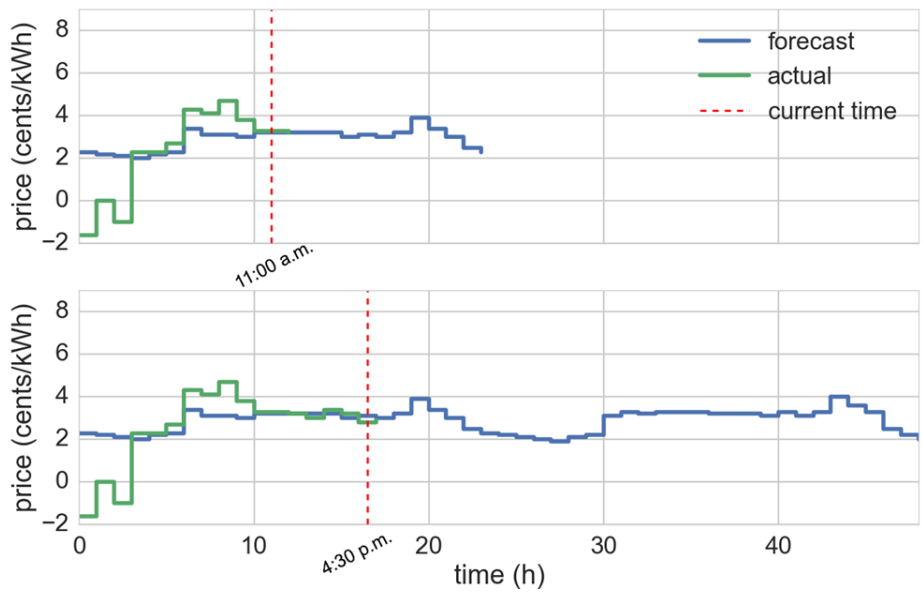

Figure 4.4. An example of the PJM forecast price availability and the RTP. Image from [55].

4.3.7 Uncertainty of Price Problem

As explored in Chapter 1, further uncertainties in the power system lie in load and PV generation. The uncertainty in price, however, is a particular problem for scheduling of residential appliances with DR. The problem of minimizing the residential customers cost of energy under the uncertainty of price is explored in detail in [55]. In summary, the customer has some appliances arriving in their queue and some schedulable to start running in a given time window—notice that there is no knowledge of appliances arriving in the future. Thus, the optimization problem of scheduling the appliances is run without precise information of the price. The vector of scheduling arriving appliances of the customer at times H_t. The ith element in the vectors correspond to the ith appliance. Scheduling appliances are removed from the vector once they have been scheduled to run in the current time. The optimization goal is to find the vector of start-times, $\hat{t}_{\text{start}}$, to minimize the total cost.

Different than the problem statement from [55], the scheduling of appliances is no longer able to be performed individually for every appliance since it is expected to be run for multiple customers. In [55], the author noticed an increase in the peak load when scheduling appliances, which is not a problem with a single customer; but considering multiple customers and all attempts to schedule most of their appliances at the expected lowest price would be a problem since the expected system valley load could become the system peak. This problem is referred to as the rebound effect [112]. A possible strategy to mitigate the rebound effect is presented in [55] that suggests imposing a maximum load constraint on the optimization.

The optimization problem needs to consider all the appliances in conjunction to avoid the rebound effect. The customers with PV rich networks are not only interested in the scheduling of appliances in regard to price, they also wish to use their load flexibility to mitigate overvoltage caused by PV in order to reduce the amount of active power curtailment. As presented in Section 1, self-consumption mitigated overvoltage. To encourage self-consumption, the optimization problem must also incorporate a lower bound for some periods of time where large PV generation is expected. Thus, being a much more complex optimization problem than the one from [55] since the appliances have to be considered in conjunction with both the lower and upper bounds.

4.4 Scheduling Appliances

4.4.1 Overview

Three distinct scheduling of appliances for HEMS are discussed in Sections 4.4.2, 4.4.3, 4.4.4, and 4.4.5 being the no scheduling, using the day-ahead price for scheduling, having knowledge of the future (i.e. to know the best performance, no real world application), and the partially observable Markov decision process respectively. These sections present the theory and assumptions of the HEMS approaches. Finally in Section 4.4.6 the formulation of the optimization problem is presented.

4.4.2 Immediate

The immediate scheduling of appliances (IMM), as the name suggests, maintains the same schedule of the appliances as their arrival to the customer queueing load model. The scheduled appliances are run at the same time they arrive at the queue. This means that none of the appliances are affected by anything other than the queueing load

model [54].

4.4.3 Assuming accurate forecast

As briefly presented in Section 4.3.6, the real-time price is not the only market in the power system. Thus, the assuming accurate forecast (AAF) makes used of the forecast price (the day-ahead cleared price) assuming there will be no deviation from the forecast price to the RTP. Thus, schedulable appliances are scheduled accordingly.

4.4.4 Theoretical Lower Bound

Since we will be performing a simulation to evaluate how to propose performance strategies without the knowledge of the RTP, the actual knowledge of the future is utilized to create a theoretical lower bound, meaning that there is no way for a system to perform better than the theoretical lower bound (LB). The LB determines the maximum gap between the methods on the assumption of knowledge of the RTP, subject to the same constraints (i.e. upper and lower bound of schedule appliances for a fair comparison).

4.4.5 Partially Observable Markov Decision Process

The partially observable Markov decision process (POMDP) from [55], [113] offers a non-myopic, receding horizon control method that balances the trade-off between immediate knowledge and the uncertainty of the future (i.e. uncertainty of the RTP). The receding horizon with RTP presents us with the known prior and current price. With historical knowledge of the distribution of the RTP, the expected RTP behavior for the future is known, but not the actual RTP. The trade-off between immediate and future decisions makes use of the Q-value approximation from Bellman's equation [114]. The appliances in the scheduling vector H_t ready to run have their individual actions a_i chosen

from the set of possible actions A (i.e., $a_i \in A$). Let $\hat{a}$ be the vector consisting of individual appliance actions (henceforth known as *the action*) to be determined by the scheduler optimization, x be the current state, x' be the next state (after taking action $\hat{a}$), $R(x,\hat{a})$ be the immediate reward for taking action $\hat{a}$ in state x, and $V^*(x)$ be the optimal cumulative reward value over the time horizon given an initial state x. We want to find the optimal action policy, $\pi^*(x)$, that maps states to actions to maximize the Q-value, $Q(x,\hat{a})$, given by the equation:

$$Q(x,\hat{a}) = R(x,\hat{a}) + \mathbb{E}[V^*(x')|x,\hat{a}]. \tag{4.4}$$

The action $\pi^*(x)$ is based on Bellman's principle [114], and given by:

$$\pi^*(x) = \underset{\hat{a}}{\operatorname{argmax}}\, Q(x,\hat{a}). \tag{4.5}$$

The home energy management system will take actions $\hat{a} = \pi^*(x)$ at each state x.

The POMDP formulation in [55] is utilized for the home energy management system. The formulation assumes two types of variables: the observable and measurements of the unobservable. The unobservables are emulated and filtered to estimate their posterior distribution. The distribution of the unobservables with the observable determines the belief state. The underlying state of the POMDP HEMS is represented by y_t for time t, having the vector $\hat{\Psi}_t$ of random variables describing the likely RTP, and the error between the utility forecast and current RTP is ε_t. Then, $y_t = (c(t), \hat{\Psi}_t, \varepsilon_t, H_t)$ is the underlying state where $\hat{\Psi}_t$ is unobservable. Unobservables measurements, future RTP available, and $\hat{\Psi}_t$ are the utility forecast price, $c_f(t,\tau)$, where

$|\hat{\Psi}_t[\tau]| = \tau_{\max}$. Given measurements $c_f(t,\tau)$, we can determine $P\big(\hat{\Psi}_t | c_f(t,\tau)\big)$ using a filtering method. The filtering method utilized is the same as in [55], [115].

Available appliance actions are to run at the current time or at a later time ($A = \{\text{run}, \text{wait}\}$). In decision events, $\hat{a}$ is determined having $|\hat{a}| = |H_t|$ and each a_i corresponds to appliance $H_t[i]$ to maximize $Q(x,\hat{a})$.

The particle filter utilized is the same proposed in [55] POMDP-GARCH, which performed best for larger uncertainty in the future RTP. The particle filter is a combination of two statistical models, the autoregresive (AR) (i.e., forecast the variable of interest using a linear combination of past values of the variable) and the generalized autoregressive conditional heteroskedasticity (GARCH) (i.e. specialized AR process that analyze time-series variance error). The AR process is given as:

$$c_{\text{ar}}(t) = k + \sum_{i=1}^{m} \big(\gamma_i c_{\text{ar}}(t-i)\big) + \varepsilon_{t-\text{ar}}. \tag{4.6}$$

Where $c_{\text{ar}}(t)$ is the cost output of the AR process, k is the AR constant, $c_{\text{ar}}(t-i)$ is the ith previous output, γ_i is the coefficient corresponding to $c_{\text{ar}}(t-i)$, m is the number of modeled coefficients, and $\varepsilon_{t-\text{ar}}$ is the error.

The AR error $\varepsilon_{t-\text{ar}}$ presented in (4.6) is then modeled by the GARCH process. Having σ_t be the standard deviation and $z_t \sim \mathcal{N}(0,1)$. The expected error from AR according to GARCH is $\varepsilon_{t-\text{ar}} = \sigma_t z_t$. A GARCH$(P,Q)$ process is fully described by $\varepsilon_{t-\text{ar}}$ and results in:

$$\sigma_t^2 = \chi + \sum_{i=1}^{P} \phi_i \sigma_{t-i}^2 + \sum_{j=1}^{Q} q_i \varepsilon_{t-j}^2. \tag{4.7}$$

Where ε^2_{t-j} be the jth previous square-error, σ_t represents the linear combination of prior inputs, χ is the GARCH constant, σ^2_{t-i} is the ith previous variance, ϕ_i is the coefficient corresponding to σ^2_{t-i}, P is the number of GARCH terms (prior variances), q_j is the coefficient corresponding to ε^2_{t-j}, and Q is the number of ARCH terms (prior square-errors). The combination of the AR with the GARCH statistical models is made by replacing the AR error term $\varepsilon_{t-\text{ar}}$ by the GARCH, resulting in:

$$c_{\text{ar}}(t) = k + \sum_{i=1}^{m} \left(\gamma_i c_{\text{ar}}(t-i)\right) + \sigma_t z_t. \tag{4.8}$$

In summary, the particles are possible samples of the future RTP generated according to (4.8) by the random sampling of z_t with $\mathcal{N}(0,1)$, meaning that every particle has a different expected RTP for the future. Every particle optimized their decision making based on their understanding of the RTP. The measurements of the unobservables are utilized to estimate the Q-value of the actions taken in the moment with their expected impact on the unknown future. For more clarifications on the action selector please refer to [55].

4.4.6 Scheduling of Appliances Optimization

Independently of how the RTP is consider in LB, AAF, and POMDP the scheduled appliances must be scheduled. To optimize the scheduled of appliances to minimize the cost of energy an optimization problem must be solved with desired understanding of the RTP.

As discussed in Section 4.3.7, the HEMS interested in minimizing the cost of energy and for individual homes has a upper bound to discourage the rebound effect and a

lower bound to encourage self-consumption from PV. To consider the discouraging and encouraging load consumption, the optimization problem is organized with "generators" with low cost C_{lower_i}, normal cost C_{RTP_i}, and high cost C_{high_i} to represent the cost for encouraging self-consumption, normal, and discouraging rebound effect, respectively. The underscore i represents a unit of time for the optimization. The load of the HEMS is $Load_i$ and the load is supplied by each of the "generators" is L_{lower_i}, L_{RTP_i}, and L_{high_i}. The available scheduled appliances are characterized by vectors, having every element of the vectors be the information of the j appliance. The set of appliances to be scheduled is J. The vectors are Ap_j, and Ad_j, being the appliance power, and duration, respectively. The As_j contains the i index were the appliance can be scheduled (i.e., the scheduling window). The optimization problem is modeled as a linear optimization for schedule of the appliances in order to minimize (4.9), being subject to the constraints (4.10). The objective function contains the portion of the load being supplied by each of the "generators" multiplied by there respective cost (4.9). The sum of supplied energy from the "generators" is constraint to be equal to the load (4.10b). The "generators" maximum capacity is enforced in (4.10c), and (4.10d). Where, LB_{lower_i} and LB_{RTP_i} represents the maximum generation capacity of the lower bound "generators", and the RTP "generators", respectively. Please notice that all the decision variables are enforced to be greater or equal to zero (4.10o). To performed the scheduling of appliances a binary variable $Bs_{i,j}$, and an integer variable $Al_{i,j}$ are utilized. The binary variable $Bs_{i,j}$ marks the start of running the appliance. Since a given appliance can only start once the summation of all the values of $Bs_{i,j}$ are made equal to one (4.10i). The integer variable $Al_{i,j}$ contains the units of energy of the appliances used through time, thus, the summation must be equal to

the appliance duration Ad_j (4.10e). The load $Load_i$ in turn must be equal to the multiplication of the $Al_{i,j}$ and appliance power Ap_j (4.10a). For $Bs_{i,j}$ to mark the start of an appliance, run the combinations of the inequality constraints from (4.10k) to (4.10m) must be satisfied. Since, the inequality constraints (4.10k) and (4.10l) require the knowledge of $Al_{i,j}$ at $i-1$ the first unit of time is a place holder. Furthermore, the receding time horizon optimization problem can have appliances with a scheduling window larger than the time currently being considered. For this reason the variable $Al_{i,j}$ is only allowed to the larger than one at the last unit of time (4.10g) and (4.10h), enabling appliances to be schedulable outside the time currently being considered. This requires that the last unit of time currently being considered to also be a place holder with an expected cost to the future. Given the place holders at the first and last unit of time the constraints (4.10f) and (4.10j) are utilized. The set N has all the values of i and the set n has all the values of i except "-1" (i.e., the last time considered). The $Al_{i,j}$ can have values larger than one at $i=-1$ (4.10h), not being useful for the inequality constraints from (4.10k) to (4.10m) for $Bs_{i,j}$. To ensure the appliance runs until its end the auxiliary variable $a_{i,j}$ is utilized. The variable $a_{i,j}$ contains the number of periods the appliance j should have run if it had started at $Bs_{i,j}$. Thus, not allowing the appliances to stop running once they have started with the equality constraint (4.10n).

$$\min_{Bs_{i,j},Al_{i,j},Load_i,L_{lower_i},L_{RTP_i},L_{high_i}} \sum_{i=0}^{N} L_{lower_i}C_{lower_i} + L_{RTP_i}C_{RTP_i} + L_{high_i}C_{high_i} \tag{4.9}$$

s.t.

$$Load_i = \sum_{j=0}^{J} Ap_j Al_{i,j} \tag{4.10a}$$

$$Load_i = L_{lower_i} + L_{RTP_i} + L_{high_i} \tag{4.10b}$$

$$L_{lower_i} \leq LB_{lower_i} \tag{4.10c}$$

$$L_{RTP_i} \leq LB_{RTP_i} \tag{4.10d}$$

$$\sum_{i=0}^{N} Al_{i,j} = Ad_j \tag{4.10e}$$

$$Al_{i,j} = 0 \qquad i,j \notin \{As_j\} \tag{4.10f}$$

$$Al_{i,j} \leq 1 \qquad i,j \in \{As_j\} \cap \{n\} \tag{4.10g}$$

$$Al_{i,j} \leq \infty \qquad i = -1, j \in \{As_j\} \tag{4.10h}$$

$$\sum_{i=0}^{N} Bs_{i,j} = 1 \tag{4.10i}$$

$$Bs_{i,j} = 0 \qquad i \in \{0,-1\} \tag{4.10j}$$

$$Bs_{i,j} \geq Al_{i,j} - Al_{i-1,j} \qquad i \notin \{0,-1\} \tag{4.10k}$$

$$Bs_{i,j} \leq 1 - Al_{i-1,j} \qquad i \notin \{0,-1\} \tag{4.10l}$$

$$Bs_{i,j} \leq Al_{i,j} \qquad i \notin \{0,-1\} \tag{4.10m}$$

$$\sum_{i=0}^{n} Bs_{i,j}a_{i,j} = \sum_{i=0}^{n} Al_{i,j} \tag{4.10n}$$

$$L_{lower_i}, L_{RTP_i}, L_{high_i}, Load_i, Al_{i,j}, Bs_{i,j} \geq 0 \tag{4.10o}$$

4.5 Simulation Setup

4.5.1 Overview

The simulation setup with the consideration of the approaches from Section 4.4 are presented here. Section 4.5.2 presents the 12 homes low voltage PV rich test system network. The appliances parameters and the customer $l(t)$ parameters for the synthetic queueing load model utilized are presented in Section 4.5.3 and 4.5.4, respectively. Section 4.5.5 presents the creation of the minimum self-consumption based on the PV irradiance forecast. Section 4.5.6 presents the linear optimization problem. Finally, in Section 4.5.7, the simulation scenarios are presented.

4.5.2 Test System

The chosen test system for testing is the 12 house radial distribution system from [38]. The low voltage system was chosen given its prior development with PV inverter controllers presented in [24]. The solving of the power flow quasi-steady-state time-series software utilized GridLAB-D utilizing Bus.py [116] to communicate with python 2.7 where the PV inverter controllers are implemented as described in [104] and the summary previously presented in Section 4.3.4. The files were manually converted to python 3.7. Every residential customer or home possesses an installed peak PV capacity of 8.4 kW. The system distribution transformer is 75 kVA, single-phase, 14.4 kV–120/240 V shown in Fig. 4.5. The feeder is 120 m long and the service entrance are connected to it by 20 m long cables. The line parameters of the benchmark feeder are provided in [38], being two live wires twisted around a grounded neutral cable (NS 90 3/0 AWG) and two wires supported by a steel grounded neutral cable (NS 90 1/0 AWG)

for the feeder and service entrance respectively. In the 8.4 kW capacity PV system, the efficiency is taken as η =16.7% and A =50.2605 m^2 [24], [117], having the power available at the PV array (PV_{power}) equal to $PV_{power} = \frac{16.7}{100}\ 50.2605\ PV_{irrad}$ where PV_{irrad} is the solar irradiance and PV_{power} unit is kW.

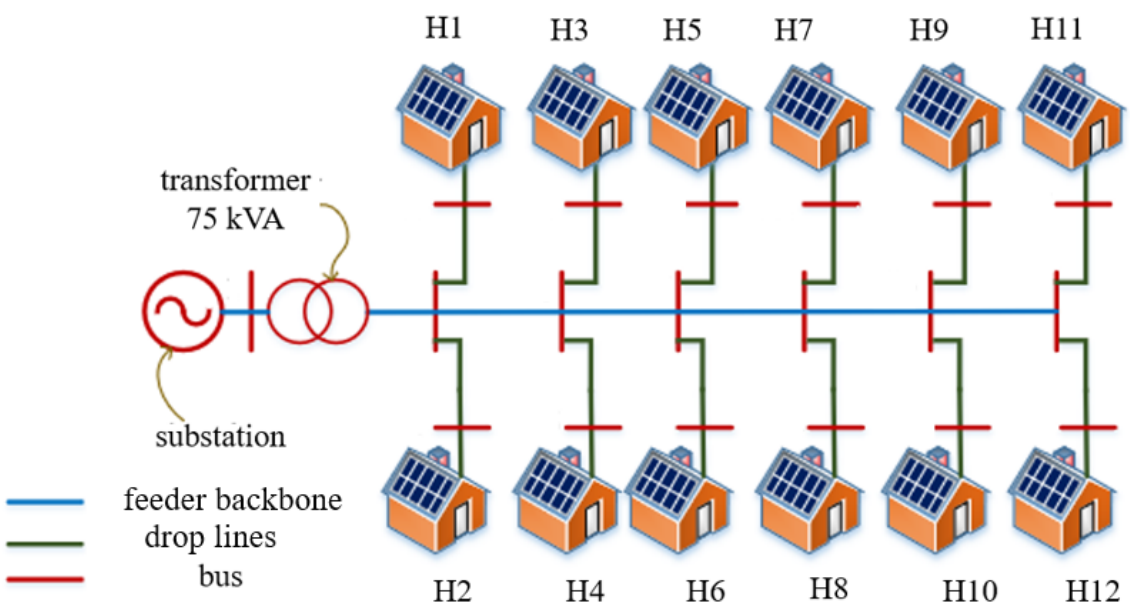

Figure 4.5. 12 house benchmark feeder with 8.4 kW grid-connected PV installed at each house.

4.5.3 Appliance Model

Section 4.3.2 presented an overview of the queueing load model for generating synthetic load profiles for energy management studies [54]. The queueing model generates the synthetic load with the arrival of appliances. Thus, the set of appliances to arrive must be generated. The appliance set is generated by performing multiple random samples. The samples generate the appliances: size, scheduling window, and ZIP characteristics. The ZIP appliance characteristics are obtained from [63]. The study was

conducted to characterize the effects of voltage variations in load consumption with field validation [62] with the intention of energy conservation using Volt/var control at the substation level. Further details are presented in [54]. The appliances are modeled as blocks of energy with specific duration and constant power draw. To generate the blocks of energy, two distinct gamma distributions were sampled to obtain the appliance duration and power. The scheduling window is generated by sampling two times: a gamma distribution for the time prior and after the appliances intended run time. However, this is only performed for schedulable appliances, which are selected by a random sample considering a desired percentage of the appliance set. Gamma distributions are continuous probability distributions in the positive real number set defined by two parameters (shape k and scale θ). The mean of a gamma distribution is $\mathbb{E}[X] = k\theta$, and the variance is $Var(X) = k\theta^2$. Thus, by defining the mean μ and standard deviation σ, the gamma parameters k and θ are computed with $k = \mu^2/\sigma^2$ and $\theta = \sigma^2/\mu$.

The appliance set ψ is generated having the gamma distribution parameters as power (W) $\mu = 500$ and $\sigma = 100$ and appliance duration (hour) $\mu = 0.5$ and $\sigma = 0.25$, as utilized in [54]. A review of home energy management systems is presented in [118] showing a comparison of multiple studies in regard to their parameters and assumptions. The mean peak reduction of the studies is 29.6%, which is chosen to be equal to the percentage of scheduling appliances. The scheduling period of time of the schedulable appliances, i.e. scheduling window, depends on the customers' willingness. A survey for customers responding to time varying price of energy is presented in [119] where 76 pricing experiments are analyzed in an attempt to identify the price responsiveness of customers. The comparison supports the possibility of achieving peak reductions of the

theoretical studies. A modeling complexity survey for home energy management systems is presented in [118] describing distinct models and the presence of demand shift. Thus, the change in demand given the price difference between the periods of time. The Ameren Illinois utilities power smart pricing report [120] presents models to attempt to characterize the price responsiveness of customers and averaged estimated demand changes for 24 hours on different days and for different types of customers. However, customers do not respond only to price changes. The experimental study by Gyamfi et al., found that extending the incentive options to reducing carbon emissions increased customer participation [121]. Exploring the different incentives for customer participation, such as functionality, price, and carbon emissions, an analytic hierarchy process for prioritizing user preferences is presented in [122]. Given the considerable deviations and assumptions, and its dependency on customers and types of incentives, the gamma distribution to define the scheduling window is chosen as duration (hour) $\mu = 2.68$ and $\sigma = 1.95$, as utilized in [16], [123].

4.5.4 Residential Customer Pattern

The residential customers' pattern comes from the xx test system described in detail in Chapter 3. One of the 193 load nodes for the system (i.e., node 143) possesses exactly 12 homes. Given that the node was chosen to characterize the low voltage network of the 12 home test system presented in Section 4.5.2, the individual home load makes the same assumptions explained in Chapter 3. Thus, the reference customer curve $l(t)$ is the same for all the 12 customers and is computed by dividing the nodal load by 12. Please keep in mind that the actual load of the 12 customers is distinct from each other. The

synthetic queueing load model utilized is the $M_t/G/C_t$.

4.5.5 PV Curtailment on the Test System

The 12 home low voltage test system was run with the load of all homes equal to $l(t)$ with the PV inverter active power curtailment local controller. May 30, 2019 has the peak solar irradiance and is evaluated in detail. Table 4.1 presents the power available in the PV array, the reference load, and the PV power curtailment by home in the 12 home system. The 12 home system is symmetrical and, in this case, the residential load is exactly the same in all homes; thus, the PV curtailment is presented for one home from the pair. The first two homes in the system have no PV curtailment and as such are not presented in the table. Half of the homes only present PV curtailment for PV generation above 4456.95 W and 4691.96 W, hours 9 and 16 respectively. With this in mind, self-consumption to avoid overvoltage in the low voltage network will be set to start for PV generation above 5 kW. The nominal power of the PV arrays is 8.4 kW, the incentive for self-consumption is made to be maximum of 10% of the schedulable load (i.e. the total schedulable load is expected to be 29.6%). A linear interpolation of the values will be utilized for the creation of the self-consumption incentive.

4.5.6 Optimization Implementation

The optimization problem is modeled as a linear optimization in python with the package Pulp [124]. The variable LB_{lower_i} represents the maximum lower bound "generators" capacity and the variable LB_{RTP_i} is the maximum bound of the RTP "generators" capacity, encouraging load consumption below LB_{lower_i} and discouraging load consumption above the LB_{RTP_i}. Load consumption above LB_{RTP_i} is discouraged to

Table 4.1. 12 home test system PV generation, load, and active power curtailment for the peak solar irradiance day

			PV curtailment in (W) by home				
hour	PV (W)	$l(t)$ (W)	H3	H5	H7	H9	H11
0	0.00	590.25	0.00	0.00	0.00	0.00	0.00
1	0.00	565.25	0.00	0.00	0.00	0.00	0.00
2	0.00	539.41	0.00	0.00	0.00	0.00	0.00
3	0.00	542.33	0.00	0.00	0.00	0.00	0.00
4	75.54	905.33	0.00	0.00	0.00	0.00	0.00
5	436.46	870.16	0.00	0.00	0.00	0.00	0.00
6	923.28	838.58	0.00	0.00	0.00	0.00	0.00
7	1628.33	771.58	0.00	0.00	0.00	0.00	0.00
8	3164.35	825.33	0.00	0.00	0.00	0.00	379.47
9	4456.95	760.75	0.00	0.00	858.94	1630.02	1972.62
10	7218.41	695.91	0.00	2158.12	3517.19	4348.34	4701.73
11	7923.46	866.87	221.98	2563.52	4000.19	4826.62	5198.65
12	8116.51	531.43	573.33	2974.82	4448.99	5296.12	5677.40
13	7839.53	1176.55	0.00	2234.96	3661.09	4469.34	4839.49
14	7142.87	994.65	0.00	1792.03	3190.28	3987.64	4352.27
15	6068.50	1016.15	0.00	928.28	2141.84	2971.24	3278.36
16	4691.96	992.70	0.00	0.00	863.83	1645.55	1964.46
17	2727.88	1179.81	0.00	0.00	0.00	0.00	0.00
18	545.57	1111.94	0.00	0.00	0.00	0.00	0.00
19	83.93	968.39	0.00	0.00	0.00	0.00	0.00
20	0.00	1215.36	0.00	0.00	0.00	0.00	0.00
21	0.00	893.93	0.00	0.00	0.00	0.00	0.00
22	0.00	722.17	0.00	0.00	0.00	0.00	0.00
23	0.00	517.50	0.00	0.00	0.00	0.00	0.00

avoid the rebound effect, but the rebound effect happens when multiple customers change their load to the same time step and now the expected valley is a peak. This will not happen in the 12 homes test case, which does not consider the change in price given the change in demand; however, it is a necessary characteristic for its deployment in a large system. Both the LB_{lower_i} and LB_{RTP_i} impact the scheduled appliances, which are 29.6% of the total. The homes load forecast is the reference load of the customer $l(t)$ as presented in Section 4.5.4. The $l(t)$ of the customers peak for the 12 homes network is 38,916 kW at the time period June 13, 2017 (2017 referes to the smart meter data presented in detailed in Chapter 3), and the distribution transformer is rated at 75 kW. Coincidental peaks of the actual customer loads are expected to be above 38,916 kW (e.g. coincidental peak for load generated for the same day is 67,540 kW). With this knowledge in mind, the LB_{RTP_i} is made equal to 60% of the $l(t)$; thus, $LB_{RTP_i} = l(t)0.6 - LB_{lower_i}$ discourages scheduled appliance load changes above two times their expected value. The cost of the C_{high_i} upper bound generator should not only discourage consumption but also follow the shape of the C_{RTP_i} to discourage consumption in regions with higher prices resulting in $C_{high_i} = max(C_{RTP_i})5.0 + C_{RTP_i}$.

The LB_{lower_i} is made as described in Section 4.5.5, being if the PV forecast generation ($PV_{forecast}$) is above 5 kW $LB_{lower_i} = l(t)0.1 \times \frac{PV_{forecast}}{8.4}$. Similarly to the cost of discouraging C_{high_i} consumption, the cost for encouraging consumption (C_{lower_i}) is made as $C_{lower_i} = min(C_{RTP_i}) - max(C_{RTP_i}) + C_{RTP_i}$. Given that the PV forecast is updated every hour, and the RTP utilized for actual billing is also updated every hour, the optimization is run hourly with the knowledge of appliances that will arrive on that hour. However, the unit of time within the optimization problem is of one minute.

4.5.7 Simulation Scenarios

The simulation scenarios compare the behavior of seven distinct scenarios. The first scenario contemplates no changes performed to the schedulable appliances and will be named Queue. The second scenario contemplates the theoretical lower bound (i.e. LB), meaning that the optimization has perfect knowledge of the future RTP. The third scenario contemplates the assuming accurate forecast (i.e. AAF), meaning that the optimization utilizes the utility forecast as there RTP. The fourth scenario contemplates the partially observable Markov decision process (i.e., POMDP), meaning that the optimization problem is run multiple times (50 times) considering the forecast error with AR GARCH statistical models. The fifth scenario contemplates the same characteristic of the second scenario, but with PV lower bound to encourage self-consumption. The sixth scenario contemplates the same characteristic of the third scenario, but with PV lower bound to encourage self-consumption. The seventh scenario contemplates the same characteristic of the fourth scenario, but with PV lower bound to encourage self-consumption. The seven distinct scenarios are simulated in GridLAB-D with the PV active power curtailment controller.

4.6 Simulation Results

The simulations were performed on May 30, 2019 (i.e. the PV peak generation day). Fig. 4.6 presented the system load without any scheduling, the total available PV generation in one home, real time price, and the forecast price. Given an overview of the day under consideration. The energy consumed by the customers on May 30, 2019 is 254.89 kWh. The total available PV generation for May 30, 2019 is 756.52 kWh. From

Fig. 4.6 in this particular day appliances will have a tendency of being scheduled for periods of low PV generation, contributing in increase PV curtailment.

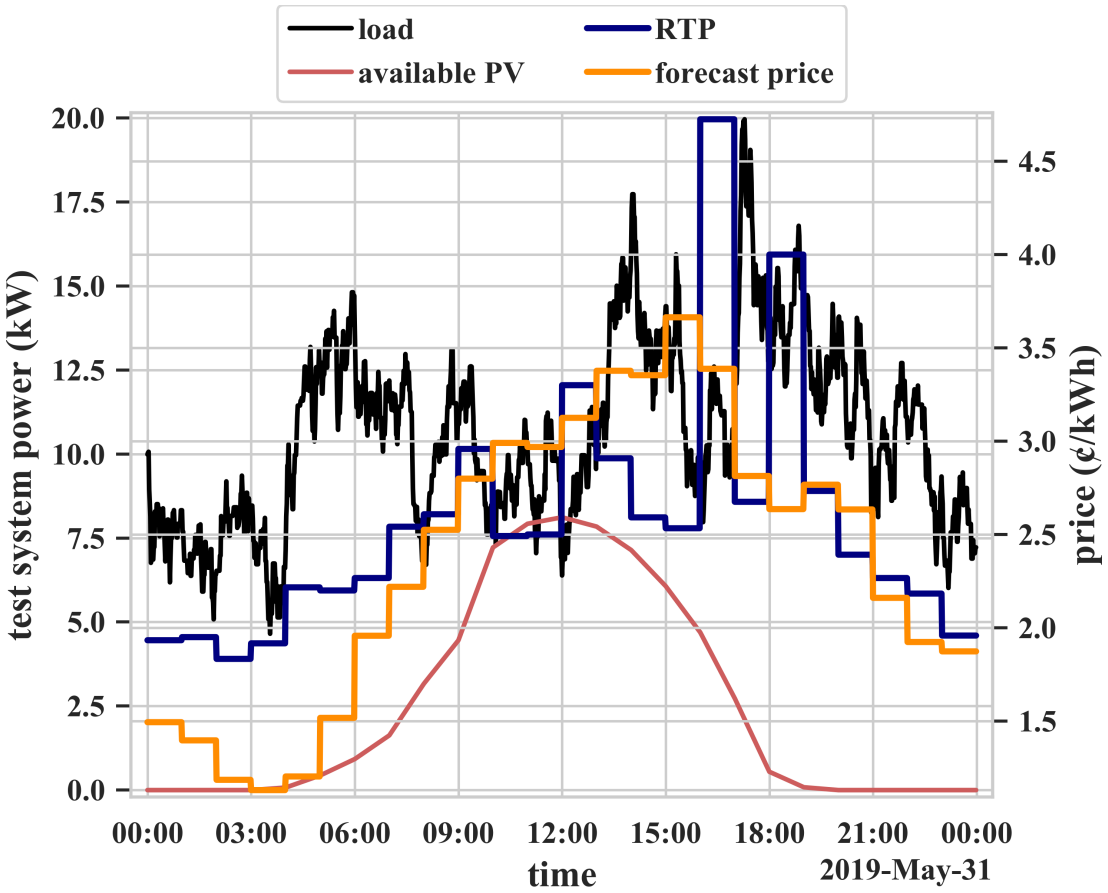

Figure 4.6. Overview of May 30, 2019 (i.e. the PV peak generation day), system load, PV generation, RTP, and forecast price. The available PV generation is for one home, since all homes have the same PV installation the total system PV installation is the home curve multiplied by twelve.

The seven distinct simulation scenarios conferred in Section 4.5.7 results summary are presented in Table 4.2. The results presented in Table 4.2 are for the behavior of the 12 homes test system. The load of all the customers are summed and multiplied by the RTP giving the energy cost no PV (¢). The load of all the customers are summed and subtracted by the sum of the system PV actual generation (i.e. considering the active

power curtailment), resulting in the total energy cost (¢). Notice that the total energy cost (¢) can be negative, meaning that the customer would be receiving an energy financial incentive. However, this does not mean that the customers would be receiving money since there are other costs than just energy as discussed in Section 4.3.6. The system curtailed PV energy (kWh) is computed by subtracting each individual PV actual generation from the total PV array availability and then aggregating it to the single value presented in Table 4.2. Notice that the scenarios that consider the PV generation with the lower bound described in Section 4.5.6, manage to reduce their PV power curtailment in relation to their counterpart. The POMDP scenario presented the largest decrease in PV curtailment of 1.33 kWh, but not considering the PV lower bound it is the one with the largest PV curtailment.

Table 4.2. Simulation results for the seven distinct scenarios

scenario	with PV lower bound	energy cost no PV (¢)	total energy cost (¢)	curtailed PV energy (kWh)
Queue	no	670.01	-944.74	191.5
LB	no	652.96	-949.39	194.49
AAF	no	666.83	-938.48	194.92
POMDP	no	660.09	-942.62	195.31
LB	yes	652.55	-951.73	193.76
AAF	yes	665.75	-941.09	194.25
POMDP	yes	658.20	-946.16	193.98

The optimization problem described in Section 4.4.6 being solved for all the scenarios is a receding time horizon where all the actions that affect the most recent hour are taken, while the other are in the waiting. This characteristic explain the apparent improve performance of the LB from without considering PV to considering PV of 0.41 ¢ for the cost of energy without considering PV. Given no scenario knows the schedulable

appliances that will be available for optimization in the next optimization period and the problem is subject to two soft constraints (i.e., the maximum "generator" capacity). The POMDP outperformed the AAF scenarios by 6.74 ¢ and 7.55 ¢, for the cases with and without PV respectively. Thus, demonstrating the value of considering the uncertainties in the decision making process. The uncertainties considered are the RTP, however the POMDP problem can be expanded to consider other sources of uncertainty. Such as, PV generation and the expected future arrival of schedulable appliances. Including more uncertainty variables to the POMDP problem does not necessarily imply that the performance will increase. It is important to keep in mind that the increase in performance by incorporating other uncertainty variables to POMDP only occurs if the forecast error is significant as demonstrated in [55]. This characteristic can also be inferred from Table 4.2, given if the price forecast were to be close enough to the RTP the results from AAF would be closer to the LB than POMDP (i.e., assuming POMDP historical data would suggest larger uncertainty than the current price forecast performance).

The changes in the system consumption, and curtailed PV power through time for May 30, 2019 is presented in Fig. 4.7, and Fig. 4.8, having the base case in relation to the scenario without considering PV, and with PV consideration, respectively. The Fig. 4.7, and Fig. 4.8 has a top plot containing the RTP and forecast price to assist in understanding the changes in consumption and curtailed power. As expected from Fig. 4.6 the appliances have a tendency of being scheduled for periods of low PV generation, contributing in increase PV curtailment. This characteristic is point out in Table 4.2 and especially in Fig. 4.7, where there is no PV self consumption incentive. Comparing Fig. 4.7, and Fig. 4.8 it becomes evident that the changes are small, however, always present for all the

scenarios.

In Fig. 4.9 gives a closer look at hour 16:00 (i.e., peak RTP) the AAF approach increased the consumption on that hour since according to the forecast price that should have been after the RTP peak price. Since POMDP considers the uncertainty of the forecast the load consumption for the hour 16:00 is much lower than AAF, and not that different from the approach that has knowledge from the future (i.e., LB). The behavior for the hour 16:00 is maintained in the cases that consider PV generation, Fig. 4.10. The comparison of the differences in between the POMDP with and without considering PV is not so visible on the Fig. 4.7, and Fig. 4.8. For example the largest difference in PV curtailment from the two POMDP scenario is 2.03 kW. The 2.03 kW PV curtailment difference happens at the time 16:15. Resampling the minute resolution PV curtailment to hourly the hour 16:00 presents a curtailment reduction of 0.86 kWh. Important to point out that some hours had the curtailment of PV increased even considering the PV on the optimization. However, on average for the day May 30, 2019 the PV curtailment is always reduced as stated in Table 4.2. A closer observation on the PV curtailment is presented on the closer look at the time periods of 10:00 to 13:00 and 15:00 to 18:00 in Fig. 4.9, and Fig. 4.8, without and with an incentive for self consumption.

Table 4.3 presents The hourly summary of the scheduling approaches. The Table 4.3 was developed in an attempt o facilitated the comparison of the scheduling approaches. Please take a closer look at the hour 16, were the POMDP is able to anticipate the likely to be RTP peak. At hour 12 the system has 95.76 kWh available PV power generation capability. Please notice that the consideration of PV generation (i.e., desire to increase self consumption) their is a increase of self consumption with POMPD at hour 12.

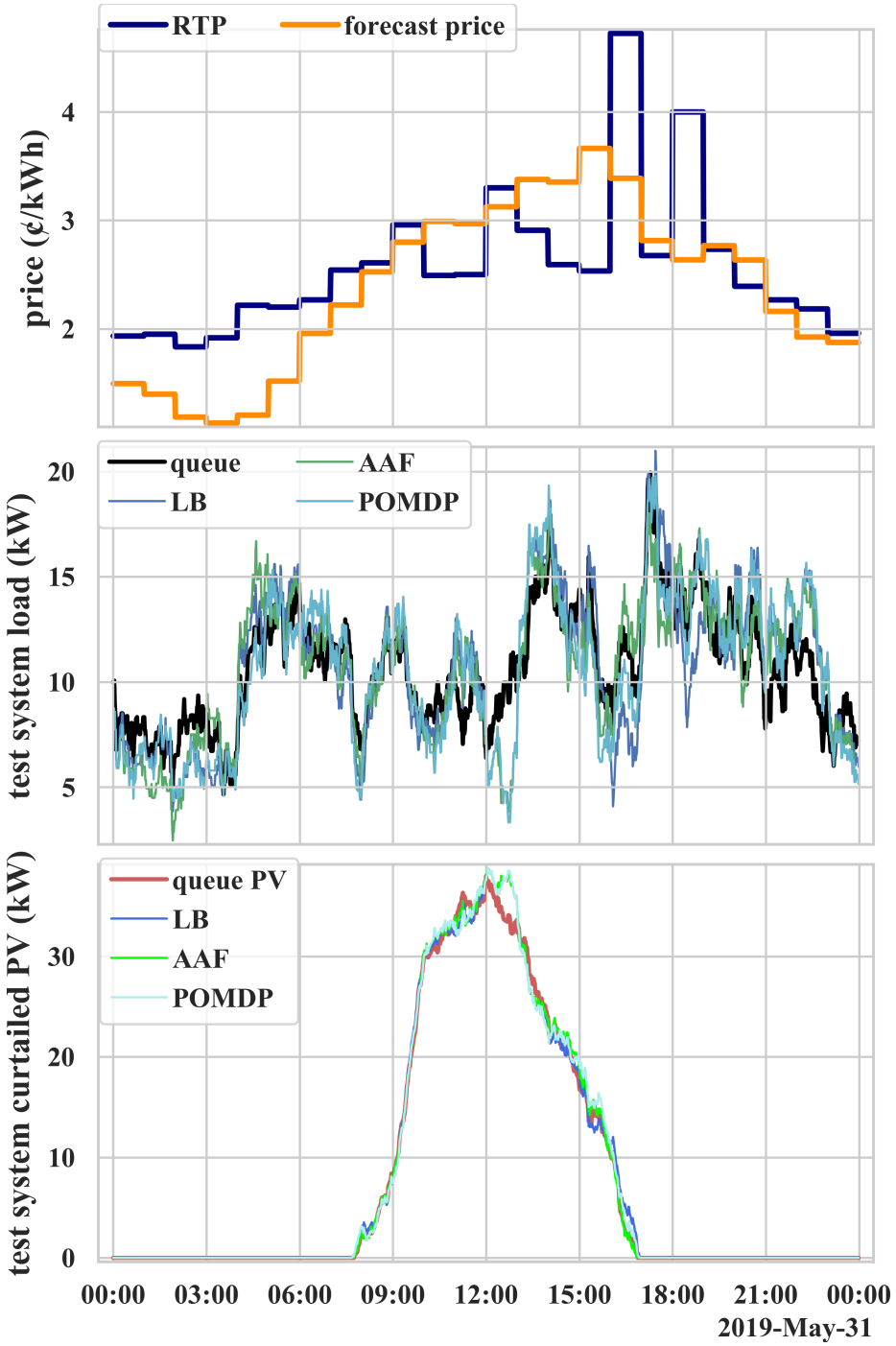

Figure 4.7. Overview of May 30, 2019 (i.e. the PV peak generation day), system load, PV generation, RTP, and forecast price. The scenario presented do not consider the PV generation (i.e., not encouraging self consumption).

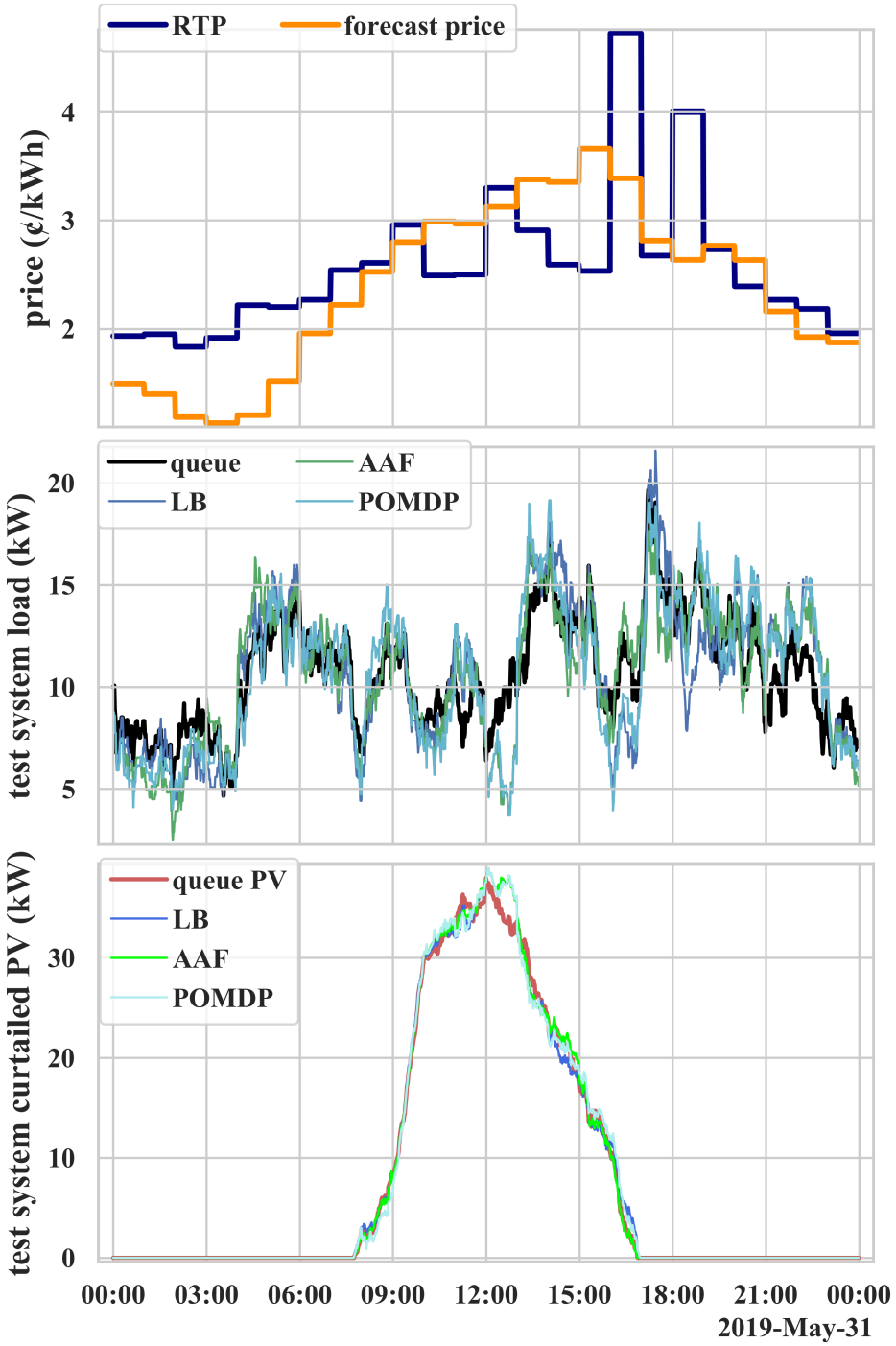

Figure 4.8. Overview of May 30, 2019 (i.e. the PV peak generation day), system load, PV generation, RTP, and forecast price. The scenario presented consider the PV generation (i.e., encouraging self consumption).

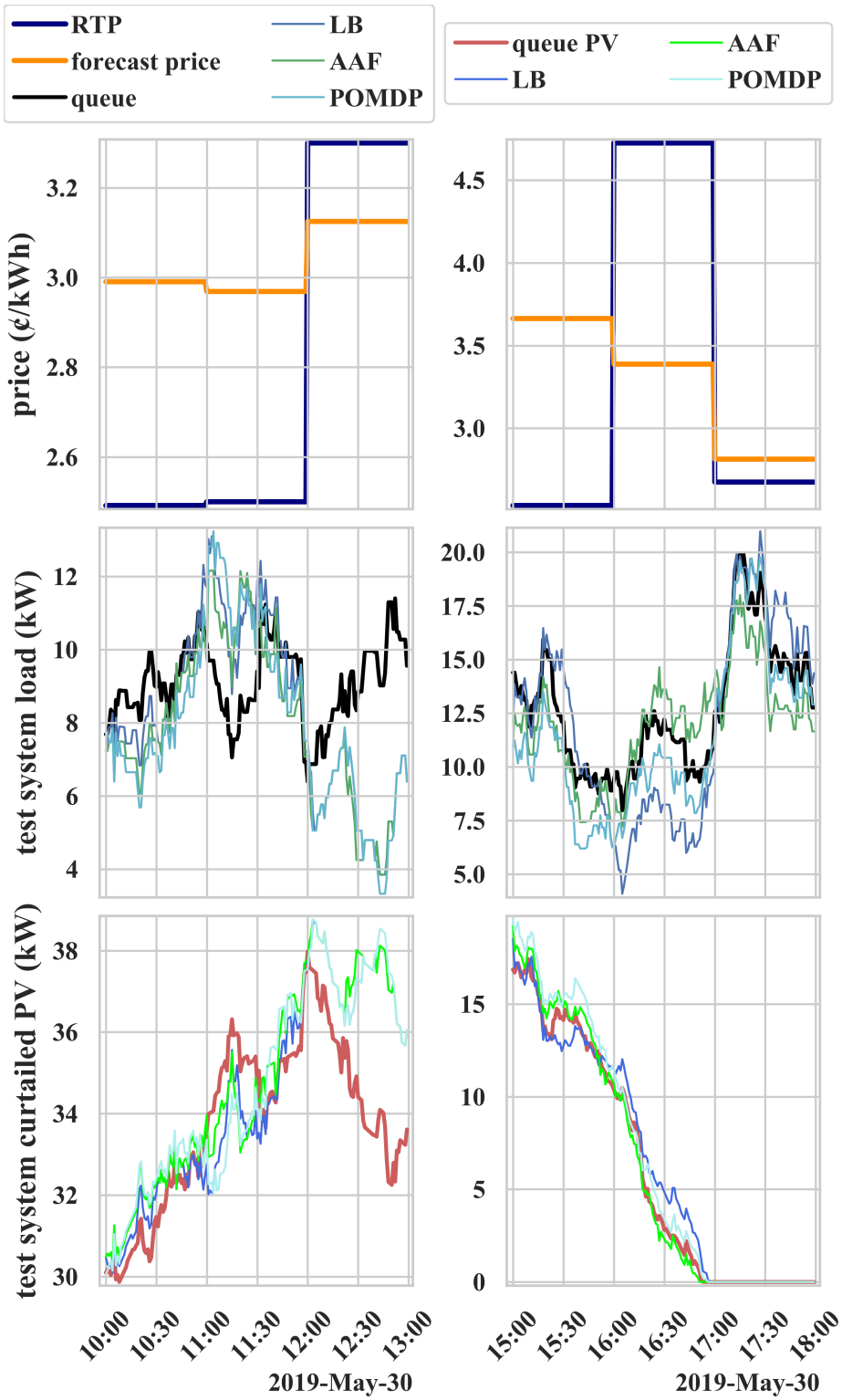

Figure 4.9. Zooming in two periods from Fig. 4.7. May 30, 2019 (i.e. the PV peak generation day), system load, PV generation, RTP, and forecast price. The scenario presented do not consider the PV generation (i.e., not encouraging self consumption).

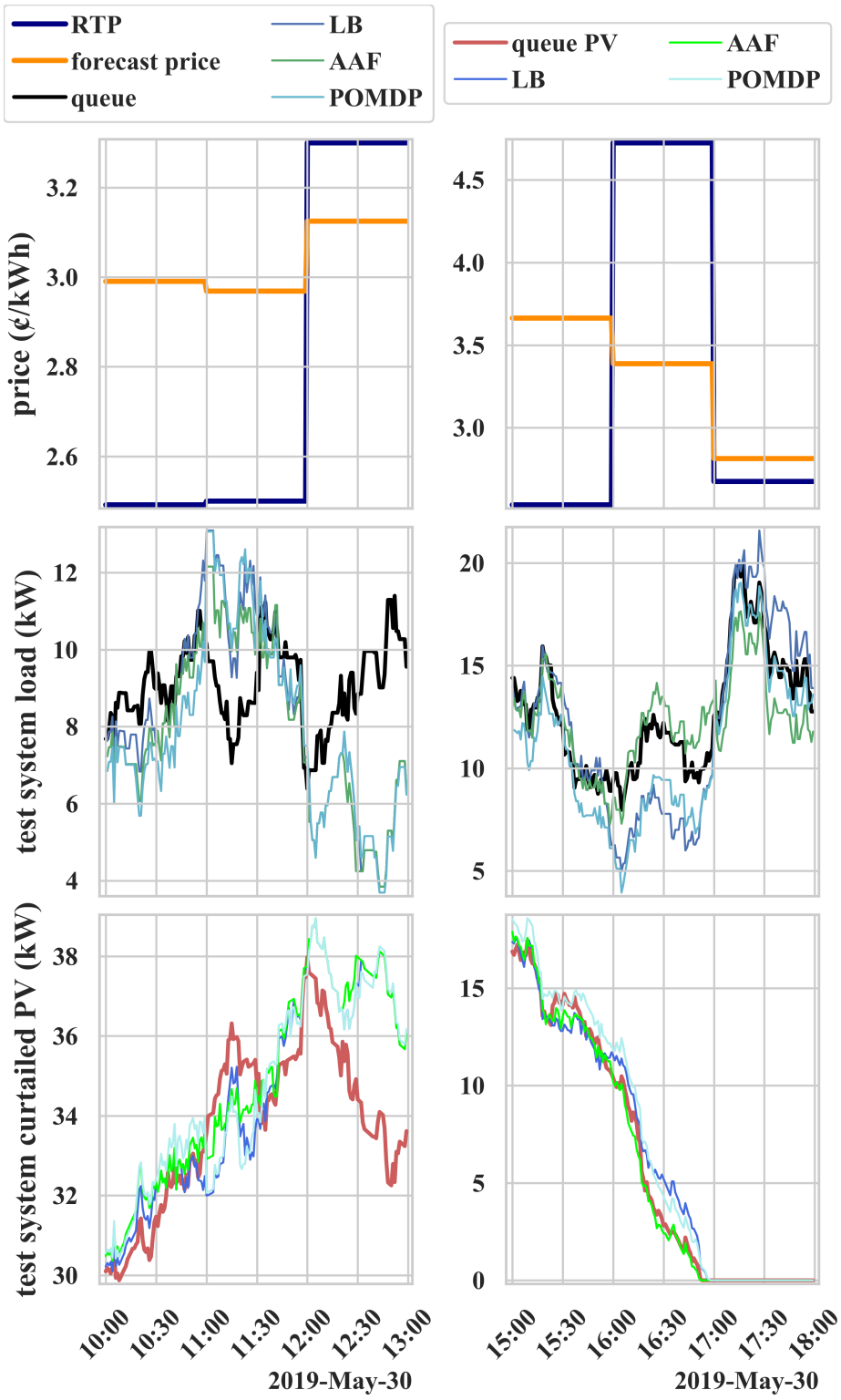

Figure 4.10. Zooming in two periods from Fig. 4.8. May 30, 2019 (i.e. the PV peak generation day), system load, PV generation, RTP, and forecast price. The scenario presented consider the PV generation (i.e., encouraging self consumption).

Table 4.3. 12 home test system hourly summary with the load, and RTP

			System load for different scheduling approaches (kWh)					
	price	queue	Not considering PV			Considering PV		
hour	(¢/kWh)	(kWh)	LB	AAF	POMDP	LB	AAF	POMDP
0	1.93	7.88	6.74	6.44	6.48	6.76	6.44	6.21
1	1.95	6.81	6.56	4.86	6.63	6.56	4.86	5.70
2	1.83	8.00	5.73	6.23	5.67	5.72	6.27	6.55
3	1.91	6.53	5.65	7.12	5.91	5.65	7.11	6.24
4	2.21	10.97	12.39	13.66	10.49	12.41	13.67	10.47
5	2.20	13.12	14.32	13.53	13.02	14.49	13.60	13.19
6	2.26	11.51	12.16	11.78	13.16	12.13	11.76	11.63
7	2.54	10.42	9.03	9.63	9.56	8.92	9.61	10.12
8	2.60	10.07	9.70	9.99	10.32	9.67	10.05	11.67
9	2.95	10.20	9.81	10.41	10.74	9.79	10.28	10.43
10	2.49	9.11	8.51	8.08	7.71	8.56	8.09	7.54
11	2.50	9.18	10.61	10.08	10.49	10.60	10.10	10.64
12	3.30	8.97	5.77	5.60	5.68	5.74	5.58	5.75
13	2.90	13.35	14.58	14.21	15.34	14.60	14.59	15.62
14	2.59	13.88	14.44	12.98	13.69	15.42	12.60	14.25
15	2.53	11.64	12.00	10.27	9.26	11.91	11.54	10.19
16	4.72	10.58	7.32	11.70	9.16	7.36	11.57	7.75
17	2.67	15.75	16.90	14.08	15.63	17.12	13.74	15.12
18	4.00	14.21	11.65	15.02	14.11	11.40	14.40	14.25
19	2.73	12.61	12.63	12.98	13.45	11.75	13.23	13.06
20	2.39	11.81	13.95	11.91	13.44	13.62	11.55	13.93
21	2.26	10.39	12.38	12.56	12.47	12.78	12.59	12.46
22	2.18	9.97	12.65	12.42	12.89	12.63	12.37	12.64
23	1.95	7.82	7.53	7.20	6.79	7.44	7.13	7.03

4.7 Discussion

The developed framework is capable of simulating HEMS with PV inverters local controllers. To the best of my knowledge, just these two characteristics make the approach presented in this chapter the first of its type. This is likely due to the fact that this simulation can only be performed by optimizing the HEMS resources and performing a dynamic simulation or the quasi-steady-state simulation considering numerical oscillations presented in [104], thus greatly increasing the complexity of the simulation. The developed framework also contains the capability to forecast PV solar generation with the statistical Markov switching model presented in [105] and the capability of performing the HEMS optimization considering the RTP uncertainty with POMDP as presented in [55]. This is different than the model in [55] that fitted the AR and GARCH only at a given time. The framework enables fitting the AR and GARCH models through time. An interesting characteristic given that the fitting of AR and GARCH models a larger importance is given on the behavior of the time series in recent time in relation to older data. This, characteristic is made possible by directly utilizing Python packages for the AR and GARCH models.

The result Section 4.5.7 presented seven distinct simulation scenarios enabling the comparison of the proposed POMDP approach. The theoretical lower bound (i.e., LB), assuming accurate forecast (i.e., AAf), and the partially observable Markov decision process (i.e., POMDP) scenarios are compared against themselves with or without the consideration of PV, and against the no scheduling. The summary of the results of the compared scenarios approaches is presented in Table 4.2. The presented results

collaborate the strategies assumptions that considering the uncertainties in RTP results in a performance increase in relation to assuming accurate forecast. The consideration of PV generation is also capable of reducing the renewable energy that has to be curtailed to avoid the problem of overvoltage. Please keep in mind the LB approach is only theoretical, and it is not possible to implement since it considers having knowledge from the future.

The presented framework demonstrates a concept that could be performed in real time. The results presented in Section 4.6 illustrate the framework's potential and are an example of the utilization of the developed synthetic residential load models in Chapters 2 and 3. The local PV inverter controller utilized is the active power curtailment; if the active-reactive droop PV inverter controller were to be utilized, it is expected that the PV curtailment would be significantly reduced as presented in [24].

4.7.1 Impact in Low Income Households

HEMS changes the load consumption to receive a financial gain of providing that flexibility. In order to provide the flexibility a home is expected to have smart appliances connected with a communication network to a controller or optimizer, as presented in this chapter. Thus, the resident may require to make the initial investment of the required structures if they are not readily available. Similarly to the installation of distributed generation such as PV, also presented in this chapter. Where the owner of the resources receives financial incentives for the generation. The impact of customer owned generation social impact is presented in [125]. The democratization of electricity systems vision for the future from [125], given that wind and solar, are available everywhere, renewable energy can be economically harnessed at small scales across the country. In [125] clams

that the larger change is not the distributed generation of renewable but the democratized network of independently-owned and widely dispersed renewable energy generators. Having the financial incentive or economic benefits as dispersed as the ownership.
In [126] presents the social impacts of community renewable energy projects, comparing two rural communities in Scottish Highlands demonstrating the resistance to change given the value the individual local communities value there current "traditional" land escape. The "destruction" of the current to open the way to the future. The different social impact given the providence of capital to the initial investment is also presented. The feeling of lack of control when the initial capital comes from abroad has the potential of increasing the resistance to change.

The democratization of the energy generation, increase of renewable generation, and improve power system performance due to load flexibility are a benefit to society as a whole. However, there is a need for the initial investment in infrastructure to directly participate in the changes. Also, participating directly or indirectly the power system is the same for all in the same region, and is affected by its changes. Thus, low income customers that are likely not to have generation nor smart appliances to change their demand based on price can be negatively impacted by the changes in the system [127]. The affordability and accessibility remain serious problems for low income households [128]. The combination of less efficient appliances, and less control over them, results in barriers to adopted demand response programs. The work in [128] presented statistical analyzes of the behavior of low income households residents. The data demonstrate significant changes in behavior by income, time of usage of energy, and available technology. The authors suggest the need to increase the knowledge of smart

meters and related technologies in low income households. An analogy to the change in electric demand by residential income, especially low income households, is presented in [129]. Where the change in the fair for the mass-transit rail system in Chicago is analyzed. Demonstrating the fight of two foresees in low income households residents that are at the same time more constrained in their budget, but also have fewer options.

In summary, HEMS approaches, and the possibility of generating energy locally, democratizing the energy generation market is amazing. However, a closer look in the impact the new technologies and approaches have on the low income households has to be given in order to not negatively affect the low income households.

CHAPTER 5 Other uses

5.1 Overview

Chapter 5 focuses on possibilities for estimating the energy utilized for space cooling and heating and making a decision for which queueing residential load model to utilize. Section 5.2 presents the beginning of the work performed for removing heating, ventilation, and air conditioning (HVAC) from the queuing load model $l(t)$ as suggested in Section 2.4.5. Section 5.3 presents a summary understanding of the queueing load models from Chapter 2, which assists in selection of an appropriate queueing load model for each given simulation or study being performed. Finally, Section 5.4 discusses the limitation of applying the developed synthetic queueing load model.

5.2 Removal of Residential Space Cooling and Heating Load from $l(t)$

Residential customers have more electric energy devices than just the appliances modeled by synthetic queueing load models. Additionally, portions of the customer load profile possess climate dependencies such as HVAC and electric water heaters. In the proposed queueing models, these are modeled as an aggregate of non-schedulable appliances rather than individual options that reflect their climate dependencies. As the energy consumption of such thermal loads change based on use and climate, the energy is not able to be directly shifted to a more opportune time (i.e. preheating or cooling a home does not imply that the same amount of energy would be used at a later time).

In the literature for modeling the HVAC load (l_{HVAC}), the building characteristics and the intrinsic non linearity of the thermal models is not considered. The simple

constraining $l^{min} < l_{HVAC} < l^{max}$ is the only requirement satisfied in the optimization problem from [17], [66]. In [66], l_{HVAC} is randomly sampled from a normal distribution where the mean is obtained from historical data and the standard deviation is given. In [17], l_{HVAC} is given and the boundaries defined as $\pm 20\%$ of the l_{HVAC}. Thus, [17], [66] present an interesting optimization problem formulation, but the load model makes simplified assumptions: load does not aggregate to system load, does not consider the change per day or season, and cannot account for customer preferences. Studies that consider customer comfort provide more reliable data [13], [16]. Thus, proper thermal models must be used and need to have their energy separated from conventional appliances and presented queuing models. The residential models presented, utilized in [13], [16], are more complex than those presented in [17], [66]. However, simplifications have been made to enable computing solutions.

Fig. 5.1 demonstrates the desired theoretical behavior of the load model. Maintaining the queueing load model arrival of appliances and having segregated at $l(t)$ from space heating and space cooling demand. Fig. 5.1 also eludes to the preferred customer temperature and outdoor temperature discussed in [17], [66].

Since HVAC loads are time and climate dependent (i.e. outdoor temperature in Fig. 5.1), demand response in residential space heating and cooling is performed by temporarily altering the temperature set point (i.e., the comfort temperature in Fig. 5.1). Not utilizing an amount of energy at a given temperature set point with specific environment conditions, such as outdoor temperature, solar irradiance, and humidity, does not imply that the same energy will be used in a different time period with distinct conditions. Furthermore, the indoor temperature is not only affected by the outdoor

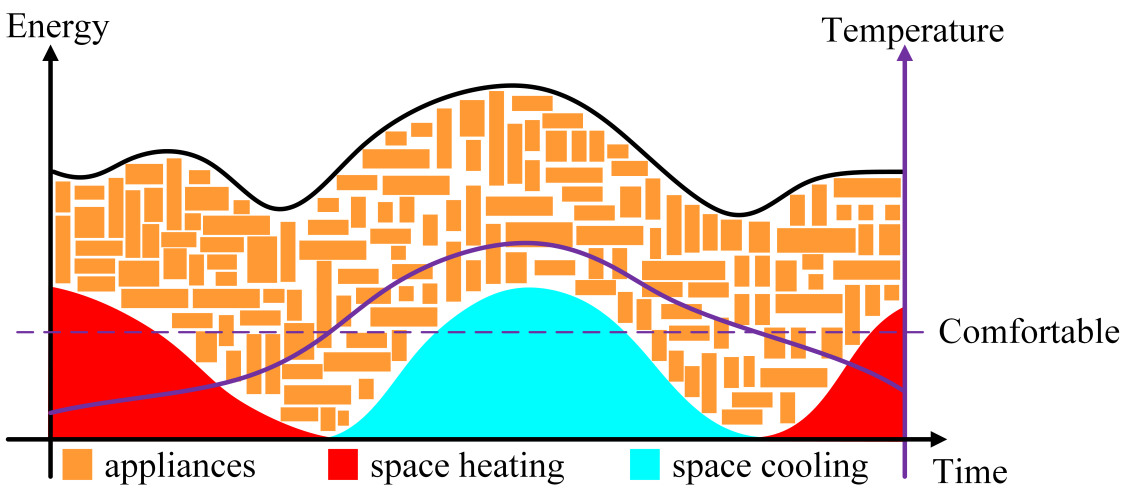

Figure 5.1. A overview of the HVAC removal is presented. The HVAC residential consumption in relation with the outdoor temperature and the temperature the residents consider comfortable is demonstrated, thus, alluding to setting of the thermostat. Please keep in mind the image does not intend to demonstrate the interaction of all the variables that contribute to HVAC consumption.

climate, temperature, and humidity, but also by the internal heat contributions, specifically inhabitants and appliances.

Detailed thermal models of residential buildings are available with the software EnergyPlus [130], which provides detailed energy requirements for heating, ventilation, and air conditioning models. EnergyPlus considers the detailed geometry of house, weather data, internal loads, temperature set points, infiltration and appliance schedule for the calculation of energy consumption in the house. EnergyPlus allows experiments to schedule the load in hourly intervals using the schedules in EnergyPlus and to create a sub hourly schedule of the appliance using an external interface.

The remainder of this section is organized as follows: Section 5.2.1 provides an approach for removing space cooling loads from $l(t)$. The approach is then tested making used of EnergyPlus home models in Section 5.2.2. Section 5.2.3 presents a discussion on

removing space heating from $l(t)$. Concluding remarks on the challenges of removing HVAC from $l(t)$ are discussed in Section 5.2.4.

5.2.1 Removal of Residential Space Cooling Load from $l(t)$

Not all appliances can be simply shifted for a different time, e.g. cooling appliances (CA). CA correspond to 17.5% of residential yearly energy use or 6.65% of total yearly energy use in the US [10]. Demand response in residential space cooling is performed by temporarily altering the temperature set; however, not using that amount of energy at a specific condition with a set temperature does not imply that the same energy will be used in a different time. For this characteristic, the cooling loads should be removed from the queueing load models $l(t)$.

Space cooling loads cannot be simply removed from $l(t)$ by utilizing the (2.10), i.e., $B_l(t) = l(t) - B(t)$, as presented in Section 2.4.5. The indoor temperature of a home is affected by the internal heat gains such as inhabitants and appliances. This means that some of the appliances are heating the home and a portion of them are cooling the home at the same time. Thus, for a given home with $l(t)$ is simulated on EnergyPlus and resulting in CA the $B_l(t)$, it cannot be computed directly by subtracting. Since the new $l(t)$ would not heat the home as much as the previous, less CA is required for cooling the home. An integrative model is required for removing the CA from $l(t)$ and also a model for estimating the amount of CA since running EnergyPlus multiple times for thousands of customers is not possible.

Generalized linear model (GLM) can be utilized with predictors that have error distribution models other than a normal distribution. EnergyPlus climate data has over 30

variables. Thus, it is necessary to choose which variables to use in fitting the GLM. In order to select the predictors in fitting the GLM, step wise selectors are utilized. The step wise selector adds and removes predictors to evaluate their impact on the model in relation to their statistical significance and/or goodness of fit. The R function made use of the Akaike information criterion (AIC), which considers the quality of the model (goodness of fit) in relation to the model complexity. AIC deals with both the risk of over-fitting and the risk of under-fitting. The interactive process of removing the cooling load from $l(t)$ is presented in Algorithm 1. Here, Φ is the climate data, k is a constant small value to slowly remove the CA, and $l_i(t)$ refers to the reference load at iteration i; thus, $\sum l_0(t)$ is the original $l(t)$. The Algorithm 1 stops subtracting small portions of $l_i(t)$ until the energy of the $l_i(t)$ and $CA_i(t)$ are equal or smaller than the energy of $\sum l_0(t)$. The effectiveness of this approach is presented in Section 5.2.2.

Algorithm 1: Interactive process for removing the space cooling load from the expected customer load $(l(t))$.

```
i = 0
l_i(t) = l(t)
CA_i(t) = ∞
while Σ l_0(t) ≤ Σ l_i(t) + CA_i(t) do
    i = i + 1
    CA_i(t) = MODEL_{GLM_cooling}(l_{i-1}(t), Φ)
    l_i(t) = l_{i-1}(t) − CA_i(t)k
end
B_l(t) = l_i(t)
```

5.2.2 Results for Removing the Space Cooling Load from $l(t)$

The openly available hourly load data ($C_L(t)$) chosen for the simulation was Commonwealth Edison Company (ComEd) [67] for the year of 2014. The load data is scaled down to make the expected individual home load with equation (2.2). The ComEd

company provides electric service to approximately 3.8 million customers across northern Illinois, or 70% of the state's population [131], and was chosen for the simulation since the region also possesses openly available climate data, which is necessary for the building model in EnergyPlus. Table 5.1 presents the characteristics of the home being simulated on EnergyPlus.

Table 5.1. House Parameters

Model attribute	**Parameters used**
Area	1517 ft^2
No. of floors	single
floor plan	3 bedrooms
HVAC system	Electric resistance heating
Window to wall ratio	7%
Glazing layer	2
Glazing material	low-e-glass
Solar heat gain coefficient	0.3
Location and weather file	Chicago, IL

The warmest day for the year of 2014 in Chicago, IL is chosen for the testing the proposed approach. Fig. 5.2 presents the total reference $l(t)$, which is in an hour resolution to the complete simulation output. The complete simulation output refers to the generated queueing load model data with $B_l(t)$ and running the simulation on EnergyPlus. As demonstrated in Fig. 5.2, the complete simulation output follows the original $l(t)$ having separated space cooling from the appliance model. Fig. 5.3 makes the comparison of the three reference curves: first, queueing original reference $l(t)$ with generated queueing load using $B_l(t)$ plus the space cooling. Second, queueing reference $B_l(t)$ with generated queueing load using $B_l(t)$. Third, the last $CA_i(t)$ from the Algorithm 1 with the space cooling from EnergyPlus to demonstrate that the generated curves follow their respective references.

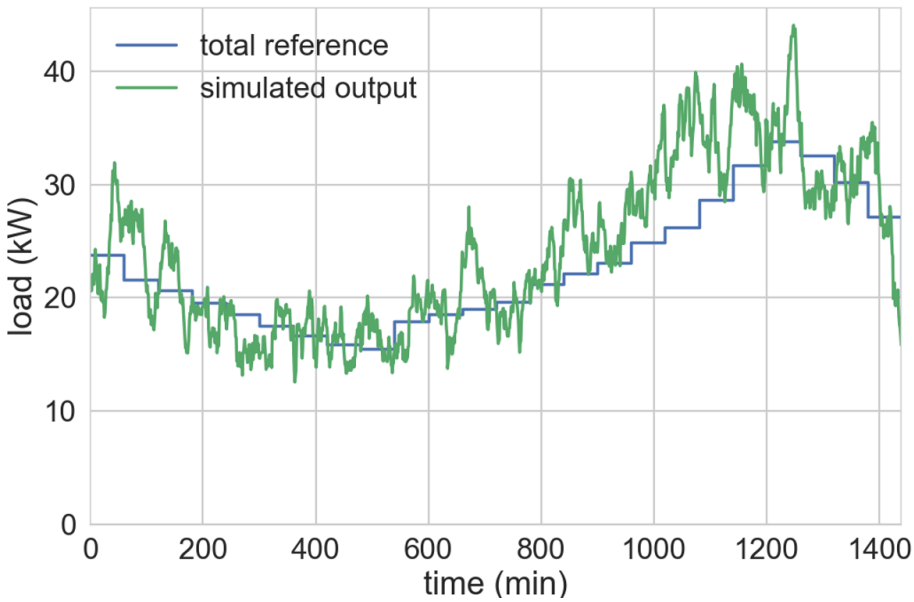

Figure 5.2. Queueing original reference $l(t)$ with the complete $B_l(t)$ plus the space cooling from EnergyPlus.

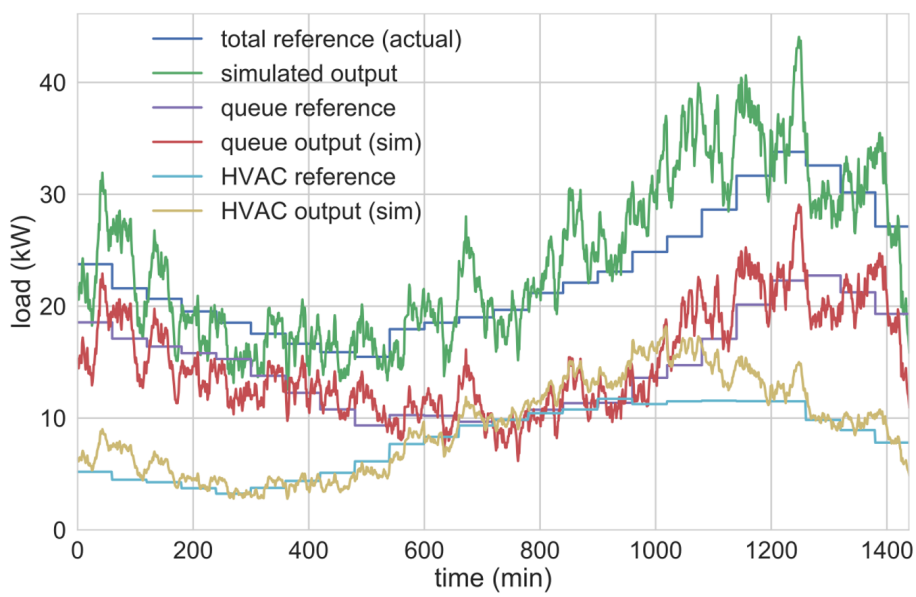

Figure 5.3. Comparison of the three reference curves: first, queueing original reference $l(t)$ with generated queueing load using $B_l(t)$, plus the space cooling. Second, queueing reference $B_l(t)$ with generated queueing load using $B_l(t)$. Third, the last $CA_i(t)$ from the Algorithm 1 with the space cooling from EnergyPlus.

5.2.3 Removal of Residential Space Heating from $l(t)$

The indoor temperature of a home is affected by the internal heat gains (inhabitants and appliances). Section 5.2.1 demonstrated an interactive process for slowly removing the CA from $l(t)$. Thus, it is not possible to remove the space heating from $l(t)$ given that the appliances running during cooling are assisting in heating the home. If a similar approach was utilized for space heating, for example, the amount of required heating load would increase at every interaction. In order to consider the space heating separately, it is required to go back to the original curve where $l(t)$ is generated and remove the space heating from the openly available hourly load data from any distribution company $C_L(t)$. Unfortunately, this results in a loop since the $l(t)$ for a home must be known to simulate the residential model in EnergyPlus and $l(t)$ changes significantly by removing the space heating from $C_L(t)$.

A completely different approach is required to enable the removal of space heating from $C_L(t)$. Understanding the challenges this implies requires recalling how the number of homes for every load node was computed in Chapter 3. The knowledge of $C_L(t)$ and the expected energy consumption for a home is utilized, which means that not only will the $l(t)$ change with the removal of space heating from $C_L(t)$ the number of homes would also change. An approach demonstrating some promise would sequentially generate data for a single home to segregate the expected energy consumption from space heating followed by updating the $C_L(t)$ until the percentage of homes with electric space heating had been generated. Unfortunately, the sequential removal of space heating for the winter period would result in negative values for $C_L(t)$. For this characteristic, it was known that the

home model in EnergyPlus would have to be changed and/or a different consideration for the number of homes with electric space heating, and/or utilized the energy of the system that is expected to be utilized for electric space heating. In order to account for this characteristic the following solutions, or a combination, may be used: change the home model in EnergyPlus, change considerations for the number of homes with electric space heating, or calculate expected energy use for electric space heating. Thus, the removal of space heating from $l(t)$ is still an open question.

5.2.4 Discussion

Section 5.2 presents the beginning of the work performed for removing the residential home HVAC load from $l(t)$ to provide different treatment than conventional appliances in DR programs. Section 5.2.1 presents a promising approach for removing space cooling load. This strategy works best for warm days with periods of time during Spring and Fall facing some challenges for utilizing the model. Thus, is not only necessary to have an accurate model for removing the space cooling but it also requires a classification model to know the periods of time to perform the removal. Classification models attempted were not accurate for the periods of Spring and Fall, especially for days that have a considerable change in temperature. This is true for both space heating and cooling. Regarding space heating, other challenges remain to be overcome as presented in Section 5.2.3; however, the most significant change occurs in regard to knowing how to properly generate the $l(t)$ and how to select the number of homes.

5.3 Queueing Load Model Choice

There are three queueing load models presented in Chapter 2, i.e. $M_t/G/\infty$, $M_t/G/C$, and $M_t/G/C_t$. The proper selection of the queuing load model to be utilized is dependent on the studies being performed. The characteristics of interest to consider for each challenge and the behavior of the tree queueing load models are described in detailed in Chapter 2. However, a simplified approached can be utilized for selecting the most appropriate queueing load model depending on the studies being performed.

The simplified selection of the queueing load model can be performed by considering three characteristics. First, the number of customers being generated. Second, utilization of the customer model in a low voltage distribution network. Third, the desired to match a small level of aggregation to the system expected behavior. The number of customers refers to the time or computational resources that will be utilized for generating the customers. The queueing load $M_t/G/\infty$ does not possess the inner loop for verifying if the appliance is able to be run given the maximum power that can be served. Algorithm loops require more computational time, so $M_t/G/\infty$ is faster than $M_t/G/C$ and $M_t/G/C_t$. Also, given a lower power supplied capacity, $M_t/G/C$ is faster than $M_t/G/C_t$.

Assuming the load being generated will be utilized in a low voltage distribution system, the queueing load models $M_t/G/C$ and $M_t/G/C_t$ should be utilized. The queueing load models $M_t/G/C$ and $M_t/G/C_t$ possess an upper bound power capacity not enabling a single customer to consume an infinite amount of energy. Customers naturally have a maximum power capacity given the appliances on their electric system. However, the utilization of the queueing load models $M_t/G/C$ and $M_t/G/C_t$ for low voltage

distribution systems comes from the need to be able to solve the power flow for all periods of time. The likelihood of a customer generated with $M_t/G/\infty$ presenting periods where a single customer surpasses the system capacity limits is small. However, if the customer generated load surpasses the system limit for a single period the simulation would not be successful. Thus, in the event were a customer surpasses the system capabilities, the customer would have to be identified and the load generated as many times as required until the customer behaves as expected. For these reasons, the load models $M_t/G/C$ or $M_t/G/C_t$ should be utilized.

The customer reference curve $l(t)$ can be generated from a system level demand or nodal level demand, as presented in Chapter 2 and Chapter 3. Chapter 2 presented the three queueing load models and Chapter 3 utilized the $M_t/G/C_t$ to generate the customer load with nodal level demand. Chapter 3 demonstrates, in Fig. 2.5, the theoretical comparison of the $M_t/G/C$ and $M_t/G/C_t$. The models $M_t/G/C$ and $M_t/G/C_t$ address the issue of unfeasible peaks. Only $M_t/G/C_t$ avoids unrealistic peaks by shifting forward appliances that arrived at a valley period that is already more that filled given the defined time varying power capacity. This characteristic naturally reduces the MAPE of the generated load in respect to the aggregated demand. The system level MAPE for $M_t/G/\infty$, $M_t/G/C$ and $M_t/G/C_t$ are 2.7973%, 2.7379%, and 2.5828% respectively. Fig. 5.4, Fig. 5.5, and Fig. 5.6 present that the larger the number of customers being aggregated to a node lode, the smaller the rate of error. Each of the figure's refers to the output of a queueing load model.

Notice that in Fig. 5.6 the MAPE for the number of homes is reduced in the earlier nodes when comparing to Fig. 5.4 and Fig. 5.5. Similarly, but to a much lower degree, this

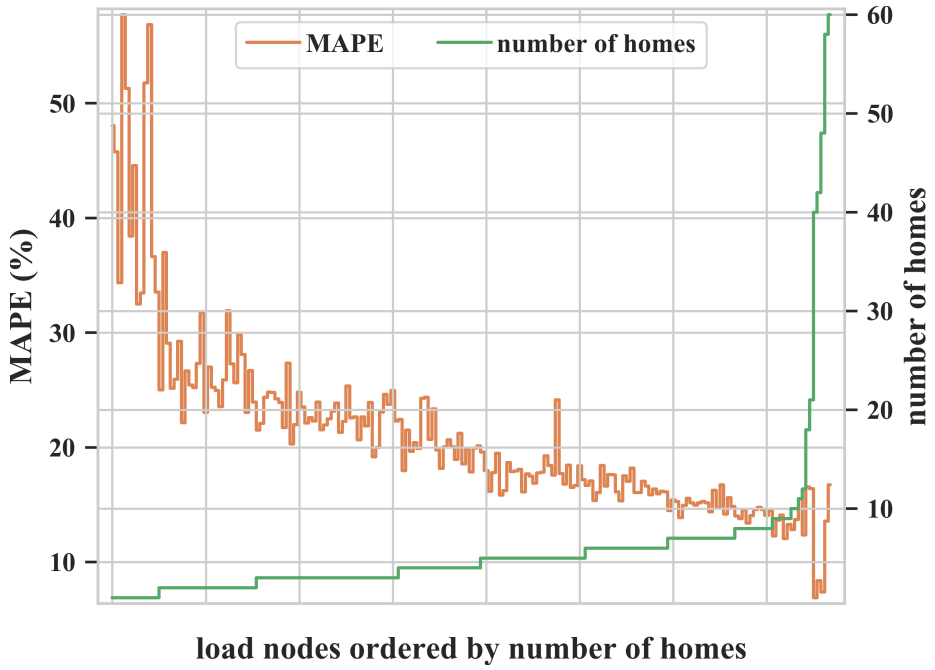

Figure 5.4. Queuing load model $M_t/G/\infty$ year MAPE by load node in relation to the number of homes.

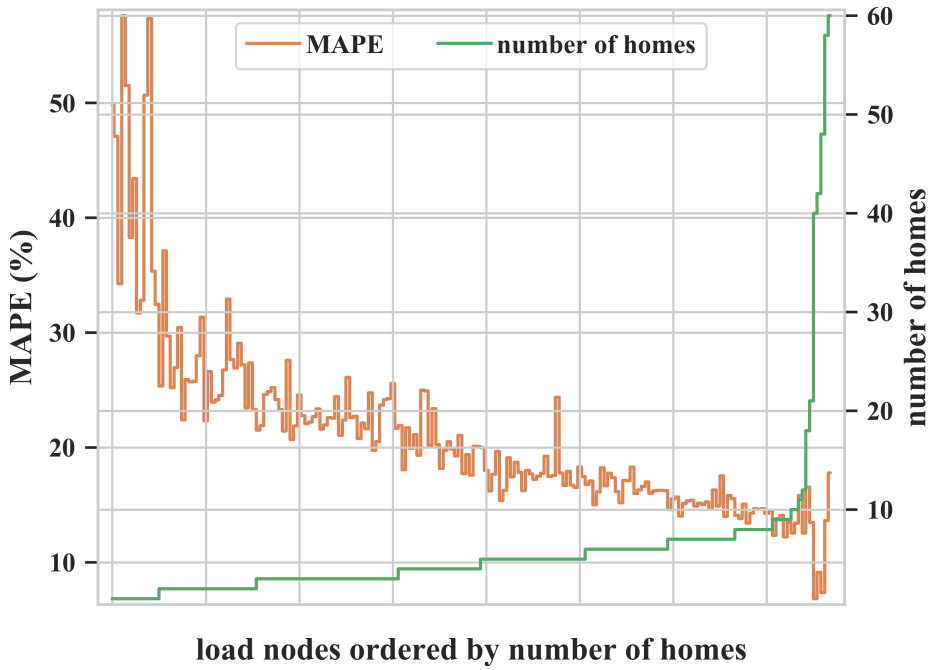

Figure 5.5. Queuing load model $M_t/G/C$ year MAPE by load node in relation to the number of homes.

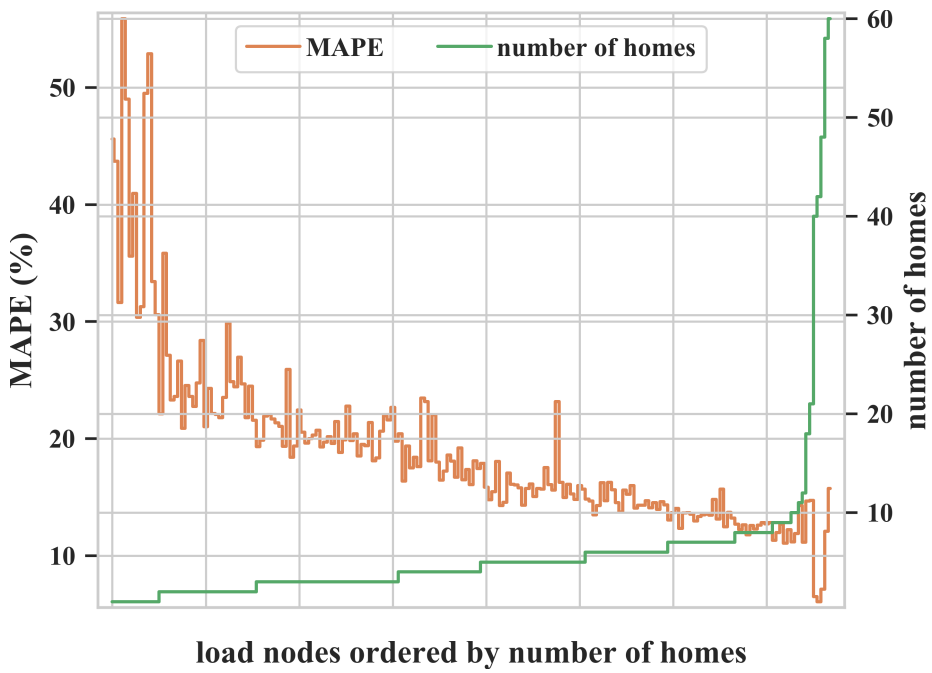

Figure 5.6. Queuing load model $M_t/G/C_t$ year MAPE by load node in relation to the number of homes. This is the same figure from Chapter 3 being placed here to facilitate the comparison

is also observed from Fig. 5.5 to Fig. 5.4. The presence of a power capacity limit improves the MAPE from the real smart meter load data to the synthetically generated queueing load model . This characteristic is expected since large deviations from the reference are not permitted at all times. In this regard, $M_t/G/C_t$ is better than the other queueing load models; however, $M_t/G/C_t$ is the model that is more computationally demanding.

5.4 Current Limitations of the Proposed Queueing Load Model

The data driven synthetic load modeling for smart city energy management studies has the potential for standardizing the studies performed on the future of smart grids. Demand response in smart grids is considered the main characteristic of the smart cities of the future. However, the proposed synthetic load model currently has limitations, the most relevant are that the:

- Appliance model lacks:
  - Frequency dependencies
    - I.e., not suitable for frequency studies.
  - Protection models
    - I.e., protection is not triggered under pre-defined conditions.
  - Harmonic distortion.
    - I.e., load do not contribute for harmonic distortion and power quality studies.
- Arrival rate of appliances is dependent only on the reference curve generated from the aggregated load. The load curve does not explicitly consider non-arriving appliances, e.g.:
  - Heating, ventilation, and air conditioning.
  - Electric vehicles.

CHAPTER 6 Conclusions and Future Work

6.1 Conclusions

The ability to control tens of thousands of residential electricity customers in a coordinated manner has the potential to enact system-wide electric load changes, such as reduce congestion and peak demand, among other benefits. To quantify the potential benefits of demand side management, synthetic load datasets that accurately characterize the system load are required. Addressing this need, data driven residential synthetic load models utilizing time-varying queueing models to characterize individual residential customers and their individual electric assets are presented and discussed in detail in this dissertation. The queueing load models presented in this paper address the challenges of unavailability and proprietary customer data by using only public available aggregated load data for a region, allowing researchers to replicate results in many studies and compare their methods to the state-of-the-art. In addition, by aggregating to a known system load curve, the economic and technical impacts of new research methods can be better evaluated. The model assumes that the aggregated distribution system behavior is known while including the stochastic nature of individual customers and their electric assets. Thus, the developed synthetic residential load model combined a top-down and bottom-up approach for modeling individual residential customers and their individual electric assets, each possessing their own characteristics. The models are general enough to incorporate other characteristics, such as non-arriving portions of customer loads (e.g. HVAC), voltage dependencies (e.g. ZIP polynomial coefficients), scheduling characteristics, and more depending on the needs of the individual researcher. The models

were validated by visualizing the differences in output between a thousand customers and by their aggregated load characteristics and follow a known system curve. As the proposed models were shown to scale in a near-linear fashion and individual customer loads can be independently generated, the methods can be used in large-scale demand side management studies (e.g. Smart City demand response) with individual customer load data that maintains the time-varying characteristics of an actual power system region. The three residential queueing load models use only publicly available data. An open-source Python tool to allow researchers to generate residential synthetic load data for their studies is made publicly available.

The developed residential synthetic queueing load models are used to create the Midwest 240-Node distribution test case system generating an appliance-level synthetic residential load for 1,120 homes for the Iowa State distribution system test case with 193 load nodes over three feeders. The Midwest 240-Node is a real distribution system from the Midwest region of the U.S. with real one-year smart meter data at the hourly aggregated node level for 2017, available in an OpenDSS model. Collecting smart meter data for 1,120 customers for a yearlong period is not trivial. Real data is subject to environmental hardships; thus, being under the influence of equipment failure, communication failure, and misused equipment. Given the environment is a distribution system in the Midwest U.S. many had access to the equipment (i.e. not a control area). The smart meter data was evaluated to identify specious behavior, which could indicate erroneous data. Two periods of strange behavior were found. The first time period is from 3,504 to 3,792 hours for the nodes 41, 154, 158, 162, and 163. The second time period is from 6,408 to 6,696 hours for the nodes 134, 140, 142, 149, 152, 180, and 183. The

strange behavior is not limited to those nodes, but unexpected behavior has a larger likelihood of being erroneous. Attempting to preserve the original smart meter data, and to reduce the chances of replacing accurate data, only the previously mentioned nodes were altered. The replacing of data makes use of a generalized linear model on the selected nodes and periods consisting of an alteration below 0.21% of the smart meter load data. The Midwest 240-Node one-year mean absolute percentage error from the smart meter to the generated is 2.5828%. The Midwest 240-Node distribution system OpenDSS model was converted to GridLAB-D to enable smart grid and transactive energy studies. The percentage wise maximum error observed on voltage magnitude from the OpenDSS to GridLAB-D model is below 0.0009%. The GridLAB-D model and the generated synthetic residential load is made publicly available. The Midwest 240-Node real distribution system with the synthetic residential load that follows the real data from smart meters is intended to be a distributed energy active consumer test system network.

The contribution of this dissertation provides the power system research community two publicly available resources: the python tool for generating residential synthetic load datasets that, when aggregated, characterizes a region of the power system and the distribution test case Midwest 240-Node which is a real system with synthetic residential load generated with the nodal information of a real yearlong smart meter data. These publicly available resources allow the researchers the capability to create their own standardized systems with no privacy concerns and a realistic standardized distribution test system to deploy their strategies.

6.2 Limitations

The data driven synthetic load modeling for smart city energy management studies has the potential for standardizing the studies performed on the future of smart grids. Demand response in smart grids is considered the main characteristic of the smart cites of the future. Being able to use the available resources in a more effective manner mixes multiple fields of study but is centered in improving the wellbeing of the residents, environment, and economy. Being a vast field of studies. Researcher that require the assumptions made on the developed model to change will face the model limitations. Important to keep in mind that one of the most interesting characteristics of the developed model is its base on actual load data.

The utilization of real load data is, in itself, one of the main benefits of the developed models, but such data may not be available for systems of interest to a given researcher. Since, the models utilize aggregated load data the unavailability is unlikely. Even if the aggregated load is not publicly available for a given system, it is possible to use a nearby system with similar climate conditions, but the knowledge of how close the replacement data was to the actual system load would be completely lost in such a scenario since a comparison would not be possible.

The usage of generic appliances, as the name states, enables the arrival of generic blocks of energy to the queueing load models. A synthetic load model assumption enables not knowing what the appliance is to perform the energy management, and demand response, meaning that the models do not know and do not attempt to represent specific appliances. This implies that appliances that have a strong tendency to be used in

sequence (e.g. clothes washer and clothes dryer) will not have such constraints to the problem being studied. If such a characteristic is necessary, the researcher would have to either incorporate the characteristic as illustrated in the Section 5.2.1 for space cooling or seek their own solution to the model limitations.

6.3 Future Work

This work presented a data driven synthetic load modeling approach for generating smart city energy management studies. Synthetic load modeling makes use of open-source data from distribution companies since it is developed for residential load modeling. The residential synthetic load data generated the test case Midwest 240-Node, which was generated having the knowledge of the nodal load of a real distribution network from the U.S. Midwest region. The test case can be utilized to perform complex energy management studies on the system 1,120 customers over three feeders. Chapter 4 illustrates how the developed synthetic load modeling can be utilized for energy management systems. The work presented in Chapter 4 can be expanded to the complete Midwest 240-Node test system and further integrated with a transmission network with a detailed model of generators, enabling studies interested in the change in price given the changes in load and incentives.

The dissertation also presents, in Section 5.2, the initial work performed to segregate space heating and cooling from the generic appliance model. Concluding the initial work would increase the relevance and complexity of the models given the natural non linearity of space heating and cooling loads in regulating the indoor temperature, which impact the customer comfort. Thus, having multiple variables competing for their

own desires and the common non linearity of the model that now would contemplate an actual home. Furthermore, more characteristics can be given to the load, such as frequency dependencies. Future work could also focus on improving the developed synthetic load models, incorporating other models and test system, and utilizing the developed models.

APPENDIX

CHAPTER A Queueing Load Model Algorithm

Algorithm 2 demonstrates the process of generating the synthetic residential load model $M_t/G/\infty$, representing the information of Fig. 2.4 in algorithm form.

Algorithm 2: $M_t/G/\infty$ residential queueing load model

Input:

- Simulation time period, $\underline{T}$ to $\overline{T}$
- Load scaling factors, b_{min} and b_{max}

Data:

- Openly available hourly load data, $C_L(t)$
- Set of appliances, ψ

```
1  t = T ; i = 0 ; Arrival = empty list
2  while t < T̄ do
3      Δt_i ← random sample exponential distribution with λ(t)
4              according to (2.1) and (2.2)
5      t ← t + Δt_i
6      if t < T̄ then
7          app ← random sample appliance from ψ
8          Arrival[i] ← app at arriving time t
9      end
10     i ← i + 1
11 end
12 return Arrival
```

Algorithm 3 demonstrates the process of generating the synthetic residential load model $M_t/G/C$ (i.e., Fig. 2.6). The initialization of variables, input, and data from Algorithm 2 is the same, thus being summarized in Line 1. Note that the procedure for the $M_t/G/C_t$ queueing load model is the same as Algorithm 3, except C is replaced with C_t,

thus the power capacity computed with (2.4) and the internal loop condition replaced by 2.5 (i.e., Line 10).

Algorithm 3: $M_t/G/C$ residential synthetic load model

```
1  Same initialization as Algorithm 2.
2  P_h(t) ← 0
3  C ← user-defined power capacity or by (2.3)
4  while t < T̄ do
5      Δt_i ← random sample exponential distribution with λ(t)
6              according to (2.1) and (2.2)
7      t ← t + Δt_i
8      if t < T̄ then
9          app ← random sample appliance from ψ
10         t_add ← t
11         while (P_h(t_add) + app_power) > C do
12             t_add ← t_add + δ
13         end
14         Arrival[i] ← app arriving at time t_add
15         P_h(t_add) ← P_h(t_add) + app_power
16     end
17     i ← i + 1
18 end
19 return Arrival
```

CHAPTER B Analyzing Time-Series Real Utility Data for a Distribution Test System

The appendix analyzes the time-series distribution test system load data (i.e., smart meter data from [93]). The data being analyzed is the first of its type. Having real year long time-series data for an actual distribution feeder. Commonly real data of a test system is not available with the exception of the test system IEEE European LV. However, the data for IEEE European LV is only for a single day. Thus, having data for a complete year is of interest. Given the privacy and technical challenges in collecting data for customers of an actual system. The year long time-series load data is analyzed and small portions suspected of being erroneous data are replaced with a generalized linear model. Less than 0.21% of the data is altered. The appendix is separated in three sections. Section B.1 presents the analyses, evaluating the smart meter data from [93]. Section B.2 presents the strategy adopted for identifying the load nodes with strange behavior, and the process for addressing it. A brief discussion on the presented approach is presented in Section B.3.

B.1 Test system load data

The authors have noticed the presence of significantly small energy consumption for the available data provided in [93]. Fig. B.1 presents the number of occurrences of nodal energy consumption of below 100, 10, and 1 Wh. The nodes that have occurrences of below 100 Wh are the 12, 32, 142, 158, 159, and 183.

The occurrences of low energy consumption for the expected home is presented in Fig. B.2. According to U.S. Energy Information Administration 2015 Residential Energy Consumption Survey [94], homes from the Midwest region have a expected yearly consumption of 9,567 kWh. Assuming the yearly consumption divided by the expected

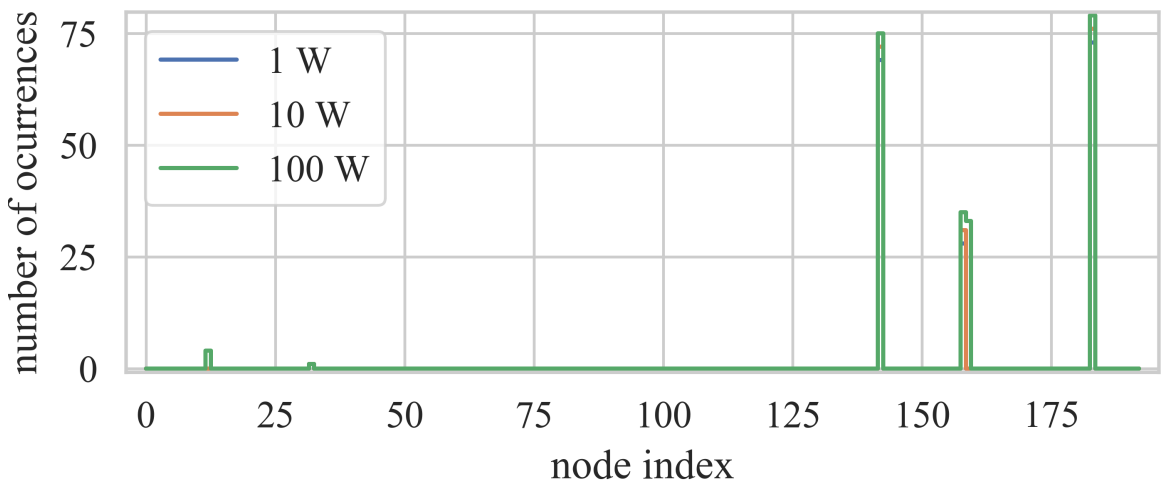

Figure B.1. Number of times the nodal energy consumption is below 100, 10, and 1 Wh.

yearly energy consumption the number of residential customers is 1,367. Which is considerably different from the 1,120 homes. However, it is also known that the consumption is climate dependent as demonstrated in [92]. Using the month of May for selecting the number of homes in load nodes the number of homes is 1,161, which is much closer from the 1,120 homes. The estimation of homes in each load node makes use of such an assumption. The energy consumption for the expected homes is computed by dividing the total nodal energy consumption by the expected number of homes. The nodes that have occurrences of below 100 Wh are the 12, 32, 38, 129, 134, 140, 142, 149, 152, 158, 159, 163, 180, and 183. However, the occurrence of low energy consumption for the node or for the estimated home energy consumption can be normal operation.

The yearly energy consumption for the load nodes 12, 32, 38, 129, 134, 140, 142, 149, 152, 158, 159, 163, 180, and 183 is furthered analyzed. During two periods of time the year long data has strange behavior. The first is in between the hours 3,504 to 3,792 in the nodes 158, and 163. The load for this nodes are presented in Fig. B.3. The second period is between the hours 6,408 to 6,696 in the nodes 134, 140, 142, 149, 152, 180, and

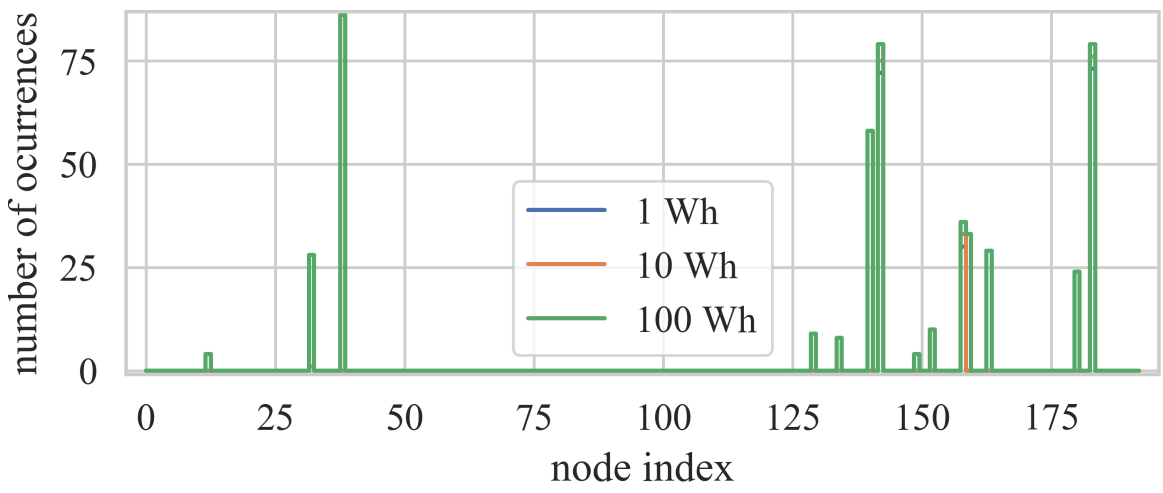

Figure B.2. Number of times the estimated home energy consumption is below 100, 10, and 1 Wh.

183. The load for this nodes are presented in Fig. B.4.

Fig. B.3 and Fig. B.4 present regions of data that for the nodes of interest that do not appear to follow there normal behavior. Fig. B.3 node 158 clearly presents two regions of constant energy consumption for over 100 hours. Furthermore, during the same period of time the load node 163 behaves strangely. Fig. B.4 nodes 142, and 183 presents two regions of constant energy consumption for over 100 hours. Similarly, during the same period of time the load nodes 134, 152, 140, 149, and 180 behaves strangely.

The load nodes 158 and 163 are considerably close to each other as shown in Fig. 3.1. Taking a closer look on the near by load nodes 157, 159, 160, 161, 164, and 165 was performed. However, only the load node 162 behaves strangely as show in Fig. B.5. Since, only 162 presents strange behavior it does not appear to be dependent on the test system location.

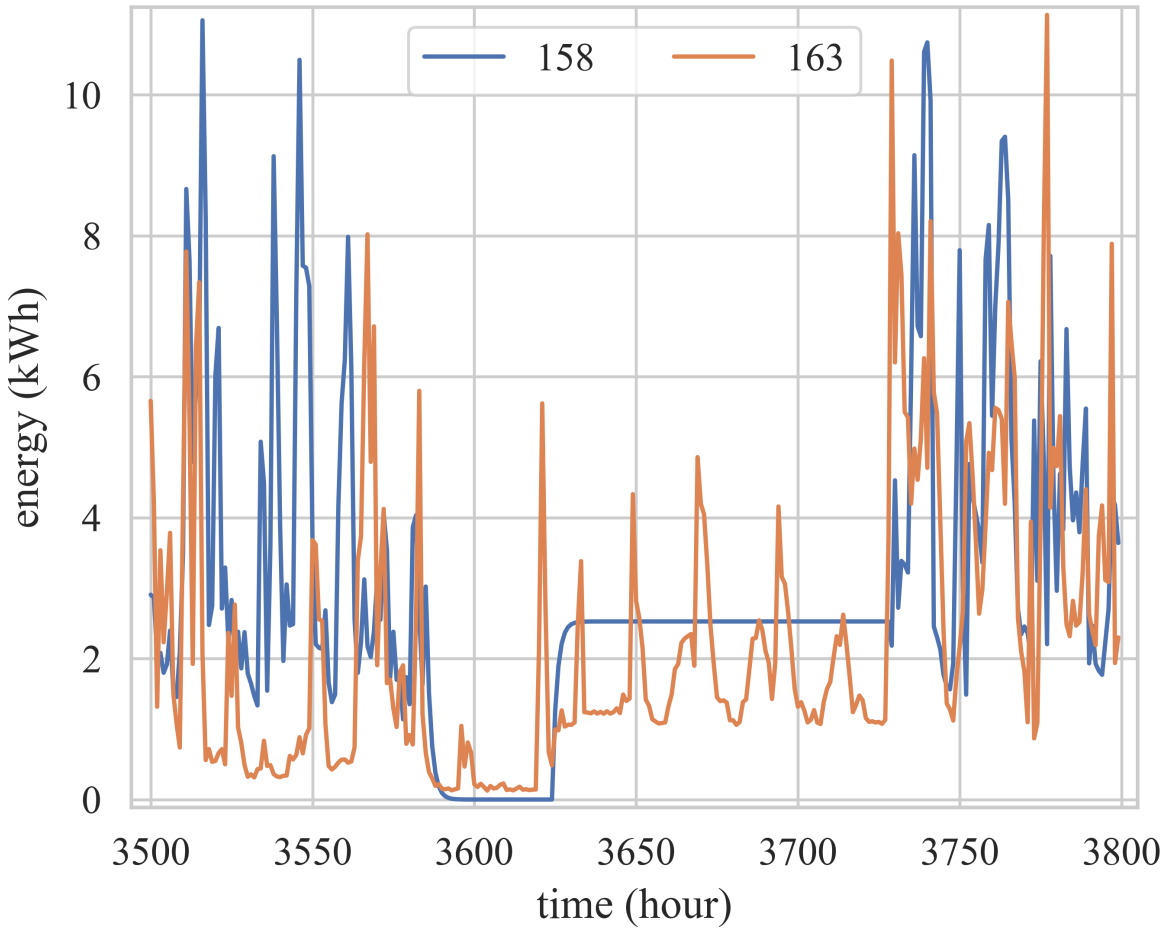

Figure B.3. Energy consumption of the load nodes 158, and 163 for the period from 3,504 to 3,792 hours.

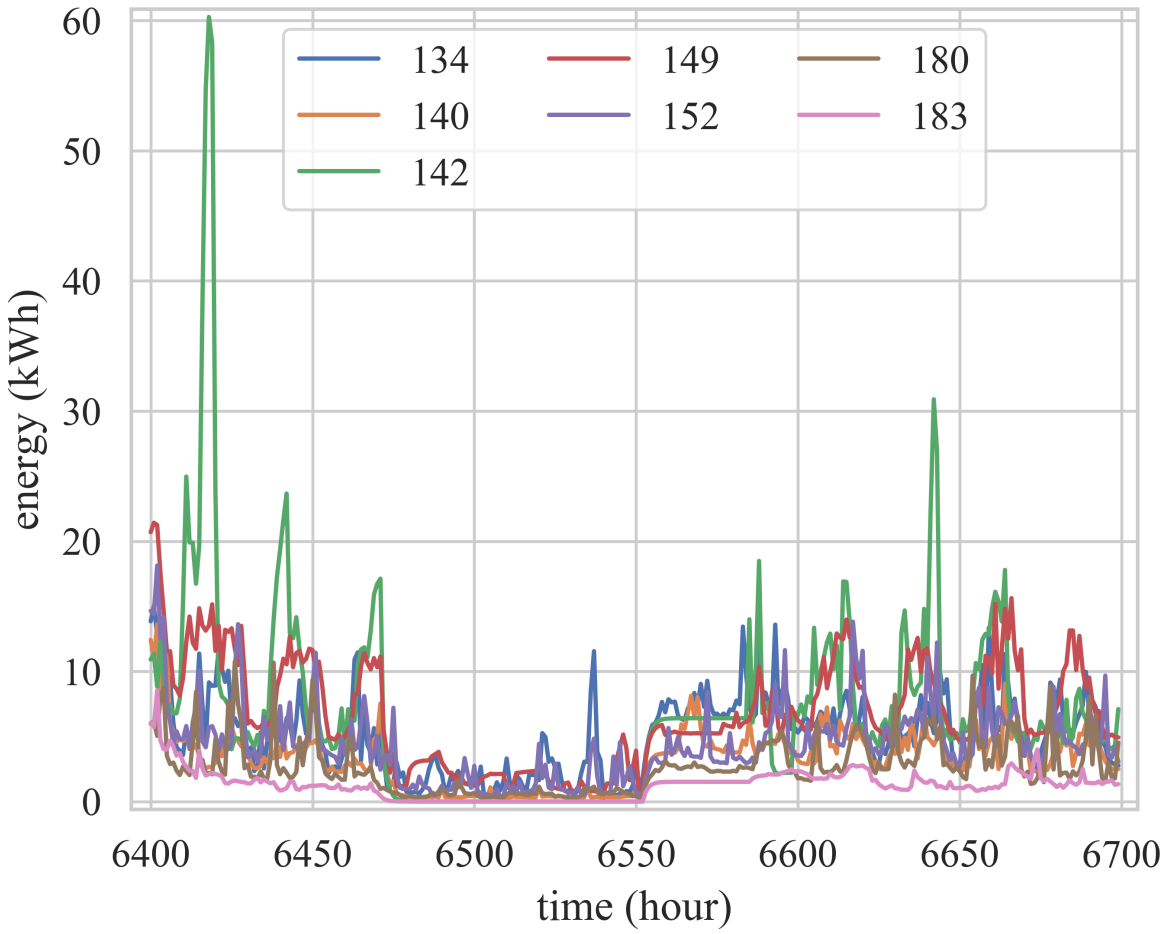

Figure B.4. Energy consumption of the load nodes 134, 140, 142, 149, 152, 180, and 183 for the period from 6,408 to 6,696 hours.

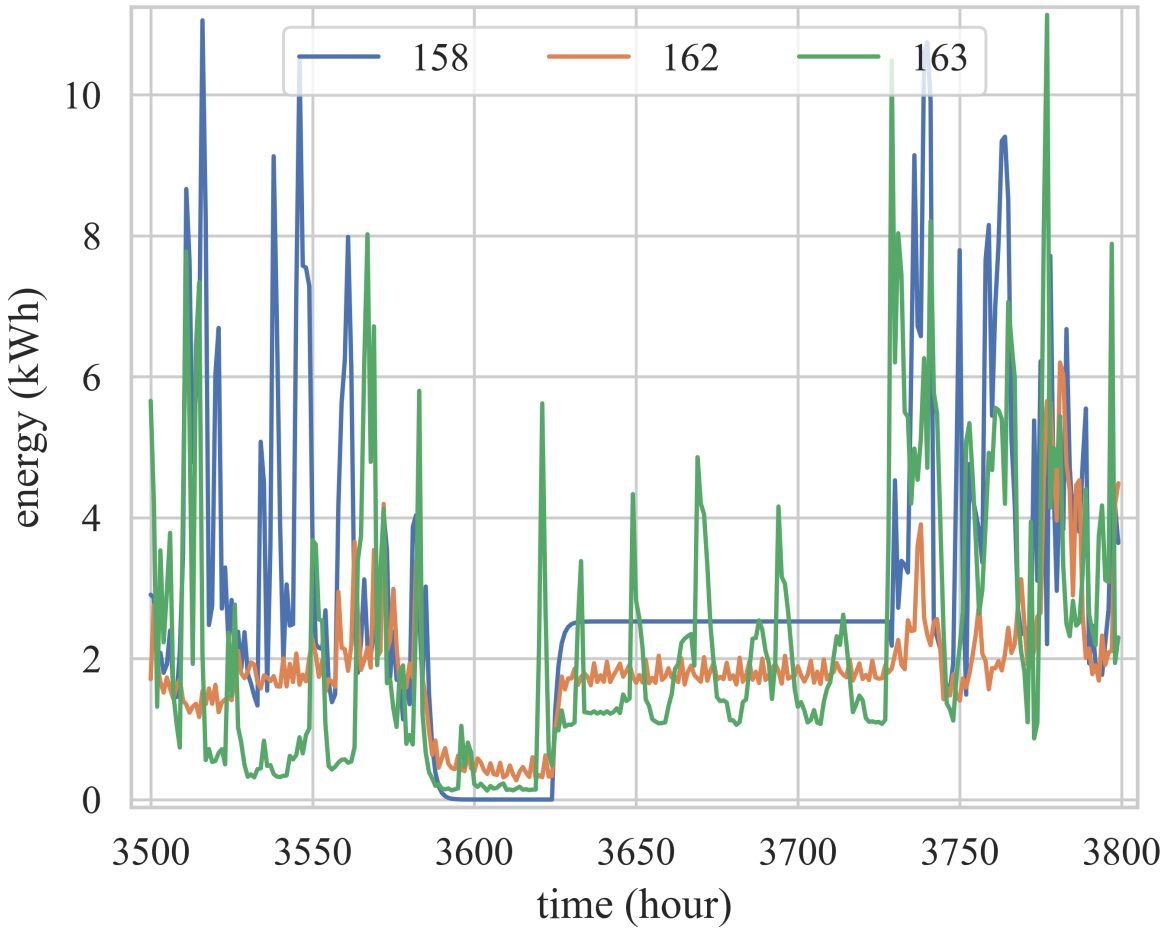

Figure B.5. Energy consumption of the load nodes 158, 162, and 163 for the period from 3,504 to 3,792 hours.

B.2 Fill in Error and Identify Nodes

In this section the generalized linear model (GLM) is made use to identify nodes with errors and to fill in the error period. In the Section B.1 two error periods have been identified, i.e., 3,504 to 3,792 hour, and 6,408 to 6,696 hour. Furthermore, some nodes have been identified and will make the baseline for identification. Given the only knowledge available is time the predictors for the model are the hour of the day and day of the week, i.e., both are classifiers with 24 and 7 possibilities respectively. The GLM equation considers the interactions of the two predictors since this greatly improves the fitted model. The model provides the average behavior of the load node for the hour of the day and day of the week. Thus, selecting the fitting regions near the period of interest is expected to provide the average behavior of the error period. The python package statsmodels [132] was utilized.

A GLM model is fitted for every node for the two periods. The fitting regions are 3,144 to 3,480 hour and 3,816 to 4,152 hour for the first period, and 6,048 to 6,384 hour and 6,720 to 7,056 hour for the second period. Thus, fitting the GLM models with 2 weeks prior and 2 weeks after the error. The first interest is to identify nodes thus the fitted models are used to compute the expected behavior on the two error periods (i.e., testing regions). The testing regions are from 3,600 to 3,700 hour and 6,450 to 6,550 hour for the first and second error regions respectively. The testing regions have been chosen imperially from the behavior demonstrated in Section B.1. Once the modes have been fitted the energy percentage error (EPE) and mean absolute percentage error (MAPE) are computed on the testing region with (B.1) and (3.1) respectively. Where t is the hour

index, A_t is the load from file, and F_t is the forecast value (i.e., the average hour and day of the week behavior from the fitted GLM model).

$$\text{EPE} = \left| \frac{\sum A_t - \sum F_t}{\sum A_t} \right| \times 100\% \tag{B.1}$$

The GLM models for the first fitted region presented a minimum, median, and maximum MAPE in p.u. of 0.1197, 0.3167, and 1.5495 respectively. The GLM models for the second fitted region presented a minimum, median, and maximum MAPE in p.u. of 0.0947, 0.2775, and 0.8846 respectively. Please keep in mind that the calculation of MAPE (3.1) is sensitive to small values of A_t, i.e., deviations for small values of A_t have a height percentage error. Fig. B.6 present the original and the GLM model on the fitted region from 3,144 to 3,480 hour and from 3,816 to 4,152 hour for the worse MAPE node of the first fitted region (node 58). Thus, illustrating the sensitive of MAPE to small values of A_t.

Evaluating the performance of Section B.1 identified nodes in regard to there EPE and MAPE boundaries to classifies problematic nodes are tested. Utilizing the data presented in Table B.1 for the performance of the GLM model for the testing region from 3,600 to 3,700 hour possible boundaries utilizing EPE and MAPE are tested and visually analyzed. The lowest EPE and MAPE from Table B.1 are utilized as starting points. The resulting classifier for the testing region from 3,600 to 3,700 hour became a combination of presenting EPE larger than 0.67 and MAPE larger than 1.75. Resulting in adding the nodes 41 and 154 which have been verified visually, as presented on Fig. B.7.

Similarly utilizing the data presented in Table B.2 for the performance of the GLM model

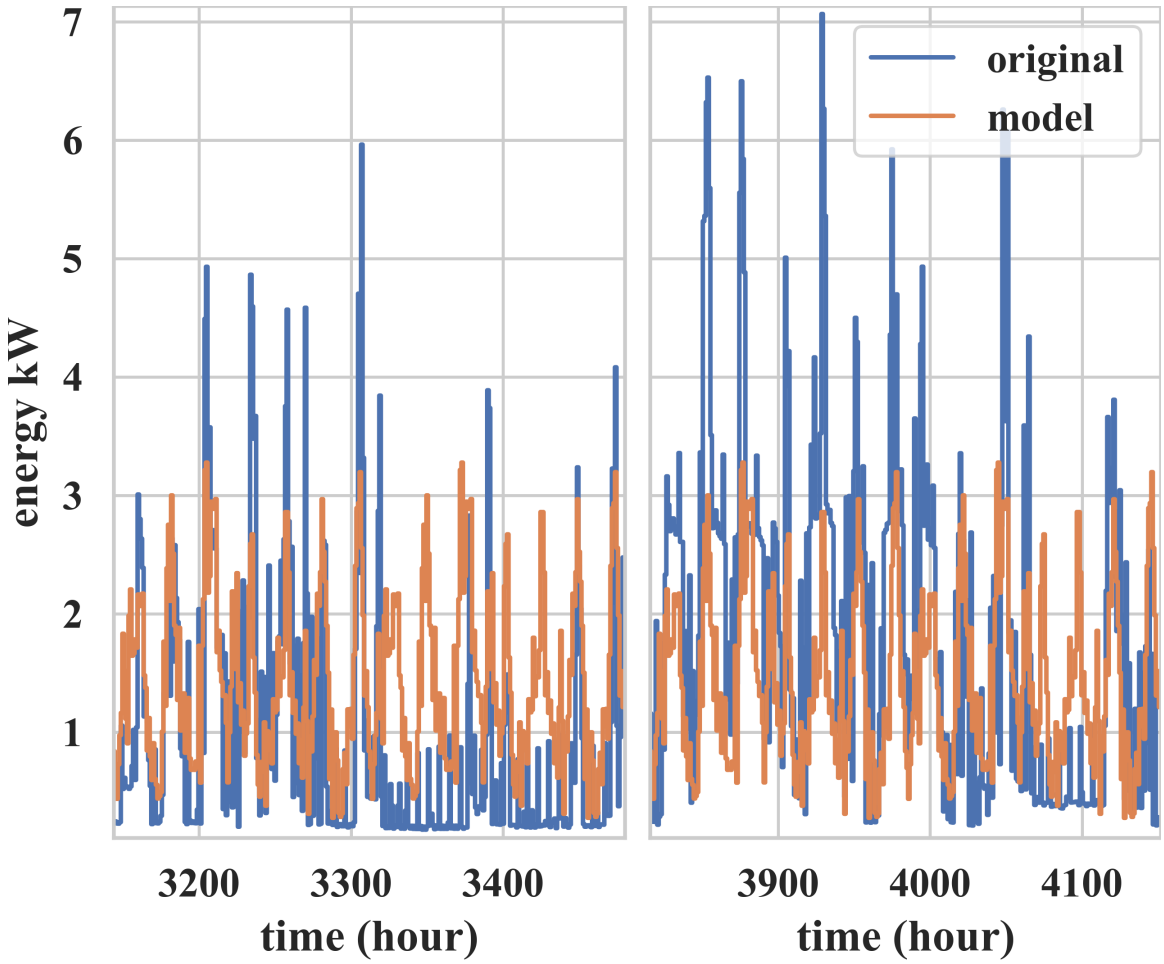

Figure B.6. Original and the GLM model energy consumption of the worse load node 58 on the fitted region from 3,144 to 3,480 hour and from 3,816 to 4,152 hour.

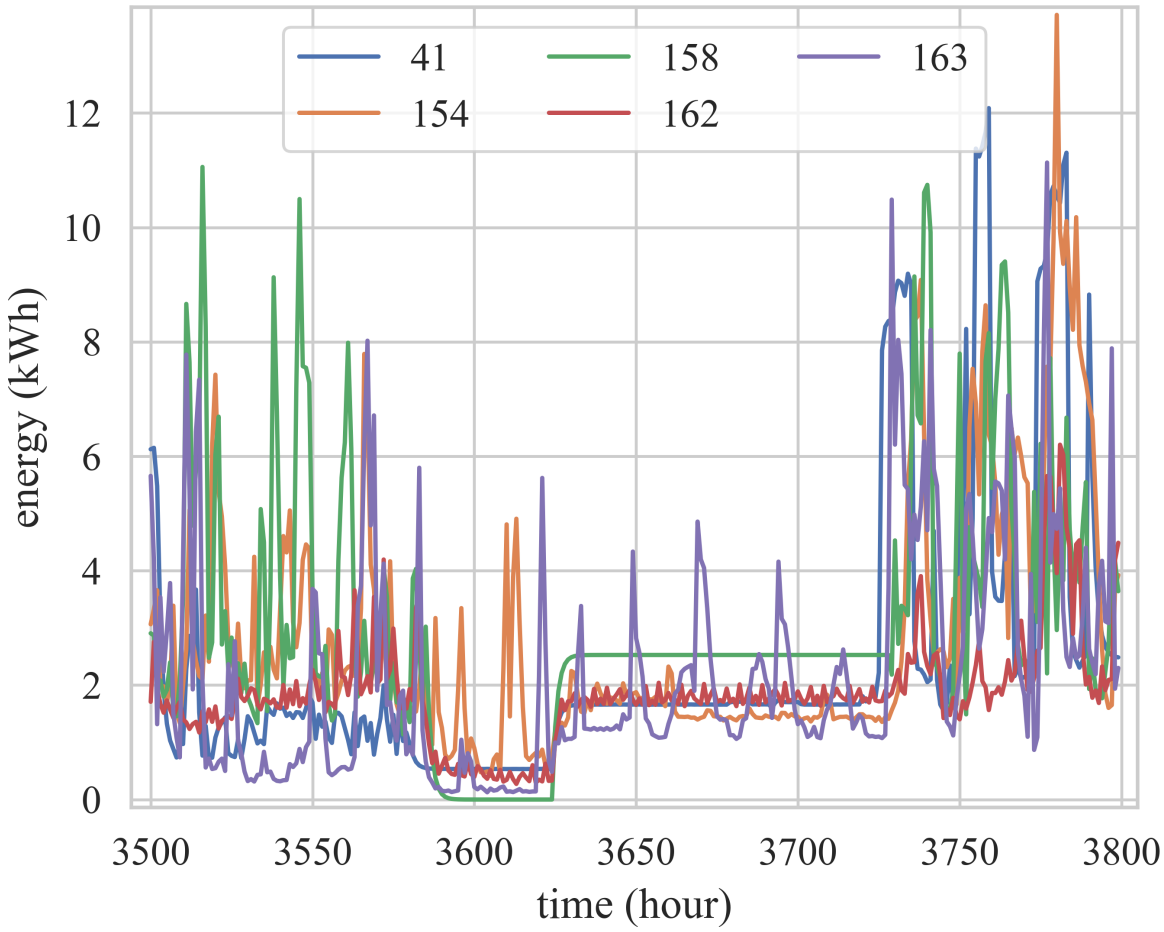

Figure B.7. Energy consumption of the load nodes 41, 154, 158, 162, and 163 for the period from 3,504 to 3,792 hours.

Table B.1. Testing region from 3,600 to 3,700 hour EPE and MAPE for Section B.1 identified nodes and added classified nodes.

Node	EPE (p.u.)	MAPE (p.u.)
158	1.0446	inf
162	0.8734	1.9536
163	1.2984	5.0156
Added node	EPE (p.u.)	MAPE (p.u.)
41	1.8870	2.9512
154	2.4025	2.9391

for the testing region from 6,450 to 6,550 hour EPE and MAPE are made. Utilizing the experience from the first testing region and visually testing multiple boundaries the same classifier was made. Nodes with problems are classified by the combination of presenting EPE larger than 1.99 and MAPE larger than 4.47. No nodes have been added.

Table B.2. Testing region from 6,450 to 6,550 hour EPE and MAPE for Section B.1 identified nodes.

Node	EPE (p.u.)	MAPE (p.u.)
134	2.2388	5.0973
140	4.1193	1.0915×10^{1}
142	7.1508	6.1649×10^{21}
149	2.3612	5.0088
152	2.1967	4.6794
180	2.6942	5.1156
183	10.3285	4.4456×10^{22}

The selection of nodes identified as errors has been performed utilizing the GLM model in regards to there EPE and MAPE. The replacing of the data suspected of being erroneous is performed by the same GLM model. The first is in between the hours 3,504 to 3,792 in the nodes 41, 154, 158, 162, and 163. The new load for this nodes are presented in Fig. B.8. The second period is between the hours 6,408 to 6,696 in the nodes 134, 140, 142, 149, 152, 180, and 183. The new load for this nodes are presented in

Fig. B.9. Comparing Fig. B.8 with Fig. B.7 and Fig. B.9 with Fig. B.4 the differences of the original data with the replaced model data are presented.

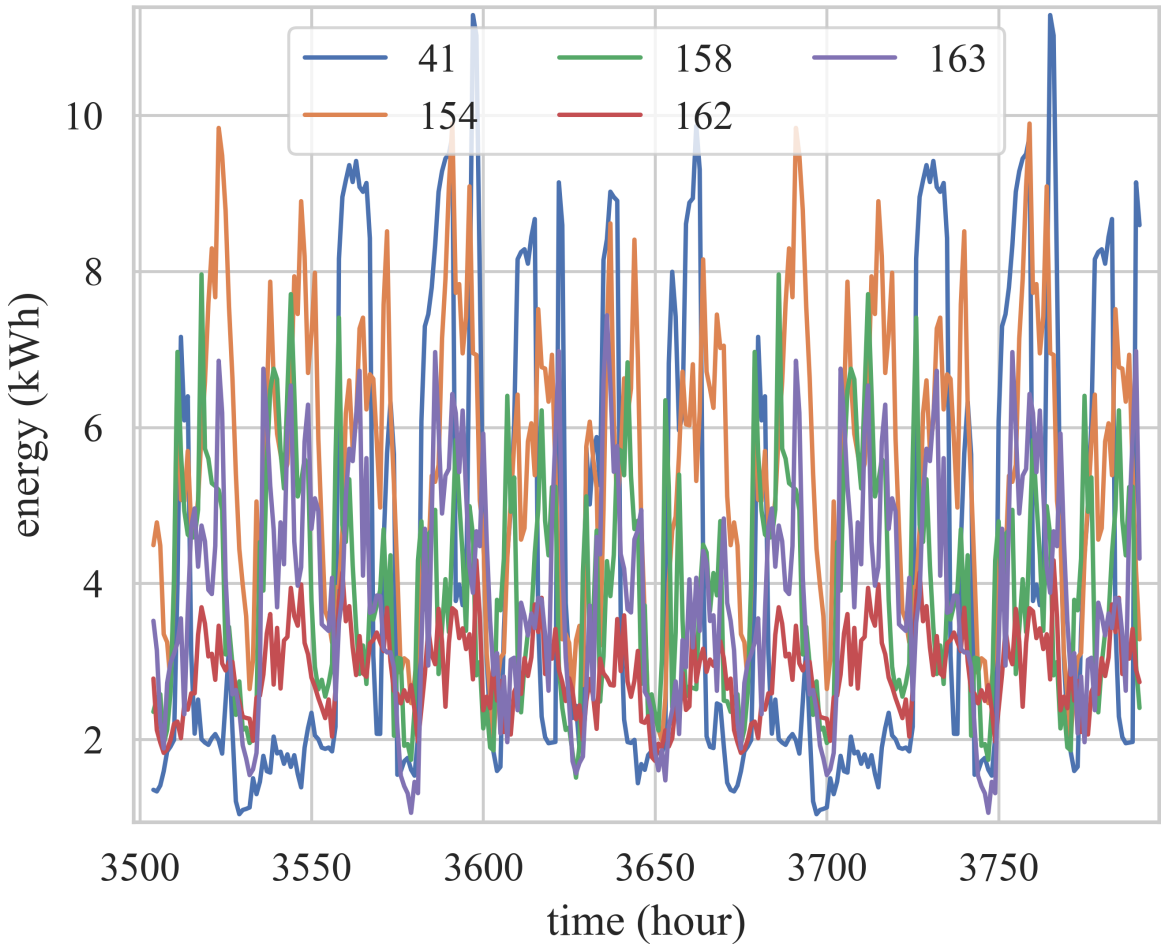

Figure B.8. Model energy consumption of the load nodes 41, 154, 158, 162, and 163 for the period from 3,504 to 3,792 hours.

B.3 Discussion

Collecting smart meters data for 1,120 customers for a year long period is not trivial. Real data is subject to the environment hardships. Thus, being under the influence of equipment failure, communication failure, and misuse of equipment. Given the environment is a distribution system in the Midwest U.S. many had aces to the equipment (i.e., not a control area). The analyze the time-series distribution test system load data

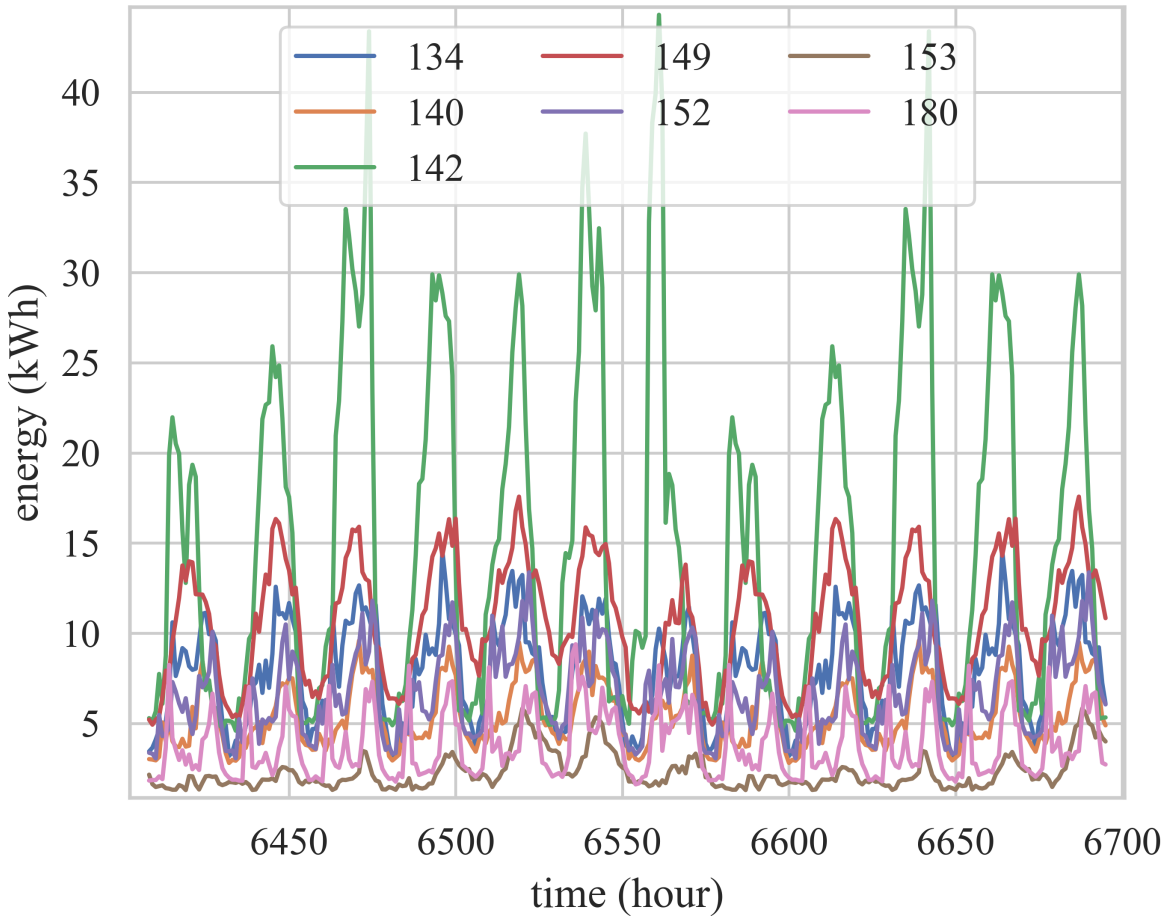

Figure B.9. Model energy consumption of the load nodes 134, 140, 142, 149, 152, 180, and 183 for the period from 6,408 to 6,696 hours.

presented here shows two time periods were some of the load nodes presents strange behavior. The first time period is from 3,504 to 3,792 hour for the nodes 41, 154, 158, 162, and 163. The second time period is from 6,408 to 6,696 hours for the nodes 134, 140, 142, 149, 152, 180, and 183. The strange behavior is not limited to nodes presented here, however, strange behavior only suggest errors on the data. The authors attempted to preserve the original data, avoiding replacing correct data from the two time periods of strange behavior. The presented replacing of data for the GLM models on the selected nodes and regions consist of an alteration lower than 0.21% of the load data.

REFERENCES

[1] U. E. I. Administration, *Most pumped storage electricity generators in the u.s. were built in the 1970s - today in energy - u.s. energy information administration (eia)*, 2019. [Online]. Available: https://www.eia.gov/todayinenergy/detail.php?id=41833.

[2] U.S. Federal Energy Regulatory Commission, *Order No. 745, Demand Response Compensation in Organized Wholesale Energy Markets*. Washington D.C.: FERC, Mar. 2011, Availble online: https://www.ferc.gov/EventCalendar/Files/20110315105757-RM10-17-000.pdf (accessed on 10 Oct. 2018).

[3] E. P. R. I. (EPRI), *Epri maps out power system of the future*, 2015. [Online]. Available: https://www.bpa.gov/news/newsroom/Pages/EPRI-maps-out-power-system-of-the-future.aspx.

[4] M. Avendaño and V. Patel, *Pandemic planning and response in electric utilities: Sce's experience - ieee smart grid*, 2020. [Online]. Available: https://smartgrid.ieee.org/newsletters/may-2020/pandemic-planning-and-response-in-electric-utilities-sce-s-experience.

[5] S. G. Liasi, A. Shahbazian, and M. T. Bina, *Covid-19 pandemic; challenges and opportunities in power systems - ieee smart grid*, 2020. [Online]. Available: https://smartgrid.ieee.org/newsletters/may-2020/covid-19-pandemic-challenges-and-opportunities-in-power-systems?utm_source=sg-monthly-may2020&utm_medium=email&utm_campaign=2020-enewsletter.

[6] A. Aziz and A. T. OO, *Covid-19 implications on electric grid operation - ieee smart grid*, 2020. [Online]. Available: https://smartgrid.ieee.org/newsletters/may-2020/covid-19-implications-on-electric-grid-operation?utm_source=sg-monthly-may2020&utm_medium=email&utm_campaign=2020-enewsletter.

[7] M. Lewis and R. Hebner, *Resilience and pandemics - ieee smart grid*, 2020. [Online]. Available: https://smartgrid.ieee.org/newsletters/may-2020/resilience-and-pandemics?utm_source=sg-monthly-may2020&utm_medium=email&utm_campaign=2020-enewsletter.

[8] T. Ding, Q. Zhou, and M. Shahidehpour, *Impact of covid-19 on power system operation planning - ieee smart grid*, https://smartgrid.ieee.org/newsletters/may-2020/impact-of-covid-19-on-power-system-operation-planning?utm_source=sg-monthly-may2020&utm_medium=email&utm_campaign=2020-enewsletter, 2020.

[9] M. Amin, *On countering multi-pronged evolving systemic threats: Covid-19 and beyond - ieee smart grid*, 2020. [Online]. Available: https://smartgrid.ieee.org/newsletters/may-2020/on-countering-multi-pronged-evolving-systemic-threats-covid-19-and-beyond?utm_source=sg-monthly-may2020&utm_medium=email&utm_campaign=2020-enewsletter.

[10] US Energy Information Administration (EIA), *Electricity Explained, Use of Electricity*, 2017. [Online]. Available: \text{https://www.eia.gov/energyexplained/index.cfm?page=electricity_use}.

[11] US Department of Energy, "Benefits of demand response in electricity markets and recommendations for achieving them," Washington, DC, Tech. Rep., 2006, pp. 1–109.

[12] V. Durvasulu and T. M. Hansen, "Benefits of a demand response exchange participating in existing bulk-power markets," *Energies, special issue on Demand Response in Electricity Markets*, vol. 11, no. 12, 21 pages, Dec. 2018.

[13] F. Elghitani and W. Zhuang, "Aggregating a large number of residential appliances for demand response applications," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 5092–5100, 2018, ISSN: 1949-3053. DOI: 10.1109/TSG.2017.2679702.

[14] T. M. Hansen, R. Roche, S. Suryanarayanan, A. A. Maciejewski, and H. J. Siegel, "Heuristic Optimization for an Aggregator-Based Resource Allocation in the Smart Grid," *IEEE Transactions on Smart Grid*, vol. 6, no. 4, pp. 1785–1794, 2015, ISSN: 19493053. DOI: 10.1109/TSG.2015.2399359.

[15] B. V. Solanki, A. Raghurajan, K. Bhattacharya, and C. A. Canizares, "Including Smart Loads for Optimal Demand Response in Integrated Energy Management Systems for Isolated Microgrids," *IEEE Transactions on Smart Grid*, vol. 8, no. 4, pp. 1739–1748, 2017, ISSN: 19493053. DOI: 10.1109/TSG.2015.2506152.

[16] E. Shirazi and S. Jadid, "Optimal residential appliance scheduling under dynamic pricing scheme via hemdas," *Energy and Buildings*, vol. 93, pp. 40 –49, 2015.

[17] M. Ye and G. Hu, "Game Design and Analysis for Price based Demand Response: An Aggregate Game Approach," *IEEE Transactions on Cybernetics*, vol. 47, no. 3, pp. 720–730, 2015. DOI: 10.1109/TCYB.2016.2524452. arXiv: 1508.02636.

[18] J. Widén and E. Wäckelgård, "A high-resolution stochastic model of domestic activity patterns and electricity demand," *Applied Energy*, vol. 87, no. 6, pp. 1880 –1892, 2010.

[19] C. M. Colson and M. H. Nehrir, "An alternative method to load modeling for obtaining end-use load profiles," in *41st North American Power Symposium*, Starkville, MS, 2009, pp. 1–5. DOI: 10.1109/NAPS.2009.5484036.

[20] M. Tasdighi, H. Ghasemi, and A. Rahimi-Kian, "Residential microgrid scheduling based on smart meters data and temperature dependent thermal load modeling," *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 349–357, 2014, ISSN: 1949-3061. DOI: 10.1109/TSG.2013.2261829.

[21] D. Zhang, N. Shah, and L. G. Papageorgiou, "Efficient energy consumption and operation management in a smart building with microgrid," *Energy Conversion and Management*, vol. 74, pp. 209 –222, 2013.

[22] M. Nijhuis, M. Gibescu, and J. Cobben, "Bottom-up markov chain monte carlo approach for scenario based residential load modelling with publicly available data," *Energy and Buildings*, vol. 112, pp. 121 –129, 2016.

[23] J. M. G. López, E. Pouresmaeil, C. A. Cañizares, K. Bhattacharya, A. Mosaddegh, and B. V. Solanki, "Smart residential load simulator for energy management in smart grids," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 2, pp. 1443–1452, 2019, ISSN: 0278-0046. DOI: 10.1109/TIE.2018.2818666.

[24] R. Mahat, K. Duwadi, F. B. Dos Reis, R. Fourney, R. Tonkoski, and T. M. Hansen, "Techno-Economic Analysis of PV InverterControllers for Preventing Overvoltage in LV Grids," in *2020 International Symposium on Power Electronics, Electrical Drives, Automation and Motion (SPEEDAM)*, 2020.

[25] L. Sherwood, *U.S. solar market trends 2013*, July 2014. [Online]. Available: {https://irecusa.org}.

[26] P. Chen and R. Salcedo and Q. Zhu and F. de Leon and D. Czarkowski and Z. Jiang and V. Spitsa and Z. Zabar and R. E. Uosef, "Analysis of Voltage Profile Problems due to the Penetration of Distributed Generation in Low-Voltage Secondary Distribution Networks," *IEEE Transactions on Power Delivery*, vol. 27, no. 4, pp. 2020–2028, 2012, ISSN: 0885-8977. DOI: 10.1109/TPWRD.2012.2209684.

[27] R. Tonkoski and L. A. C. Lopes, "Voltage Regulation in Radial Distribution Feeders with High Penetration of Photovoltaic," in *2008 IEEE Energy 2030 Conference*, Nov. 2008, 7 pp.

[28] D. Cheng, B. A. Mather, R. Seguin, J. Hambrick, and R. P. Broadwater, "Photovoltaic (PV) Impact Assessment for Very High Penetration Levels," *IEEE Journal of Photovoltaics*, vol. 6, no. 1, pp. 295–300, Jan. 2016.

[29] J. von Appen, M. Braun, T. Stetz, K. Diwold, and D. Geibel, "Time in the Sun: The Challenge of High PV Penetration in the German Electric Grid," *IEEE Power and Energy Magazine*, vol. 11, no. 2, pp. 55–64, Mar. 2013.

[30] J. M. Guerrero, F. Blaabjerg, T. Zhelev, K. Hemmes, E. Monmasson, S. Jemei, M. P. Comech, R. Granadino, and J. I. Frau, "Distributed Generation: Toward a New Energy Paradigm," *IEEE Industrial Electronics Magazine*, vol. 4, no. 1, pp. 52–64, Mar. 2010.

[31] E. Demirok, D. Sera, R. Teodorescu, P. Rodriguez, and U. Borup, "Clustered PV Inverters in LV networks: An Overview of Impacts and Comparison of Voltage Control Strategies," in *2009 IEEE Electrical Power & Energy Conference (EPEC)*, Oct. 2009, 6 pp.

[32] J. P. Lopes, N. Hatziargyriou, J. Mutale, P. Djapic, and N. Jenkins, "Integrating Distributed Generation into Electric Power Systems: A review of Drivers, Challenges and Opportunities," *Electric Power Systems Research*, vol. 77, no. 9, pp. 1189–1203, July 2007.

[33] C. L. Masters, "Voltage Rise: Rhe big issue when connecting embedded generation to long 11 kV overhead lines," *Power Engineering Journal*, vol. 16, no. 1, pp. 5–12, 2002, ISSN: 0950-3366. DOI: 10.1049/pe:20020101.

[34] M. E. Elkhatib, R. El-Shatshat, and M. M. A. Salama, "Novel Coordinated Voltage Control for Smart Distribution Networks With DG," *IEEE Transactions on Smart Grid*, vol. 2, no. 4, pp. 598–605, 2011, ISSN: 1949-3053. DOI: 10.1109/TSG.2011.2162083.

[35] O. Homaee, A. Zakariazadeh, and S. Jadid, "Real-time voltage control algorithm with switched capacitors in smart distribution system in presence of renewable generations," *International Journal of Electrical Power & Energy Systems*, vol. 54, pp. 187 –197, 2014, ISSN: 0142-0615. DOI: https://doi.org/10.1016/j.ijepes.2013.07.010.

[36] R. Kabiri, D. G. Holmes, B. P. McGrath, and L. G. Meegahapola, "LV Grid Voltage Regulation Using Transformer Electronic Tap Changing, With PV Inverter Reactive Power Injection," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 3, no. 4, pp. 1182–1192, 2015, ISSN: 2168-6777. DOI: 10.1109/JESTPE.2015.2443839.

[37] Y. Wang, P. Zhang, W. Li, W. Xiao, and A. Abdollahi, "Online overvoltage prevention control of photovoltaic generators in microgrids," *IEEE Transactions on Smart Grid*, vol. 3, no. 4, pp. 2071–2078, 2012, ISSN: 19493053. DOI: 10.1109/TSG.2012.2222679.

[38] R. Tonkoski, L. A. Lopes, and T. H. El-Fouly, "Coordinated active power curtailment of grid connected PV inverters for overvoltage prevention," *IEEE Transactions on Sustainable Energy*, vol. 2, no. 2, pp. 139–147, 2011.

[39] M. Maharjan, "Voltage regulation of low voltage distribution networks," M.S. Thesis, South Dakota State University, 2017, pp. 1–81.

[40] P. Jahangiri and D. C. Aliprantis, "Distributed Volt/VAr control by PV inverters," *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 3429–3439, 2013, ISSN: 08858950. DOI: 10.1109/TPWRS.2013.2256375.

[41] J. E. Quiroz, M. J. Reno, O. Lavrova, and R. H. Byrne, "Communication requirements for hierarchical control of volt-var function for steady-state voltage," in *2017 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, 2017, 5 pp. DOI: 10.1109/ISGT.2017.8086007.

[42] F. Olivier, P. Aristidou, D. Ernst, and T. Van Cutsem, "Active Management of Low-Voltage Networks for Mitigating Overvoltages Due to Photovoltaic Units," *IEEE Transactions on Smart Grid*, vol. 7, no. 2, pp. 926–936, 2016, ISSN: 19493053. DOI: 10.1109/TSG.2015.2410171.

[43] K. E. Antoniadou-Plytaria, I. N. Kouveliotis-Lysikatos, P. S. Georgilakis, and N. D. Hatziargyriou, "Distributed and decentralized voltage control of smart distribution networks: Models, methods, and future research," *IEEE Transactions on Smart Grid*, vol. 8, no. 6, pp. 2999–3008, 2017.

[44] "Residential demand response model and impact on voltage profile and losses of an electric distribution network," *Applied Energy*, vol. 96, pp. 84 –91, 2012, Smart Grids, ISSN: 0306-2619. DOI: https://doi.org/10.1016/j.apenergy.2011.12.076. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0306261911008798.

[45] M. Shamshiri, C. K. Gan, and R. Omar, "Assessment of distribution networks performance considering residential photovoltaic systems with demand response applications," *Journal of Renewable and Sustainable Energy*, vol. 9, no. 4, p. 045 502, 2017. DOI: 10.1063/1.4993048. eprint: https://doi.org/10.1063/1.4993048. [Online]. Available: https://doi.org/10.1063/1.4993048.

[46] F. R. S. Sevilla], D. Parra, N. Wyrsch, M. K. Patel, F. Kienzle, and P. Korba, "Techno-economic analysis of battery storage and curtailment in a distribution grid with high pv penetration," *Journal of Energy Storage*, vol. 17, pp. 73 –83, 2018, ISSN: 2352-152X. DOI: https://doi.org/10.1016/j.est.2018.02.001. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2352152X17302591.

[47] L. Xiong, P. Li, Z. Wang, and J. Wang, "Multi-agent based multi objective renewable energy management for diversified community power consumers," *Applied Energy*, vol. 259, p. 114 140, 2020, ISSN: 0306-2619. DOI: https://doi.org/10.1016/j.apenergy.2019.114140. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0306261919318276.

[48] A. Bernstein, L. Reyes-Chamorro, J.-Y. L. Boudec], and M. Paolone, "A composable method for real-time control of active distribution networks with explicit power setpoints. part i: Framework," *Electric Power Systems Research*, vol. 125, pp. 254 –264, 2015, ISSN: 0378-7796. DOI: https://doi.org/10.1016/j.epsr.2015.03.023. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0378779615000905.

[49] B. Zhang, A. Y. S. Lam, A. D. Domínguez-García, and D. Tse, "An optimal and distributed method for voltage regulation in power distribution systems," *IEEE Transactions on Power Systems*, vol. 30, no. 4, pp. 1714–1726, 2015.

[50] S. J. Steffel, P. R. Caroselli, A. M. Dinkel, J. Q. Liu, R. N. Sackey, and N. R. Vadhar, "Integrating solar generation on the electric distribution grid," *IEEE Transactions on Smart Grid*, vol. 3, no. 2, pp. 878–886, 2012.

[51] O. Malík and P. Havel, "Active demand-side management system to facilitate integration of res in low-voltage distribution networks," *IEEE Transactions on Sustainable Energy*, vol. 5, no. 2, pp. 673–681, 2014.

[52] E. Yao, P. Samadi, V. W. S. Wong, and R. Schober, "Residential demand side management under high penetration of rooftop photovoltaic units," *IEEE Transactions on Smart Grid*, vol. 7, no. 3, pp. 1597–1608, 2016.

[53] A. Losi, P. Mancarella, and A. Vicino, *Integration of demand response into the electricity chain: challenges, opportunities, and Smart Grid solutions*. John Wiley & Sons, 2015.

[54] F. B. dos Reis, R. Tonkoski, and T. Hansen, "Synthetic residential load models for smart city energy management simulations," *IET Smart Grid*, vol. 3, 342–354(12), 3 2020. DOI: 10.1049/iet-stg.2019.0296. [Online]. Available: https://digital-library.theiet.org/content/journals/10.1049/iet-stg.2019.0296.

[55] T. M. Hansen, E. K. P. Chong, S. Suryanarayanan, A. A. Maciejewski, and H. J. Siegel, "A Partially Observable Markov Decision Process Approach to Residential Home Energy Management," *IEEE Transactions on Smart Grid*, vol. 9, no. 2, pp. 1271–1281, 2018, ISSN: 1949-3053. DOI: 10.1109/TSG.2016.2582701.

[56] W. Whitt and X. Zhang, "A data-driven model of an emergency department," *Operations Research for Health Care*, vol. 12, pp. 1 –15, 2017.

[57] O. Hafez and K. Bhattacharya, "Queuing analysis based pev load modeling considering battery charging behavior and their impact on distribution system operation," *IEEE Transactions on Smart Grid*, vol. 9, no. 1, pp. 261–273, 2018, ISSN: 1949-3053. DOI: 10.1109/TSG.2016.2550219.

[58] S. G. E. Eick, A. M. William, and W. Whitt, "$M_t/G/\infty$ queues with sinusoidal arrival rates," *Management Science*, vol. 39, no. 2, pp. 241–252, 1993.

[59] R. Lambiotte, L. Tabourier, and J.-C. Delvenne, "Burstiness and spreading on temporal networks," *The European Physical Journal B*, vol. 86, no. 7, p. 320, 2013.

[60] M. Pipattanasomporn, M. Kuzlu, S. Rahman, and Y. Teklu, "Load profiles of selected major household appliances and their demand response opportunities," *IEEE Transactions on Smart Grid*, vol. 5, no. 2, pp. 742–750, 2014, ISSN: 19493053. DOI: 10.1109/TSG.2013.2268664.

[61] B. R. Scalley and D. G. Kasten, "The effects of distribution voltage reduction on power and energy consumption," *IEEE Transactions on Education*, vol. 24, no. 3, pp. 210–216, 1981, ISSN: 0018-9359. DOI: 10.1109/TE.1981.4321493.

[62] M. Diaz-Aguiló and J. Sandraz and R. Macwan and F. de León and D. Czarkowski and C. Comack and D. Wang, "Field-Validated Load Model for the Analysis of CVR in Distribution Secondary Networks: Energy Conservation," *IEEE Transactions on Power Delivery*, vol. 28, no. 4, pp. 2428–2436, 2013, ISSN: 0885-8977. DOI: 10.1109/TPWRD.2013.2271095.

[63] A. Bokhari, A. Alkan, R. Dogan, M. Diaz-Aguiló, F. de León, D. Czarkowski, Z. Zabar, L. Birenbaum, A. Noel, and R. E. Uosef, "Experimental Determination of the ZIP Coefficients for Modern Residential, Commercial, and Industrial Loads," *IEEE Transactions on Power Delivery*, vol. 29, no. 3, pp. 1372–1381, 2014, ISSN: 0885-8977. DOI: 10.1109/TPWRD.2013.2285096.

[64] R. Tonkoski, D. Turcotte, and T. H. M. EL-Fouly, "Impact of high PV penetration on voltage profiles in residential neighborhoods," *IEEE Transactions on Sustainable Energy*, vol. 3, no. 3, pp. 518–527, 2012, ISSN: 1949-3029. DOI: 10.1109/TSTE.2012.2191425.

[65] J. Hu, M. Marinelli, M. Coppo, A. Zecchino, and H. W. Bindner, "Coordinated voltage control of a decoupled three-phase on-load tap changer transformer and photovoltaic inverters for managing unbalanced networks," *Electric Power Systems Research*, vol. 131, pp. 264 –274, 2016.

[66] K. Ma, G. Hu, and C. J. Spanos, "Distributed energy consumption control via real-time pricing feedback in smart grid," *IEEE Transactions on Control Systems Technology*, vol. 22, no. 5, pp. 1907–1914, 2014, ISSN: 10636536. DOI: 10.1109/TCST.2014.2299959.

[67] PJM, *PJM estimated hourly load*. [Online]. Available: http://www.pjm.com/markets-and-operations/energy/real-time/loadhryr.aspx.

[68] "American national standard for electric power systems and equipment," According to American National Standards Institute, Standard C84.1, 2011.

[69] F. Bereta dos Reis, *Synthetic Residential Load Models for Energy Management Benchmarks*, will be published on paper acceptance. [Online]. Available: https://github.com/FBdR05/Synthetic-Residential-Load-Models-for-Energy-Management-Benchmarks.

[70] Fatih Birol, "Digitalization & Energy," International Energy Agency (IEA), Tech. Rep., 2017, pp. 1–188.

[71] N. Good, K. A. Ellis, and P. Mancarella, "Review and classification of barriers and enablers of demand response in the smart grid," *Renewable and Sustainable Energy Reviews*, vol. 72, pp. 57 –72, 2017, ISSN: 1364-0321. DOI: https://doi.org/10.1016/j.rser.2017.01.043. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1364032117300436.

[72] G Strbac, M Aunedi, D Pudjianto, P Djapic, S Gammons, and R Druce, "Understanding the balancing challenge," *DECC, London*, 2012.

[73] A. Pratt, D. Krishnamurthy, M. Ruth, H. Wu, M. Lunacek, and P. Vaynshenk, "Transactive home energy management systems: The impact of their proliferation on the electric grid," *IEEE Electrification Magazine*, vol. 4, no. 4, pp. 8–14, 2016.

[74] S. Chen and C. Liu, "From demand response to transactive energy: State of the art," *Journal of Modern Power Systems and Clean Energy*, vol. 5, no. 1, pp. 10–19, 2017.

[75] B. P. Bhattarai, J. Alam, J. Hansen, K. Schneider, N. Radhakrishnan, A. Somani, and W. Du, "Enhancing distribution system resiliency through a novel transactive energy systems framework," in *2019 IEEE Power Energy Society General Meeting (PESGM)*, 2019.

[76] L. Park, Y. Jang, S. Cho, and J. Kim, "Residential demand response for renewable energy resources in smart grid systems," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 6, pp. 3165–3173, 2017.

[77] S. Nan, M. Zhou, and G. Li, "Optimal residential community demand response scheduling in smart grid," *Applied Energy*, vol. 210, pp. 1280 –1289, 2018, ISSN: 0306-2619. DOI: https://doi.org/10.1016/j.apenergy.2017.06.066. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S030626191730819X.

[78] D. Li, W. Chiu, H. Sun, and H. V. Poor, "Multiobjective optimization for demand side management program in smart grid," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 4, pp. 1482–1490, 2018.

[79] P. Paudyal, P. Munankarmi, Z. Ni, and T. M. Hansen, "A hierarchical control framework with a novel bidding scheme for residential community energy optimization," *IEEE Transactions on Smart Grid*, vol. 11, no. 1, pp. 710–719, 2020.

[80] *IEEE PES Distribution Systems Analysis Subcommittee radial test feeders*, 2015. [Online]. Available: http://sites.ieee.org/pes-testfeeders/resources.

[81] V. Krishnan, B. Bugbee, T. Elgindy, C. Mateo, P. Duenas, F. Postigo, J. Lacroix, T. G. S. Roman, and B. Palmintier, "Validation of Synthetic U.S. Electric Power Distribution System Data Sets," *IEEE Transactions on Smart Grid*, pp. 1–1, 2020.

[82] C. Mateo, F. Postigo, F. de Cuadra, T. Gómez, T. Elgindy, P. Dueñas, B. Hodge, V. Krishnan, and B. Palmintier, "Building Large-Scale U.S. Synthetic Electric Distribution System Models," *IEEE Transactions on Smart Grid*, 2020, pre-print.

[83] C. Coffrin, D. Gordon, and P. Scott, "NESTA, the NICTA energy system test case archive," *arXiv preprint arXiv:1411.0359*, 2016.

[84] A. B. Birchfield, E. Schweitzer, M. H. Athari, T. Xu, T. J. Overbye, A. Scaglione, and Z. Wang, "A metric-based validation process to assess the realism of synthetic power grids," *Energies*, vol. 10, no. 8, p. 1233, 2017.

[85] A. B. Birchfield, T. Xu, K. M. Gegner, K. S. Shetye, and T. J. Overbye, "Grid structural characteristics as validation criteria for synthetic networks," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 3258–3265, 2017.

[86] R. Espejo, S. Lumbreras, and A. Ramos, "Analysis of transmission-power-grid topology and scalability, the european case study," *Physica A: Statistical Mechanics and its Applications*, vol. 509, pp. 383 –395, 2018, ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2018.06.019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S037843711830743X.

[87] R. Espejo, S. Lumbreras, and A. Ramos, "A complex-network approach to the generation of synthetic power transmission networks," *IEEE Systems Journal*, vol. 13, no. 3, pp. 3050–3058, 2019.

[88] G. Prettico, F. Gangale, A. Mengolini, A. Lucas, and G. Fulli, "Distribution system operators observatory: From european electricity distribution systems to representative distribution networks," *Luxembourg: Publications Office of the European Union*, p. 50, 2016. DOI: 10.2790/701791.

[89] C. Mateo, G. Prettico, T. Gómez, R. Cossent, F. Gangale, P. Frías, and G. Fulli, "European representative electricity distribution networks," *International Journal of Electrical Power & Energy Systems*, vol. 99, pp. 273 –280, 2018, ISSN: 0142-0615. DOI: https://doi.org/10.1016/j.ijepes.2018.01.027. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S014206151731801X.

[90] M. Grzanic, M. G. Flammini, and G. Prettico, "Distribution network model platform: A first case study," *Energies*, vol. 12, no. 21, p. 4079, 2019.

[91] E. Schweitzer, A. Scaglione, A. Monti, and G. A. Pagani, "Automated generation algorithm for synthetic medium voltage radial distribution systems," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 7, no. 2, pp. 271–284, 2017.

[92] F. Bu, Y. Yuan, Z. Wang, K. Dehghanpour, and A. Kimber, "A time-series distribution test system based on real utility data," *arXiv preprint arXiv:1906.04078*, 2019.

[93] Z. Wang, *Iowa Distribution Test Systems*, 2019. [Online]. Available: http://wzy.ece.iastate.edu/Testsystem.html.

[94] U.S. Energy Information Administration, 2015 Residential Energy Consumption Survey, *Use of energy explained Energy use in homes*, 2015. [Online]. Available: https://www.eia.gov/energyexplained/use-of-energy/electricity-use-in-homes.php.

[95] D. P. Chassin, J. C. Fuller, and N. Djilali, "GridLAB-D: An Agent-Based Simulation Framework for Smart Grids," *Journal of Applied Mathematics*, vol. 2014, H. Jia, Ed., p. 492 320, 2014, ISSN: 1110-757X. DOI: 10.1155/2014/492320. [Online]. Available: https://doi.org/10.1155/2014/492320.

[96] *NREL/ditto: DiTTo is a Distribution Transformation Tool that aims at providing an open source framework to convert various distribution systems modeling formats.* 2018. [Online]. Available: https://github.com/NREL/ditto.

[97] *NREL/glm: A fast GridLAB-D to JSON (and back) parser.* 2019. [Online]. Available: https://github.com/NREL/glm.

[98] W. H. Kersting, *Distribution system modeling and analysis*. CRC press, 2012.

[99] *GridLAB-D model with the synthetic load data is made publicly available*, Will be published on paper acceptance. The authors are selecting the venue.

[100] *GridLAB-D model with the synthetic load data is made publicly available*, 2020. [Online]. Available: db.bettergrids.org.

[101] S.-E. Razavi, E. Rahimi, M. S. Javadi, A. E. Nezhad, M. Lotfi, M. Shafie-khah, and J. P. Catalão, "Impact of distributed generation on protection and voltage regulation of distribution systems: A review," *Renewable and Sustainable Energy Reviews*, vol. 105, pp. 157 –167, 2019, ISSN: 1364-0321. DOI: https://doi.org/10.1016/j.rser.2019.01.050. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1364032119300668.

[102] A. Zakariazadeh, O. Homaee, S. Jadid, and P. Siano, "A new approach for real time voltage control using demand response in an automated distribution system," *Applied Energy*, vol. 117, pp. 157 –166, 2014, ISSN: 0306-2619. DOI: https://doi.org/10.1016/j.apenergy.2013.12.004. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0306261913009884.

[103] P. H. Divshali, B. J. Choi, and H. Liang, "Multi-agent transactive energy management system considering high levels of renewable energy source and electric vehicles," English, *IET Generation, Transmission & Distribution*, vol. 11, 3713–3721(8), 15 2017, ISSN: 1751-8687. [Online]. Available: https://digital-library.theiet.org/content/journals/10.1049/iet-gtd.2016.1916.

[104] K. Duwadi, F. B. Dos Reis, R. Mahat, R. Fourney, R. Tonkoski, T. M. Hansen, and B. P. Bhattarai, "Numerical oscillation prevention for pv inverter controllers in quasi-steady-state simulators," in *2019 IEEE Power Energy Society General Meeting (PESGM)*, 2019, pp. 1–5.

[105] A. Bajracharya, "Intra-Day Solar Irradiance Forecasting for Remote Microgrids Using Hidden Markov Model," Master's thesis, South Dakota State University, South Dakota, 2019.

[106] A. Shakya, S. Michael, C. Saunders, D. Armstrong, P. Pandey, S. Chalise, and R. Tonkoski, "Solar irradiance forecasting in remote microgrids using markov switching model," *IEEE Transactions on Sustainable Energy*, vol. 8, no. 3, pp. 895–905, 2017.

[107] A. Shakya, "Implementation of solar irradiance forecasting using markov switching model and energy management system," Master's thesis, South Dakota State University, South Dakota, 2016.

[108] M. B. Kursa, W. R. Rudnicki, *et al.*, "Feature selection with the boruta package," *J Stat Softw*, vol. 36, no. 11, pp. 1–13, 2010.

[109] ComEd, *Five minute prices — comed's hourly pricing program*. [Online]. Available: https://hourlypricing.comed.com/live-prices/five-minute-prices/.

[110] ComEd, *Answers — comed's hourly pricing program*. [Online]. Available: https://hourlypricing.comed.com/faqs/?question=what-is-capacity-charge.

[111] ComEd, *View a Sample Residential Bill*, 2020. [Online]. Available: https://www.comed.com/MyAccount/MyBillUsage/Pages/SampleResidentialBill2.aspx.

[112] R. Roche, S. Natarajan, A. Bhattacharyya, and S. Suryanarayanan, "A framework for co-simulation of AI tools with power systems analysis software," in *23rd International Workshop on Database and Expert Systems Applications (DEXA)*, Sept. 2012, pp. 350–354.

[113] E. K. P. Chong, C. M. Kreucher, and A. O. Hero III, "Partially observable Markov decision process approximations for adaptive sensing," *Discrete Event Dynamic Systems,* special issue on Optimization of Discrete Event Dynamic Systems, vol. 19, no. 3, pp. 377–422, Sep. 2009.

[114] R. E. Bellman, *Dynamic Programming*. Princeton University Press, Princeton, NJ, 1957.

[115] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEE Proceedings F (Radar and Signal Processing)*, vol. 140, no. 2, pp. 107–113, Apr. 1993.

[116] T. M. Hansen, B. Palmintier, S. Suryanarayanan, A. A. Maciejewski, and H. J. Siegel, "Bus.py: A GridLAB-D communication interface for Smart distribution Grid simulations," in *2015 IEEE Power Energy Society General Meeting*, 2015, 5 pp. DOI: 10.1109/PESGM.2015.7286003.

[117] SOLAR ELECTRIC SUPPLY, *8.4 KW SolarWorld Sunmodule Plus SW 280 Mono Solar Panel System*, 2017. [Online]. Available: https://www.solarelectricsupply.com/8-4-kw-solaria-powerxt-350r-pd-all-black-solar-panel-system.

[118] M. Beaudin and H. Zareipour, "Home energy management systems: A review of modelling and complexity," *Renewable and Sustainable Energy Reviews*, vol. 45, pp. 318–335, 2015, ISSN: 13640321. DOI: 10.1016/j.rser.2015.01.046. [Online]. Available: http://dx.doi.org/10.1016/j.rser.2015.01.046.

[119] A. Faruqui and J. Palmer, "The discovery of price responsiveness–a survey of experiments involving dynamic pricing of electricity," *Available at SSRN 2020587*, 2012.

[120] D. Violette, B. Provencher, M. Klos, R. Freeman, P. Steele-Mosey, D. Clark, and D. Klos, "Power Smart Pricing 2009 Annual Report Submitted to Ameren Illinois Utilities," Summit Blue Consulting, Boulder, CO, Tech. Rep., 2010, p. 119.

[121] S. Gyamfi, S. Krumdieck, and T. Urmee, "Residential peak electricity demand response—highlights of some behavioural issues," *Renewable and Sustainable Energy Reviews*, vol. 25, pp. 71 –77, 2013, ISSN: 1364-0321. DOI: https://doi.org/10.1016/j.rser.2013.04.006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1364032113002578.

[122] R. Kadavil, S. Lurbé, S. Suryanarayanan, P. A. Aloise-Young, S. Isley, and D. Christensen, "An application of the analytic hierarchy process for prioritizing user preferences in the design of a home energy management system," *Sustainable Energy, Grids and Networks*, vol. 16, pp. 196 –206, 2018, ISSN: 2352-4677. DOI: https://doi.org/10.1016/j.segan.2018.07.009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2352467717301893.

[123] D. Zhang, N. Shah, and L. G. Papageorgiou, "Efficient energy consumption and operation management in a smart building with microgrid," *Energy Conversion and Management*, vol. 74, pp. 209 –222, 2013, ISSN: 0196-8904. DOI: https://doi.org/10.1016/j.enconman.2013.04.038. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0196890413002355.

[124] S. Mitchell, C.-M. Duquesne, and F. Peschiera, *Coin-or/pulp: A python linear programming api*, 2010. [Online]. Available: https://github.com/coin-or/pulp.

[125] J. Farrell, "Democratizing the electricity system: A vision for the 21st century grid. new rules project," *Institute for Local Self Reliance. Jun. http://www. newrules. org/energy/publications/democratizingelectricity-system-vision-21st-century-grid*, 2011.

[126] B. van Veelen and C. Haggett, "Uncommon ground: The role of different place attachments in explaining community renewable energy projects," *Sociologia Ruralis*, vol. 57, pp. 533–554, 2017.

[127] B. Alexander, "Smart meters, real time pricing, and demand response programs: Implications for low income electric customers," *Oak Ridge National Laboratory*, pp. 1–73, 2007.

[128] X. Xu and C. fei Chen, "Energy efficiency and energy justice for u.s. low-income households: An analysis of multifaceted challenges and potential," *Energy Policy*, vol. 128, pp. 763 –774, 2019, ISSN: 0301-4215. DOI: https://doi.org/10.1016/j.enpol.2019.01.020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0301421519300205.

[129] C. Miller and I. Savage, "Does the demand response to transit fare increases vary by income?" *Transport Policy*, vol. 55, pp. 79 –86, 2017, ISSN: 0967-070X. DOI: https://doi.org/10.1016/j.tranpol.2017.01.006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0967070X16302852.

[130] D. B. Crawley, L. K. Lawrie, F. C. Winkelmann, W. Buhl, Y. Huang, C. O. Pedersen, R. K. Strand, R. J. Liesen, D. E. Fisher, M. J. Witte, and J. Glazer, "Energyplus: Creating a new-generation building energy simulation program," *Energy and Buildings*, vol. 33, no. 4, pp. 319 –331, 2001, Special Issue: BUILDING SIMULATION'99, ISSN: 0378-7788. DOI: https://doi.org/10.1016/S0378-7788(00)00114-6. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0378778800001146.

[131] Commonwealth Edison Company, *Company Information*. [Online]. Available: https://www.comed.com/AboutUs/Pages/CompanyInformation.aspx (visited on 01/02/2018).

[132] S. Seabold and J. Perktold, "statsmodels: Econometric and statistical modeling with Python," in *9th Python in Science Conference*, 2010.