A Profit-Neutral Double-price-signal Retail Electricity Market Solution for Incentivizing Price-responsive DERs Considering Network Constraints

Mengmeng Cai

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

Doctor of Philosophy
In
Electrical Engineering

Saifur Rahman

Manisa Pipattanasomporn

Robert Broadwater

Guoqiang Yu

Chandan K.Reddy

May 11, 2020
Arlington, Virginia

Keywords: Retail Electricity Market, Load Forecasting, Battery Arbitrage, Bi-level Optimization, Deep Learning

A Profit-Neutral Double-price-signal Retail Electricity Market Solution for Incentivizing Price-responsive DERs Considering Network Constraints

Mengmeng Cai

Abstract

Emerging technologies, including distributed energy resources (DERs), internet-of-things and advanced distribution management systems, are revolutionizing the power industry. They provide benefits like higher operation flexibility and lower bulk grid dependency, and are moving the modern power grid towards a decentralized, interconnected and intelligent direction. Consequently, the emphasis of the system operation management has been shifted from the supply-side to the demand-side. It calls for a reconsideration of the business model for future retail market operators. To address this need, this dissertation proposes an innovative retail market solution tailored to market environments penetrated with price-responsive DERs. The work is presented from aspects of theoretical study, test-bed platform development, and experimental analysis, within which two topics relevant to the retail market operation are investigated in depth.

The first topic covers the modeling of key retail market participants. With regard to price-insensitive participants, fixed loads are treated as the representative. Deep learning-based day-ahead load forecasting models are developed in this study, utilizing both recurrent and convolutional neural networks, to predict the part of demands that keep fixed regardless of the market price. With regard to price-sensitive participants, battery storages are selected as the representative. An optimization-based battery arbitrage model is developed in this study to represent their price-responsive behaviors in response to a dynamic price. The second topic further investigates how the retail market model and pricing strategy should be designed to incentivize these market participants. Different from existing works, this study innovatively proposes a profit-neutral double-price-signal retail market model. Such a design differentiates elastic prosumers, who actively offer flexibilities to the system operation, from normal inelastic consumers/generators, based on their sensitivities to the market price. Two price signals, namely retail grid service price and retail energy price, are then introduced to separately quantify values of the flexibility, provided by elastic participants, and the electricity commodity, sold/bought to/from inelastic participants. Within the proposed retail market, a non-profit retail market operator (RMO) manages and settles the market through determining the price signals and supplementary subsidy to minimize the overall system

cost. In response to the announced retail grid service price, elastic prosumers adjust their day-ahead operating schedules to maximize their payoffs. Given the interdependency between decisions made by the RMO and elastic participants, a retail pricing scheme, formulated based on a bi-level optimization framework, is proposed. Additional efforts are made on merging and linearizing the original non-convex bi-level problem into a single-level mixed-integer linear programming problem to ensure the computational efficiency of the retail pricing tool.

Case studies are conducted on a modified IEEE 34-bus test-bed system, simulating both physical operations of the power grid and financial interactions inside the retail market. Experimental results demonstrate promising properties of the proposed retail market solution: First of all, it is able to provide cost-saving benefits to inelastic customers and create revenues for elastic customers at the same time, justifying the rationalities of these participants to join the market. Second of all, the addition of the grid service subsidy not only strengthens the profitability of the elastic customer, but also ensures that the benefit enjoyed per customer will not be compromised by the competition brought up by a growing number of participants. Furthermore, it is able to properly capture impacts from line losses and voltage constraints on the system efficiency and stability, so as to derive practical pricing solutions that respect the system operating rules. Last but not least, it encourages the technology improvement of elastic assets as elastic assets in better conditions are more profitable and could better save the electricity bills for inelastic customers. Above all, the superiority of the proposed retail market solution is proven. It can serve as a promising start for the retail electricity market reconstruction.

A Profit-Neutral Double-price-signal Retail Electricity Market Solution for Incentivizing Price-responsive DERs Considering Network Constraints

Mengmeng Cai

General Audience Abstract

The electricity market plays a critical role in ensuring the economic and secure operation of the power system. The progress made by distributed energy resources (DERs) has reshaped the modern power industry bringing a larger proportion of price-responsive behaviors to the demand-side. It challenges the traditional wholesale-only electricity market and calls for an addition of retail markets to better utilize distributed and elastic assets. Therefore, this dissertation targets at offering a reliable and computational affordable retail market solution to bridge this knowledge gap.

Different from existing works, this study assumes that the retail market is managed by a profit-neutral retail market operator (RMO), who oversees and facilitates the system operation for maximizing the system efficiency rather than making profits. Market participants are categorized into two groups: inelastic participants and elastic participants, based on their sensitivity to the market price. The motivation behind this design is that instead of treating elastic participants as normal customers, it is more reasonable to treat them as grid service providers who offer operational flexibilities that benefit the system efficiency. Correspondingly, a double-signal pricing scheme is proposed, such that the flexibility, provided by elastic participants, and the electricity commodity, generated/consumed by inelastic participants, are separately valued by two distinct prices, namely retail grid service price and retail energy price. A grid service subsidy is also introduced in the pricing system to provide supplementary incentives to elastic customers. These two price signals in addition to the subsidy are determined by the RMO via solving a bi-level optimization problem given the interdependency between the prices and reaction of elastic participants.

Experimental results indicate that the proposed retail market model and pricing scheme are beneficial for both types of market participants, practical for the network-constrained real-world implementation, and supportive for the technology improvement of elastic assets.

Acknowledgements

I would like to take this opportunity to express my gratitude to all those persons who have offered me invaluable support during my Ph.D. and contributed to the finish of this dissertation.

To begin with, I would like to thank my advisor, Dr. Saifur Rahman, for providing me an excellent platform and opportunities working on projects that broadened my horizon and consolidated my skills. His selfless share of life experience taught me how to be a qualified, ingenuous and rigorous researcher. Under his encouragement and guidance, I have fully enjoyed the time of exploring my research. I also could never reach this far without the support and help from Dr. Manisa Pipattanasomporn. I appreciate the patient and time she has spent on me for ensuring the quality and influence of my work. In addition to that, I would like to thank Dr. Robert Broadwater, Dr. Guoqiang Yu and Dr. Chandan Reddy for serving in my Ph.D. committees and continuously providing me valuable feedbacks and suggestions on my dissertation.

I also could not thank enough my mentors, Dr. Rui Yang and Dr. Bishnu Bhattarai, and managers, Dr. Yingchen Zhang and Mr. Robert Pratt, who have guided me during my internships at NREL and PNNL. It has been a great pleasure working with them on two exciting projects. I enjoyed every discussion and brainstorms I had with them as well as other teammates: Fernando, Sarmad, Ahmad, Jacob and Jeffrey. They are all smart and inspiring persons for me to learn from.

Moreover, I would like thank Rajendra for being my first close friend at ARI. The creative ideas come out from his mind have been a great source of happiness to me. And many thanks to all my friends at ARI, Dr. Murat, Yonael, Ganaa, Aditya, Ashraful, Nazmul, Tao, Zejia, Desong, Jason, Abdullah, Musaed, Massoud, Avijit, Shibani, Shuyang, Dr. Jiang, Hongxun, Jia, Grace, Dr. Usama, Sneha and Imran for always being there to help me and listen to my happiness or sadness.

Last, but definitely not the least, I am greatly indebted to my family. I would like to thank my parents, Huaining and Zhi, for teaching me the attitude of always being curious and searching for truths. Not to mention their selfless devotion on providing me the warmest harbor and the best education. And my biggest thank to my husband, Xiangyu, for being my strongest backbone. He is my compass when I am lost in my journey. Thanks to his unconditional support, care and tolerance, I feel like I can overcome all the difficulties. I am also sincerely grateful for his family for supporting both of us pursuing our Ph.D. degree.

TABLE OF CONTENTS

List of Figures ix
List of Tables xi
1. Introduction 1
1.1 Background 1
1.2 Motivation 3
1.3 Objectives 3
1.4 Tasks 5
2. Literature research 7
2.1 Types of DERs 7
2.2 Applications of DERs 10
2.3 Distribution system with high DERs penetration 12
2.4 Load forecasting techniques 13
2.5 Wholesale electricity market 14
2.6 Attempted retail market deregulation 16
2.7 Knowledge gaps 20
3. Retail market environment setup and test-bed system development 21
3.1 Retail market model 21
3.1.1 Market hierarchy 21
3.1.2 Assumptions made in the market 22
3.1.3 Categories of market participants and the double-signal pricing scheme 22
3.1.4 Decision makers in the market 23
3.1.5 Market settlement 24
3.2 Test-bed system development 25
3.2.1 Framework of the test-bed system 25
3.2.2 Modeling of the three-phase unbalanced distribution system 26
3.2.3 Development of three functional agents 27
4. Deep learning-based load forecasting models 29
4.1 Problem formulation 29
4.2 Seasonal ARIMAX model 29
4.3 Gated RNN models 31
4.4 Gated CNN models 32
4.5 Experimental setting 33

4.5.1 Datasets 34
4.5.2 Weather relevant feature selection 35
4.5.3 Data preprocessing 35
4.5.4 Model validation 36
4.5.5 Implementation 36
4.6 Performance analyses 37
4.6.1 Accuracy of one-hour prediction models 37
4.6.2 Accuracy of 24-hour prediction models 38
4.6.3 Performance comparison of day-ahead load forecasts 39
4.6.4 Computational efficiency 43
4.6.5 Robustness analysis 44
4.6.6 Conclusion 45
5. Day-ahead battery arbitrage model 46
5.1 Modeling of single battery storage arbitrage behavior 46
5.2 Sensitivity analyses of the price-responsive behavior 49
6. Bi-level retail pricing mechanism 53
6.1 Problem statement 53
6.2 Linearization of the upper-level problem 55
6.3 Transformation of the bi-level optimization problem 57
6.4 Linearization of the merged single-level problem 59
7. Retail market simulation 62
7.1 Experimental design 62
7.2 Parameter configuration 62
7.3 Rationality of customers to participate in the retail market 64
7.4 Impacts of the retail grid service subsidy 67
7.5 Impacts of network constraints 68
7.6 Sensitivity analysis 71
7.6.1 Sensitivity of battery parameters 71
7.6.2 Sensitivity of the market parameter 75
7.7 Proposed vs. alterative pricing schemes 77
8. Summary, conclusions and future work 80
8.1 Summary 80
8.2 Conclusions 81

8.2.1 Deep learning models for day ahead load forecasting 81
8.2.2 Optimization-based battery arbitrage model for modeling the price-responsive behavior of batteries 81
8.2.3 Profit-neutral double-signal retail market model and pricing scheme for optimal system operation 81
8.3 Future work 82
References 84

LIST OF FIGURES

Figure 1-1. Illustration of the emerging smart grid framework 2

Figure 2-1. Long-term prediction of renewable energy generation [9] 8

Figure 2-2. Smart building with active EMS [26] 9

Figure 2-3. Price elasticity curve 17

Figure 2-4. Hierarchical structure of the retail market 17

Figure 3-1. Structure of the retail market model 21

Figure 3-2. Illustration of market settlements under three different market settings 24

Figure 3-3. Framework of the test-bed platform 25

Figure 3-4. Structure of the OpenDSS model and its interactions with the OpenDSS Communicator 27

Figure 3-5. Python classes of three functional agents 28

Figure 4-1. Gated RNN model for day-ahead load forecasts 32

Figure 4-2. Gated CNN model for day-ahead load forecasts 33

Figure 4-3. Load profiles of three buildings 34

Figure 4-4. Comparison of the RMSE boxplots of 24-hour DL-based models across all scenarios 39

Figure 4-5. Comparing day-ahead load forecasting results of different models (SARIMAX, GRNN1, GCNN1, GRNN24 and GCNN24) 41

Figure 4-6. Hour-of-day indexed error analysis for SARIMAX, GCNN1 and GCNN24 42

Figure 4-7. Learning curves comparison between GRNN and GCNN models 44

Figure 5-1. Optimized battery scheduling under dynamic price signals 48

Figure 5-2. Projected Lithium-ion battery pack price fall 48

Figure 5-3. Comparison among different $DkWh$ values 49

Figure 5-4. Comparison among different C values 50

Figure 5-5. Comparison among different $Rmax$ values 50

Figure 5-6. Battery scheduling under the price signal C 51

Figure 5-7. Battery scheduling under the price signal D 52

Figure 7-1. Optimized retail market solutions under different subsidy rates (No Network) 66

Figure 7-2. Financial analyses against the number of elastic customers before and after adding the grid service subsidy 68

Figure 7-3. Optimized retail market solutions under different levels of network constraints 69

Figure 7-4. Sensitivity analysis of $DkWh$ 72

Figure 7-5. Retail market solutions under different $DkWh$ 73

Figure 7-6. Sensitivity analysis of C 74

Figure 7-7. Sensitivity analysis of $Rmax$ 74

Figure 7-8. Sensitivity analysis of $C\beta$ 76

Figure 7-9. Four types of retail market and corresponding market solutions 78

LIST OF TABLES

Table 3-1. Types of market participants 22

Table 3-2. Decision makers in the retail market 23

Table 4-1. Time-series characteristics of load profiles 30

Table 4-2. Pearson correlation coefficients between different weather relevant variables and building load 35

Table 4-3. Load forecasting model configuration 36

Table 4-4. MAPE (%) of one-hour prediction models under different testing cases 37

Table 4-5. CV (%) of one-hour prediction models under different testing cases 37

Table 4-6. Comparison of the averaged RMSE (kW) for different testing datasets 41

Table 4-7. Computation time for each load forecasting model (in seconds) 43

Table 4-8. Averaged RMSE (kW) for five testing days under different noise levels 45

Table 5-1. Definitions of parameters applied in the battery arbitrage problem 47

Table 5-2. Parameter configurations 49

Table 5-3. Optimized financial outcomes of the battery under different parameter settings 51

Table 5-4. Optimized financial outcomes of the battery under different price signals 52

Table 7-1. Three levels of network constraints 63

Table 7-2. Common parameters shared by 20 batteries 63

Table 7-3. Unique parameters of different batteries 64

Table 7-4. Financial analyses for inelastic customers under different subsidy rates 67

Table 7-5. Financial analyses for elastic customers under different subsidy rates 67

Table 7-6. Financial analyses under three different network constraint conditions 71

Table 7-7. Financial analyses under three different values of $C\beta$ 75

Table 7-8. Financial analyses of four retail pricing mechanisms 79

1. INTRODUCTION

1.1 Background

Providing electricity with reliability, efficiency and sustainability is the unremitting mission of the electric power industry. Due to the limited availability of electricity storage, there has to be a perfect match between electricity demand and supply in real-time. However, it is greatly challenged by the complex grid topology and uncertain system behaviors. A poorly managed power grid may easily encounter instability issues ranging from voltage violation, frequency deviation to cascading failures. To avoid these from happening, advanced techniques such as voltage control, frequency regulation and demand response have been implemented ensuring harmonic interactions among system components. Beyond stability, economy is another critical concern of the system operation. On the one hand, dispatches of various energy resources have to be optimized to lower the system cost. On the other hand, appropriate pricing mechanisms have to be designed to reward most cost-efficient system behaviors. Meanwhile, under the pressure of environment degradation, efforts putting on clean energy technology to continuously cut the carbon intensity become commonplace. According to the Sustainable Energy in America Factbook published in 2019, renewables, including hydro, contributed nearly 18% of U.S. electricity generation in 2018, up from 11% in 2009. Total renewable installation capacity doubled to hit 261GW, up from 131GW in 2009 [1]. Due to the continuous support from federal and state governments and the extensive cooperation within the industry, these numbers are expected to climb even higher.

During the past three decades, several developments and emerging technologies have reshaped the modern power industry. The enactment of the Public Utilities Regulatory Policies Act (PURPA) back in 1978 launched the reconstruction process of the modern electricity industry in the United States. PURPA and subsequent acts accelerated the transition from a vertically integrated utility into unbundled and deregulated power system where generation, transmission and distribution are separately managed by different entities. A successful electricity market deregulation played a key role in such a transition [2]. It broke up the monopoly of traditional vertical-structured utilities and brought competition to the market. Under such an open market environment, electricity is exchanged in a way that leads to the highest social welfare. Generating companies compete with each other to improve their price advantages via technological or manufactural innovation; Consumers are guided to actively participate in the demand side management to further improve the system efficiency.

Along with the electricity market deregulation is the progress made by distributed energy resources (DERs) and the information & communication (I&C) technology. DERs refer to generating units located close to end-users. Compare to the traditional centralized generation, it brings benefits like lower marginal cost, lower transportation cost and lower gas emission. The emerge of the DER technology decentralizes the power system and cuts down the dependency of Microgrids on the bulk grid. Meanwhile, there is a widespread implementation of the I&C technology, enhancing awareness and automation of the system operation. At the system-level, advanced metering

infrastructure (AMI) is deployed to ensure quality of the grid service and provide the foundation of autonomous system control. At the end-users-level, building automation technologies, including home energy management systems (HEMS) and building energy management systems (BEMS)[3], [4], connect the grid and end-users through sensors and controllers. It turns buildings from pure consumers into prosumers who actively participate in Demand Response (DR) programs to facilitate demand-side management.

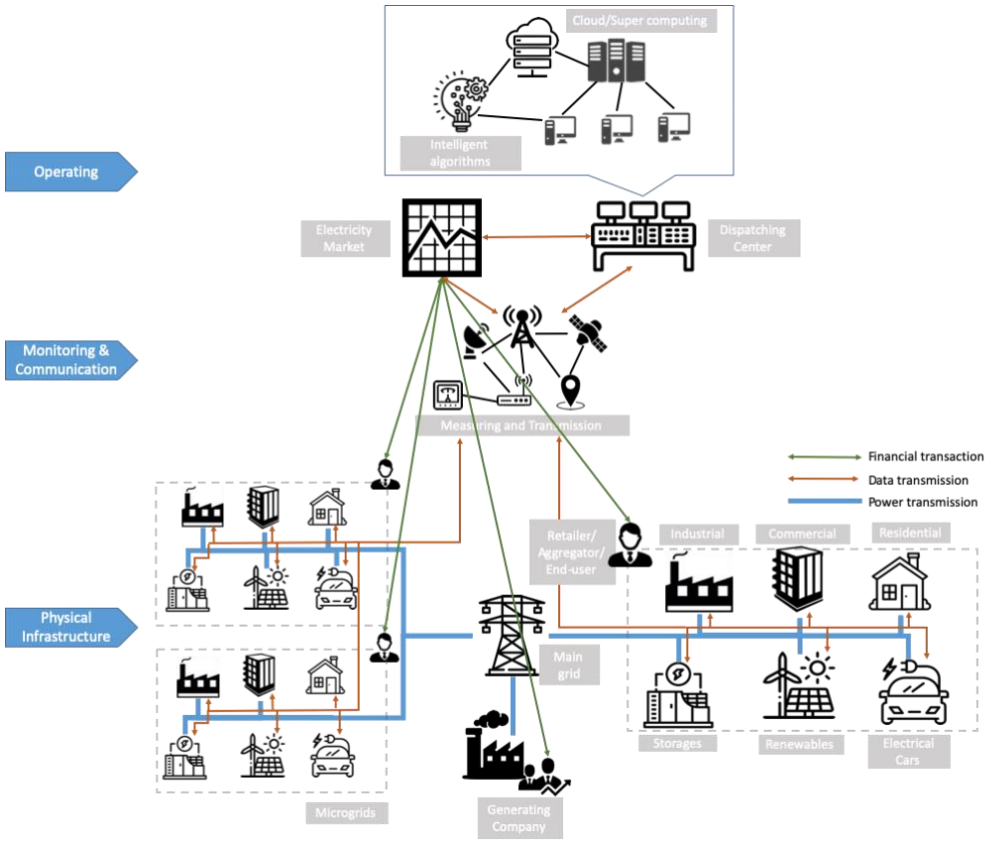


Figure 1-1. Illustration of the emerging smart grid framework

Figure 1-1 illustrates what the future smart grid looks like and how electricity, data and financial messages are transmitted across key system components. At the physical infrastructure layer, the regional power grid can be further broken down into multiple microgrids (MGs), whose dependency on the main grid are weaken, as their demands can be partially served by local resources, including renewables, storages and DR products. At the monitoring & communication layer, an extensive installation of measuring devices and communication network enhances both inter- and intra-connections of MGs. At the operating layer, electricity market and dispatching center work together to ensure the reliable and economic operation of the system. Advanced algorithms [5], [6] as well as computing power are equipped at this layer for handling sophisticated and uncertain application scenarios. To cooperate with the decentralized structure at the infrastructure layer, future electricity market is expected to evolve into a multi-layer hierarchy with market participants at different scales, e.g. individual end-users, aggregators or microgrid operator, participate in markets at different scales, e.g. local market, regional market or wholesale market.

To sum up, under the emerging power industry evolution, the whole power system is moving towards a decentralized, interconnected, intelligent and proactive direction.

1.2 Motivation

The ongoing decentralization and digital transformation marks a new era of the power system modernization. However, before these game-changing technologies, e.g. DERs and advanced distribution management system (ADMS), get fully involved in the picture of future smart grid, some key challenges need to be addressed first, especially at the distribution level:

First, the proactive participation of DERs in the system operation brings a higher proportion of uncertainty. It thus emphasizes the forecasting capability at lower-scale, e.g. building-/appliance-level. On the one hand, end-users need to better understand the energy behavior of their buildings or devices; on the other hand, the system operator requires a better visibility of behind-the-meter resources and the ability to properly value them. However, conventional tools applied for system-level load forecasting are greatly challenged by the nonlinear interdependency and high uncertainty of lower-level load profiles. It calls for an exploration on advanced mathematic tools equipped with more powerful forecasting capacity.

Second, as the distribution system becomes more active rather than passive, a reconsideration of the electricity market design, especially at retail-level, becomes a must. It is necessary to introduce competition also at the distribution level so as to encourage the engagement of DERs. However, compared to the mature implementation of wholesale markets, design of the retail electricity market is more complicated due to a higher penetration of price-responsive behaviors. How to properly incentivize market participants while avoiding the price volatility is still an open topic to be discussed.

Third, present distribution infrastructure, mainly designed for unidirectional power flow and well-controlled resources, is challenged by a large-scale integration of DERs in terms of intermittent generation, reverse power flow and voltage violation. How to take advantage of DERs in promoting the system efficiency while respecting the physical operating rules is of significance to be investigated, since uncoordinated DERs' operations may actually jeopardize rather than benefit the system.

Luckily, in past decades, remarkable achievements have been made by intelligent algorithms, including optimization, model predictive control, machine learning, deep learning, stochastic programming, game theory and so on, in solving various engineering challenges. Their applications on managing the power system have also been demonstrated in the literature. It provides the cornerstone of this dissertation for tackling emerging challenges faced by the high DER-penetrated power system.

1.3 Objectives

Based on the previous discussion, this study aims at providing a reliable and computational affordable retail market solution for addressing challenges raised by the increasing penetration of price-responsive DERs. The overall objective is broken down into five sub-objectives:

Objective 1: To develop deep-learning based day-ahead load forecasting models for predicting fixed electricity demands in the retail market.

Accurate day-ahead demand prediction ensures the optimal system resources scheduling. As mentioned previously, the load forecasting at lower-scale becomes more important nowadays due to the growing penetration of DERs and the active participant of end-users in the system operation. The deep learning (DL) technique has emerged recently as a promising alterative of conventional mathematical tools for handling the time-series analysis, especially for series with complex interdependency and high uncertainty. Despite the powerfulness of DL, being able to grasp the best DL model for certain application scenario is not easy, as there is a large variety of network structures. Therefore, the first sub-objective in the study is to design DL-based models tailored for day-ahead building-level load forecasting utilizing commonly applied DL techniques, and conduct a comprehensive comparison among proposed DL models versus its conventional counterpart.

Objective 2: To model the price-responsive behavior of the battery storage and analyze how market price impacts its decision making.

For better designing the price incentive in the retail market, it is critical to understand how price-sensitive market participants respond to different market prices. Whereas the behavior of price-insensitive customers can be captured by typical forecasting models, the behavior of price-sensitive customers is usually modeled by formulating an optimization problem. Therefore, the second sub-objective in this study is to model the price responsive behavior of battery storages, as representative elastic customers, using the optimization approach and analyze what factors drive their decision makings.

Objective 3: To propose a retail electricity market model with two price signals distinguishing between price sensitive and price insensitive market participants.

Despite the agreement made on the importance of retail electricity markets, discussions on how to structure the market and define market rules remain open. Therefore, the third sub-objective in the study is to analyze drawbacks of existing electricity market designs and propose an alternative retail electricity market solution beneficial for different types of market participants instead of only for the retailer itself.

Objective 4: To derive a bi-level day-ahead retail pricing scheme based on the proposed retail market model.

A well-designed pricing scheme is the key to a successful retail electricity market design. This study is particularly interested in the day-ahead retail pricing mechanism, as it plays an important role in eliminating the real-time retail price volatility, yet is seldom discussed in the literature. Given conflict interests of the retail market operator and price-sensitive market participants, this study proposes to apply the bi-level optimization technique in formulating the day-ahead retail pricing problem. To make sure the proposed retail pricing scheme is ready for the real-world implementation, network constraints are also considered in the problem formulation.

Objective 5: To convert the bi-level day-ahead retail pricing problem into a MILP problem.

The expense used to solve the bi-level day-ahead retail pricing problem is extremely high due to the existence of the three-phase unbalanced power flow model and coupling of the upper- and lower-level problems. It prevents the retail pricing tool to be widely applied. Therefore, the last sub-objective in the study is to fully convert the original problem into a MILP problem through utilizing the duality theory and linearization techniques.

1.4 Tasks

In order to fulfill objectives outlined above, following key tasks are identified to be accomplished in the study:

Task 1: Define a retail market environment, based on which subsequent tasks are carried out.

- Define the hierarchy of the retail market as well as its interactions with the wholesale market and end-users.
- Categorize market participants into two groups based on their sensitivity towards the market price.
- Identify decision makers in the retail market, analyze their objectives, decision variables and decision constraints.
- Propose a double-signal pricing scheme with an addition of grid service subsidy.

Task 2: Develop a Python-based test-bed system for simulating both physical behaviors of the power grid and financial interactions inside the retail market.

- Develop an IEEE 34-bus system model using OpenDSS and export system parameters from the OpenDSS model.
- Develop an inelastic participant agent, using Python class, which can be populated to predict fixed electricity demand/generation at different locations. (Take fixed load as the representative inelastic participant)
- Develop an elastic participant agent, using Python class, which can be instantiated under different equipment parameters to model the price-responsive behavior of battery storages. (Take battery storage as the representative elastic participant)
- Develop a retail market operator (RMO) agent responsible for determining market prices and managing the retail market operation.

Task 3: Propose DL-based day-ahead load forecasting models to be implemented by the inelastic participant agent.

- Collect building-level load time-series data from online resources and preprocess the data using data cleaning and data normalization techniques.
- Propose DL-based models for day-ahead load forecasting based on two popular DL frameworks: recurrent neural network (RNN) and convolutional neural network (CNN).

- Construct the RNN and CNN models under both direct multi-step and recursive multi-step manners.
- Evaluate performance of proposed DL models against traditional time-series method: seasonal autoregressive integrated moving average with exogenous variable（ARIMAX), from perspectives of accuracy, computational efficiency, robustness and generalizability.

Task 4: Develop a mathematical model representing the price-responsive behavior of battery storages to be implemented by the elastic participant agent.

- Formulate a battery arbitrage problem to represent the price-responsive behavior of battery storages in response to a day-ahead retail price.
- Run simulations to understand how the day-ahead price signal and battery parameters influence operating decisions made by the profit-oriented battery storage.

Task 5: Propose a bi-level day-ahead retail pricing scheme to be implemented by the RMO agent.

- Formulate the day-ahead retail pricing problem using the bi-level optimization technique.
- Migrate the battery arbitrage problem formulated in task 4 into lower-level problems.
- Formulate the upper-level problem as a day-ahead system resource scheduling problem.
- Merge the upper-level problem with lower-level problems through the Karush–Kuhn–Tucker (KKT) theorem.
- Linearize the merged single-level problem into a MILP problem.

Task 6: Conduct a comprehensive case study using the Python-based test-bed system.

- Verify the rationality of both inelastic and elastic customers to participate in the market.
- Compare retail market solutions obtained under different subsidy rates.
- Compare retail market solutions obtained under different levels of network constraints.
- Analyze how key equipment and market parameters impact the result of market settlement.
- Compare performance of the proposed retail pricing scheme against other alternatives.

2. LITERATURE RESEARCH

In this section a comprehensive literature research is conducted to better understand the state-of-the-art and knowledge gaps of the investigated topic.

Sections 2.1 and 2.2 review the development and application of different types of DERs. Section 2.3 shows how the distribution system deals with instability issues brought up by a high DER penetration. Section 2.4 reviews existing techniques for load forecasting. The established experience of wholesale market operation is introduced in Section 2.5 as it inspires the design of retail electricity market. Existing attempts made for designing the retail electricity market are summarized in Section 2.6. Finally, Section 2.7 identifies knowledge gaps.

2.1 Types of DERs

Distributed energy resources (DER) refer to energy resources that provide, consume and possibly store energy locally. These resources are generally small in scale and sited close to end-users. The DER technology is transforming the electric power sector as it decreases the reliance of distribution system on centralized upstream power supply and allows previously passive consumers to become proactive prosumers. Typical DERs include solar photovoltaic (PV), wind turbines, combined heat and power (CHP), diesel engines, micro-turbines, natural gas-fueled fuel cells, storage devices, electric vehicles (EVs), demand response (DR), etc., [7] which basically fall into three categories:

Non-dispatchable DERs are resources whose generations are usually intermittent, driven by uncontrollable factors (e.g. weather). Wind turbine and solar photovoltaic are two representative non-dispatchable DERs. They provide energy from wind and solar, which are cheap yet not continuously available. Also, their generations are time-dependent and vary across locations. Due to the government and industry incentives, deployment of the wind and solar technologies has gained a huge momentum during the past decade and is foreseen to contribute to more than 70% of the system generation in 2050 (see Figure 2-1).

- **Wind Turbine:** Most wind turbines installed today are large and tied directly to transmission system. In 2018, wind turbines in the United States were the source of about 6.6% of total U.S. utility-scale electricity generation. Electricity generation from wind in the United States increased significantly from about 6 billion kWh in 2000 to about 275 billion kWh in 2018 [8].
- **Solar Photovoltaic:** There are two types of PV systems, grid-tied system and stand-alone DC system. Most PV today are grid-tied. Whereas small-scale PV systems are interconnected with distribution system, utility-scale PV power plants are directly integrated to the transmission system. The U.S. Energy Information Administration (EIA) estimates that electricity generated by small-scale and utility-scale PV systems increased to 30 and 64 billion kilo-watt-hours (kWh) in 2018 compared to 11 billion kWh in 2014 and 76 million kWh in 2008 [8].

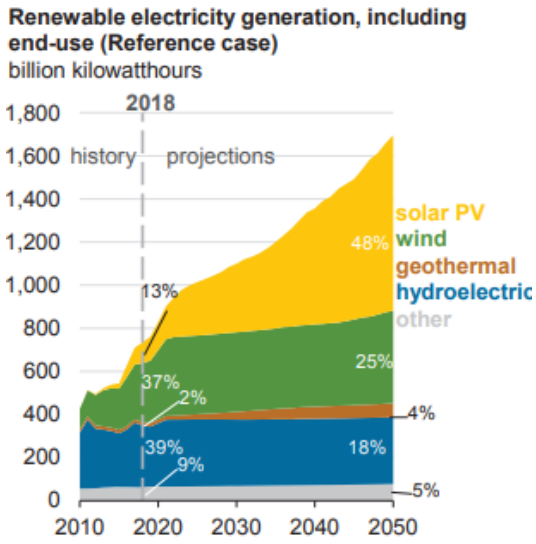


Figure 2-1. Long-term prediction of renewable energy generation [9]

Dispatchable DERs are resources that can be used on demand and be dispatched at the request of system operators. Its power injection/consumption can be precisely adjusted, as opposed to non-dispatchable DERs. CHP, diesel engines, micro-turbines, natural gas-fueled fuel cells and storage devices all fall into this category. During the past decade, deployment of the dispatchable DERs also expanded largely. For instance, since 2010, 728 new combined heat and power (CHP) systems have been installed across the United States, adding nearly 2 GW capacity to the U.S. grid. And many of these units are connected to a local distribution system [10].

- **Battery storage:** As a unique type of dispatchable DERs, battery storage can be both a load or a source. It is considered as an excellent solution to provide energy system flexibility and reduce renewable generation intermittency. Depending on the scale, battery systems are divided into two groups: utility-scale battery systems (with power capacity greater than 1 MW) and small-scale battery systems (with power capacity smaller than 1 MW). Decreased price and supportive policy have driven the deployment of both types of battery storages. As stated in the U.S. battery storage market trend report, installed capacity of large-scale battery systems has nearly doubled every two years since 2011. By the end of 2017, 708 MW was in operation. Similarly, the size of small-scale storage doubled from 66 MW, as reported by U.S. utilities in 2016 [11], to more than 120MW in 2018 [12]. Despite the higher capacity share of utility-scale battery systems, there is a greater expansion trend of the behind-the-meter storages. The installation capacity of behind-the-meter storages in 2018 almost tripled that of 2017, and matched the utility-scale storages investment for the second consecutive year [13]. This change may be a result of increased electric vehicle (EV) ownerships and favorable policies that encourage storages to be located near renewables.

Demand Response (DR): are resources offering flexibility via changing electricity usage from normal patterns in response to different financial rewards. Based on type of the reward, DR can be classified as price-based or incentive-based. Price-based DR relies on dynamic pricing, based on which users vary their consumptions across time to cut down their electricity bills. In this case, users own the full autonomy in deciding the tradeoff between their electricity need and money

saving. Incentive-based DR, on the other hand, relies an agreement made between system operators and electricity customers. Such an agreement predefines incentive payment and specifies the load control requirement. Users who have signed up for this program are usually required to either shift the time for using their most energy-intensive applications (Load shifting) or curtail the consumption during specific period to not exceed certain limit (Load curtailment) in an effort to diminish the system peak demand.

Based on data collected in EIA's annual survey of electric power sales, revenue and energy efficiency [8], in 2017, there were about 9.44 million customers in the United States that participated in demand response programs, with most (88%) of them falling into the residential sector. On average, each residential customer contributed to about 114 kilowatt-hours (kWh) energy saving and 1.1 kilowatt (kW) peak demand reduction, and, in turn, received about $35 reward. Despite the small share of commercial and industrial participants (11 percent and less than 1 percent, respectively), they provided more than half of the total peak demand reduction for much larger incentive payments.

Smart buildings with active energy management system (EMS), as depicted in Figure 2-2, are excellent DR resources [14]. First, buildings account for 40% of the energy consumption in the U.S., and most of them do not consume power in a very efficient way [15], resulting in a huge space for energy saving. Second, by slightly changing electricity usage patterns of the buildings, a great amount of energy saving or peak reduction can be achieved without significantly impacting user comforts. Studies have extensively investigated how the electric appliances, include Heating Ventilation and Air Conditioning (HVAC) [16][17][18][19][20], water heater [21][22], dishwasher [23] or EV, that commonly available in the buildings, can work independently or cooperatively to provide DR services [24]. Even more flexibilities can be offered with the addition of household PV-battery systems [25][26].

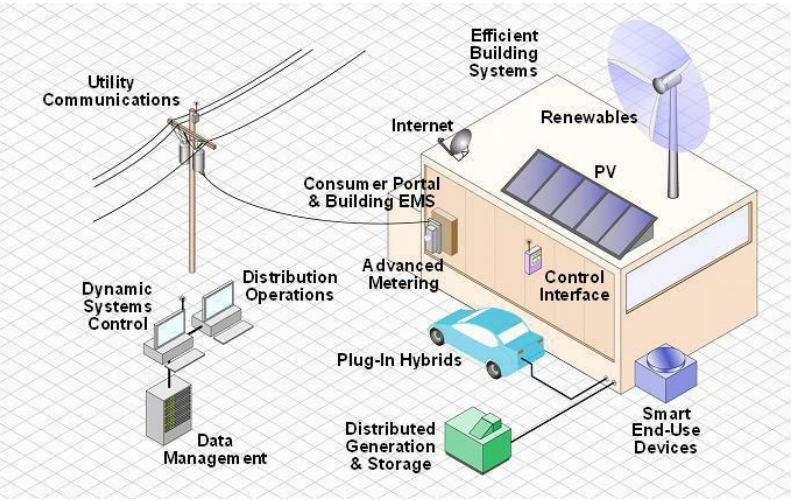


Figure 2-2. Smart building with active EMS [27]

2.2 Applications of DERs

An extensive range of grid services can be offered by DERs. Some of the most widely applied services are summarized below:

- **Frequency regulation:** is a tool that helps correct the short-term changes in electricity use or generation to maintain the system frequency of 60 Hertz. System operators provide market-based incentives to the frequency regulation participants in return for their fast responses to an automated signal. By following the control signal, DERs help the system track the moment-to-moment fluctuations in customer electricity use and correct unintended deviation between scheduled and actual generations. A variety of DERs are applied in the literature for frequency regulation. Collections of homogeneous loads, such as electric vehicles (EVs) and thermostatically controlled loads (TCLs), were discussed in [28], [29], [30], [31], [32] and [33]. [34] proposed the use of aggregators for integrating heterogeneous loads, such as plug-in EVs, TCLs and micro-turbines, to provide frequency regulation based on model predictive control (MPC). In [35], a stochastic frequency regulation method was proposed to facilitate the system operation under uncertainty. Authors in [32] reviewed three different dynamic demand control strategies for frequency regulation using residential and commercial loads. They are respectively centralized control, decentralized control and hybrid control. The work in [36] analyzed the challenges faced by frequency regulation using DERs. Authors in [37] proposed a framework for using aggregated DERs to provide frequency regulation while considering system physical constraints. The real-time regulation signal allocation problem for multiple aggregators was also investigated in [38].
- **Load following:** requires participants to inject or absorb power for compensating load variations, also known as a form of ramp rate controls. It works similar to the frequency regulation, but has slower response time requirement (usually larger than 10 min). Such applications are studied in [39], [40], [41] and [42], adopting dispatchable DERs, e.g. micro-turbines, fuel cells and energy storage system (ESS).
- **Generation reserve:** refers to synchronized electricity capacity that is not currently in use but may become available within a short delay, when a significant frequency disturbance occurs. For instance, when there is an unexpected failure of committed generator, generation reserves need to quickly respond and provide back-up. Based on the requirement of response speed, generation reserves can be grouped into primary reserves (response within 10 min) and supplemental reserves (response within 10-30min). Unlike frequency regulation and load following which deal with the system normal operation, generation reserve is only activated when system encounters contingency issues. It is usually considered together with the normal operation as a co-optimization problem. Authors in [43] considered a decentralized scheduling problem of DERs in a day-ahead distribution market that clears both energy and reserve offers at the same time. Reference [44] modeled the bidding strategy of a distribution company in the wholesale energy and reserve market as a bi-level optimization problem. A two-stage stochastic programming problem was formulated in [45] for solving the virtual power plant (VPP) trading in a joint energy and reserve market. Authors in [46] specifically discussed how TCLs can be modeled and utilized for providing operating reserve. Conditional value-at-risk (CVaR)

was applied to control risk of the profit variability. Not only dispatchable DERs, but also non-dispatchable DERs are discussed for providing such co-optimization services [47].

- **Voltage regulation:** aims at maintaining voltage level of the system to ensure quality of the delivered power [48][49]. It is also called reactive power support as it requires the participant to serve as a source or a sink of reactive power. DERs for voltage regulation are also extensively studied in the literature, within which PV and Battery storage have gained the most attention. Authors in [50] have investigated how smart inverters can be applied to relieve the voltage instability brought about by intermittent PV generations, based on which, a market-based compensation scheme was proposed for incentivizing the PV panel owners. A PV generation probabilistic model was proposed in [51], based on which, authors developed a probabilistic strategy for optimized use of PV generation via treating PV generators as capacitors or inductors. In [52], a combined method applying the battery energy management of plug-in electric vehicles (PEVs) and the active power curtailment of PV panels was proposed to help regulate voltage in a low-voltage distribution system. By implementing the multi-agent framework, [53] proposed an optimal voltage regulation approach with participations of multiple microgrids. Heterogeneous DERs for voltage regulation with continuous or discrete control command were also covered in [54].
- **System peak shaving:** is the process of reducing the amount of energy purchased during peak demand hours. It is of significant importance from the economic aspect, as it reduces or defers the need of building new power plants or purchasing capacity from the capacity market. A day-ahead battery storage scheduling problem was analyzed in [55] for minimizing the load variance in a low-voltage unbalanced distribution system. Authors in [56] applied MPC to coordinate heterogeneous flexible DERs to reduce the peak load. In [57], a novel Artificial Neural Network (ANN) based controller was proposed based on factors of load condition, storage capacity, and state-of-charge (SoC). It is designed to coordinate PV systems, V2G (vehicle to grid)-capable EVs, and ESS for reshaping the load curves.
- **Congestion management:** is the process of adjusting power injections or absorptions of multiple DERs, such that power flows in transmission or distribution lines can be reconfigured to eliminate congestion concerns. It allows deferral or complete avoid of congestion-related upgrading costs. Market-based congestion management approaches were proposed in [58] and [59]. By employing the distribution-level locational marginal price, the proposed market mechanism is able to reflect the real congestion cost and further alleviate potential congestions in the system.
- **Black start:** describes a process of re-energizing the transmission and distribution system under a catastrophic failure of the power grid (also named blackout event). Only DERs equipped with the black start capability are able to provide such services. A few papers discussed the possibility of applying DERs for black start services, within which PV-Plus-Battery system has gained the most attention. [60] talked about a black start strategy deploying PV-Plus-Battery systems applicable for both single-phase and three-phase systems. [61] emphasized the optimal sizing and siting problems of PV-Plus-Battery systems for improving the power system resilience. A detailed operation scheme and a

restoration procedure were discussed in [62] and [63] to help distribution systems survive in the islanding mode and restore to normal operation as soon as possible.

- **Energy management:** refers to the optimized coordination among multiple DERs and other resources that satisfies the system electricity need at a minimum operation cost. Based on control horizon, it may be categorized as a real-time dispatch problem or a time-dependent scheduling problem. Under a well-designed energy management framework, resources with different characteristics can be guided to optimize the system performance. Energy management problem in an islanded microgrid was investigated in [64]. A three-party energy management framework, among the power grid, residential units and shared facility controllers, was constructed as a non-cooperative Stackelberg game in [65]. Authors in [66] proposed a distributed online energy management algorithm to deal with highly uncertain generations. The reinforcement learning technique applied for energy management was also discussed in [67].
- **Arbitrage:** is an application particularly applicable for the battery storage under dynamic pricing mechanism. Profits occur when batteries charge at lower electricity price and discharge at higher electricity price. Recently, significant increases in prices and price volatility of natural gas and electricity have raised interest in potential economic opportunities of electricity storages. The potential arbitrage revenue of ESS in the PJM market was analyzed in [68] and [69], indicating that many factors impact the value of battery storages. For examples, hours of storage, efficiency, imperfect forecasting and price variation. Authors in [70] discussed multiple applications of battery storages, and were particularly interested in co-optimization between frequency regulation and energy arbitrage. Battery degradation was also taken into account when designing the energy arbitrage strategy [71].

2.3 Distribution system with high DERs penetration

Despite benefits brought by DERs, the growth of grid-connected DERs can be both a blessing and a curse. Increased deployment of DERs results in a more flexible and decentralized grid. DER owners are interested in providing grid services to system operators, e.g. Independent System Operator (ISO) and Distribution System Operator (DSO), to minimize their electricity costs or maximize their revenues. At the same time, system operators are interested in integrating DERs to achieve optimal system operation. However, there is a lack of visibility between these two parties. On the one hand, DER owners lack the awareness of system operation. It is hard to ensure the exported energy of itself is feasible and deliverable. On the other hand, system operators lack the understanding of how the operation of multiple DERs will impact their system. Furthermore, DER availability is harder to predict than traditional generators, adding more complexity to the system control.

Many concerns are raised in the literature regarding high DER penetration. Two reactive power control strategies were proposed in [72] and [73] aiming at addressing the under-/over- voltage issues raised by high PV penetration, based on a nonlinear and a linear grid models. Authors in [74] found that voltage issues brought about by large-scale PV systems installed in the rural area cannot be effectively solved by reactive power compensation, as the R/X ratio of distribution

system is high. Rather, a voltage regulation approach that coordinates upstream step voltage regulator (SVR), downstream PV inverters and BESS was proposed in [74].

Another well-known consumption saving approach which is often used in centralized distribution network structures but could not serve in decentralized systems is the Conservation Voltage Reduction (CVR), relying on step-voltage regulation. To address the challenge brought by decentralized structure, a three-phase AC optimal power flow (OPF)-based CVR scheme, using on-load tap changers and capacitors, was developed in [75] that worked even under high PV penetration. In [76], a robust energy management framework was developed with emphasis on tackling the intermittent availability of renewables. Not only the average-case costs of DG, BSS and flexible loads, but also the worst-case costs stemming from the uncertainty of RES were considered to be minimized.

2.4 Load forecasting techniques

Load forecasting plays a vital role in managing the operation and planning of power systems. It helps electric utilities or independent system operators to make important decisions including purchasing and generating electric power, load switching and infrastructure development. From the perspective of prediction horizon, it can be divided into short-term (one hour to one week), mid-term (one week to one year) and long-term (more than one year) load forecasting problems. Based on the scale of the investigated object, it can be divided into system-level, building-level and appliance-level load forecasting problems. The growth of the DER integration results in new trends in the study of load forecasting. First, building-level and appliance-level load forecasting become more important, as it is essential to understand behaviors of individual DERs. Second, system-level load forecasting becomes more challenging as existence of DERs add up uncertainty at the system-level.

A large variety of statistical tools have been investigated with regard to load forecasting, ranging from the simplest multiple linear regression (MLR) [77], [78], autoregressive integrated moving average (ARIMA) [79], [80], support vector regression (SVR) [81], [82] to the most sophisticated artificial neural network (ANN) [83], [84], [85]. Hybrid prediction model combining the capabilities of these non-DL techniques is also discussed in [86]. With the huge success Deep Learning (DL) techniques made in solving complex statistical problems, researchers also started searching for the DL-driven solutions for load forecasting applications [87], [88]. Essentially, deep neural networks boost the power of ANN via deepening its layers and leveraging its structures. These learning methods have been widely implemented to solve natural language processing and speech/image recognition problems. Not surprisingly, DL techniques in application of time-series day-ahead load forecasting are starting to gain the attention of the research community. Within the non-DL scope, ARIMAX has overall competitive advantages for load forecasting. Compared to MLR, it is more capable of capturing the temporal dependency; compared to SVR and ANN, it shows better interpretability. Unlike other supervised learning models, ARIMAX is tailored for time-series modeling, where the sequence of inputs matters. As the counterpart of ARIMAX in the DL community, recurrent neural network (RNN) is used to tackle datasets with sequential correlations. Authors in [89] proposed two RNN structured models for medium-to-long term predictions. Authors in [90] proposed a pooling deep recurrent neural network (PDRNN) model to forecast household load. Nevertheless, computational efficiency was not compared. Also, this

method requires an access to neighbors' load data, which is not practical. Authors in [91] innovatively created an autoregressive-like double-CNN model for asynchronous time-series forecasts. Experimental results demonstrated its predominant capability at handling various kinds of asynchronous datasets, including household load data. However, prior knowledge on the next day's weather condition is missing in the problem context [92]. Moreover, authors in [91] only analyzed the performance of single-step building-level load forecasts, yet did not examine multi-step performance. The application of deep learning on feature extraction was also discussed in [93], indicating a significant improvement in prediction accuracy after applying an unsupervised deep learning model. It is also found in [93] that supervised deep learning did not show obvious advantages over other conventional tools. However, the discussion was only limited to the perceptron deep neural network, excluding the families of RNN and CNN networks.

2.5 Wholesale electricity market

To better design retail market rules for active and decentralized distribution systems, it is necessary to understand first how the wholesale market works. Despite the difference among market rules, there is a common two-settlement structure applied across all the wholesale markets. A two-settlement market refers to a combination of a forward (day-ahead) market and a spot (real-time) market with each producing its own financial settlement.

The forward market schedules the production and consumption before the operating day to help avoid price volatility, whereas the spot market reconciles any deviations between day-ahead commitments and the actual real-time demand and production to ensure the system security. Based on the generation and demand bids, the system operator settles the forward market with cleared day-ahead prices plus cleared quantities of generation and demand. Such a settlement can be either financially or physically binding. At the spot market, scheduled suppliers must produce the committed quantity or, otherwise, buy power to balance their positions. Similarly, scheduled buyers earn the right to consume cleared quantities at cleared day-ahead prices, and, if needed, purchase extra needs from the spot market. Despite the commitment made in forward market, the scheduled generation usually does not match the actual real-time load. Spot market plays a significant rule on balancing such a difference. It produces a separate, secondary financial settlement for only the deviated part from the day-ahead commitments. Generators and customers can offer/purchase more energy into/from the spot market. Deviations from day-ahead obligations will be settled at the real-time prices.

Such a two-settlement structure brings two major benefits to the electricity market. First, it helps reduce the price volatility. Unexpected system behaviors, like large demand variance and generation failure, result in price volatility. Without having the commitments made by generator units and customers in the forward market, system behaviors will be more fluctuated and unpredictable. Therefore, existence of the forward market provides a hedge against the spot price risk. It improves the operational certainty for system operator, secure the financial benefit of generator companies and helps the customers reduce their vulnerability to price fluctuations. Furthermore, such a mechanism can help mitigate the market power manipulation. Market under the two-settlement design works the most effective and stabilized when generators and customers follow their patterned behaviors. Manipulating the market power will cause unexpected price

fluctuation, and in turn harms the interests of suppliers and buyers themselves. Therefore, the incentive to manipulate the market is limited.

Different from other financial markets, operation of the electricity market, especially the spot market, requires a constant interaction between the market and the underlying physical power system. A typical framework of the electricity market management system usually comprises three components: system planning and operation (SPO), market operation (MO) and financial settlement (FS) [2]. The SPO fulfills system measurements and provides essential operating parameters, such as load forecast, system topology, power flow analysis and constraints management to the MO. Based on the information obtained from SPO, the MO determines generation schedules back to the SPO and publishes the cleared prices to the FS. Generation schedules and cleared prices are usually obtained by solving an optimization problem within which there are four common components: Unit Commitment (UC), Economic Dispatch (ED), Security Constraints (SC) and Pricing. Note that SC is usually combined solved with the UC and ED as SCUC and SCED problems.

Unit commitment produces the optimized on/off status of each generating unit at each control interval. The unit operation cost for serving forecasted system net load is minimized during the UC while taking unit operational restrictions (e.g. ramping speed, minimum up/down time) and system constraints (e.g. energy balance, transmission line congestion) into account. Both the day-ahead market (DAM) and real-time market (RTM) require solving the UC problem over a 24-h and 5-min horizon. The nonconvex nature and integer control variables make the UC problem hard to solve. Among all the techniques, MILP is the most commonly used one for solving the UC problem, whose efficiency is heavily relied on how the problem is formulated.

On top of the on/off decisions made by the UC, economic dispatch process determines the specific generation output of each generation unit at each control interval. It aims at minimizing the total dispatch cost based on unit price curves submitted by generator units, while subjecting to the equipment and system constraints. Depending on nature of the cost curve, the ED problem can be formulated as a linear or a quadratic program. Moreover, since the availability of units at each interval are determined, the ED problem is a pure linear or nonlinear problem without integer control variables.

The Security Constraints component further examines the UC and ED solutions against N-1 contingencies. For each contingency, a power flow problem is solved to check and identify any network constraint violations. The violated constraints are linearized and fed back to the next iteration of the ED problem, forming the inner loop of SCED. If the SCED cannot resolve all security constraint violations, then the remaining violated constraints are fed into the next iteration of UC until all violations are resolved, forming the SCUC problem. Security analysis could be time-consuming due to the large number of contingencies. Because simulations of different contingencies are independent, parallel computing techniques are often adopted.

Whereas SCUC and SCED help yield a secure commitment and dispatch solution, market pricing is the key for a well-designed market. It serves as signal that guides the behaviors of market participants. Under a proper pricing system, the units will have no incentive to deviate from the ISO's schedule, which is desirable as it enforces market participants to bid truthfully. Moreover,

customers will be encouraged to consume energy more efficiently, with less energy efficient consumptions being charged at a high rate. Most wholesale market nowadays follows a Locational Marginal Price (LMP) pricing mechanism, which is calculated as the marginal cost for supplying the next incremental load at different location. The calculation used to determine LMPs takes a comprehensive consideration of actual operating conditions: power flow, system balance, voltage profile and line congestion, therefore, helps enhance efficiency and reliability of the electric grid. Based on the LMPs, system operator could understand where the system is stressed and therefore allocate more resources to the location; Customers, especially industrial customers, could chose to locate at a place with lower prices to reduce their energy cost. In addition, since congestion impacts the LMP as an important factor, LMPs also precisely indicate the need for infrastructure upgrade with a high congestion price. LMPs are usually obtained within an optimal power flow (OPF) framework, which basically falls into two categories: fully-structured AC-OPF and linear-approximated linear-OPF (including DC-OPF). Due to the great characteristics of convergence and computational efficiency, linear-OPF is widely applied in the wholesale markets.

2.6 Attempted retail market deregulation

It is clear that distribution system could not fully take advantage of the technical and economic benefits brought about by DERs without an efficient retail market design. Currently, efforts have been devoted to partially deregulate the retail electricity market by providing diverse choices of retail programs. Retail markets in over 40 states and districts of United States have opened the net metering and DR programs to the customers [94]. They are encouraged to provide self-production or grid services in exchange of financial compensations, which vary with different state legislation and regulatory policies. However, these programs only provide limited flexibility and will become insufficient as the implementation of DERs keep expanding. Hereby, it is requested to build a next-generation retail market with innovative business model and strategic pricing scheme to accommodate greater DERs penetration.

Despite the advance wholesale market made on managing the operation and planning of transmission system, its successful experiences cannot be directly borrowed by the retail market liberalization. First, most existing wholesale market designs focus on the generation side, however, a higher emphasis has to be put on the demand side of the retail market [95]. Second, price-responsive behaviors of the DERs largely challenge the price volatility in the retail market, which are generally not encountered in the wholesale market [96]. Third, the high resistance to inductance ratio and unbalanced characteristics of the distribution network [97], plus complicated physical dynamics of various DERs make the system physic models extremely computationally demanding. Finally, the interaction between the wholesale market and the retail market has not been explored enough, as it is not a concern when designing the wholesale-only market. Papers [2], [98], [99] provided the vision of next-generation retail electricity market and summarized expected characteristics of the market to be flexibility, diversity, decentralization and competitiveness. Under such a consensus, many attempts have been made in the literature to construct the future retail electricity market.

Modeling price-responsive operations of DERs is a crucial step in designing the retail market. There are two commonly applied practices in the literature. The first practice directly applies price elasticity curve to express the price-responsive behavior, which maps relationship between the

willingness of a DER to transact energy and the unit price. Figure 2-3 shows a typical example of a demand elasticity curve. It tells that the customer is willing to purchase more energy as the price goes higher. Both supply and demand elasticity curves build basis of the double auction-based transactive energy system (TES) researches [100][101][102]. They precisely quantify flexibilities to be offered by the market players under different prices, based on which, market organizer clears the market at a point where supply and demand reach a balance and the social welfare is maximized. The second practice obtains price-responsive behaviors within an optimization framework, where a user-defined objective function is optimized in response to a time-varying price signal. The solution offers optimized dispatch schedule for the DER. Whereas in the first practice price-sensitive behavior can be formulated as an explicit equation of price, it is not the case for optimization-based modeling approach. However, the second practice could better capture the dynamic behavior of the DER and is therefore applied more in the day-ahead market setting. In comparison to that, the elasticity curve cannot reflect the temporal dependency of energy consumption/generation, and is therefore mostly applied in the real-time market setting.

It is worth mentioning that, if all the market players only participate in the real-time market and no commitment is made in the day-ahead market, price-responsive DERs would constantly change their behaviors in response to dynamic real-time price, which in turn impacts the future price and jeopardizes market stability. Therefore, in this study, a higher emphasis is put on the day-ahead retail market design, as there is no other good solution available for the volatility problem faced by the real-time pricing under wide-deployment of price-sensitive DERs [103][104].

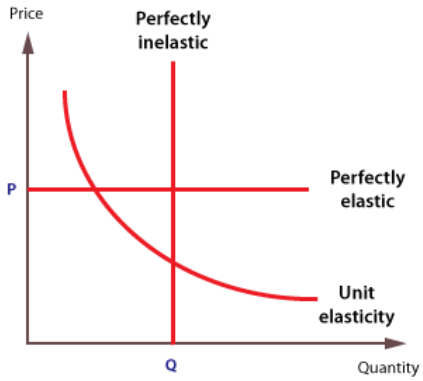


Figure 2-3. Price elasticity curve

Figure 2-4. Hierarchical structure of the retail market

The retail market design is usually considered as a hierarchical decision-making problem, as shown in Figure 2-4, where two autonomous decision makers: retailer/DSO and market player, with possible conflicting interests, interact with each other, as well as with the external wholesale environment. Retailer/DSO plays a role of intermediary agent between the end-users and wholesale market. On the one hand, the energy price goes from the upstream wholesale market down to the end-users, on the other hand, the energy volume is aggregated from the downstream market player layer up to the wholesale market. Multi-level/Bi-level programming is mentioned in the literature as an ideal approach to formulate such a problem, due to interdependency between decision-makings inside the hierarchical structure.

Although similar bi-level/multi-level structure is applied across literatures, the underlying market rules and business models are designed to be quite different. In [105], a bi-level problem was formulated to analyze the competitive situation between the DERs aggregator and the central generation unit. It follows a Stackelberg game structure, where the central generation unit plays a role of leader and the DERs aggregator acts as a follower, both aiming at maximizing their own profit. The leader acts first to pick a profile margin, and the follower acts after to decide the load curtailment and DG outputs. Such a decentralized problem formulation was compared with the centralized case where all the resources, including DERs and central generating unit, are assumed to be managed by a same entity. Experimental results indicate that decentralized problem setting allows higher level of load curtailment. The bi-level optimization technique is also applied in [106] for proposing the bidding and pricing strategy of load servicing entity in a competitive retail electricity market. Authors in [96] proposed a two-stage two-level retail market operation problem. Two levels stand for two autonomous decision makers: retailer and customers. Two stages represent a day-ahead pricing step and an energy resources dispatch step. Uncertainty of the wholesale price is also considered in [96] by converting the energy resource dispatch problem into a min-max problem that seeks to maximize profit under the worst case. Case studies proved effectiveness of the two-stage two-level retail market operation approach. Authors in [107] focused their work on the interaction between multiple utilities and multiple customers in a single-period retail market operation, where utility companies compete with each other via optimizing their price strategies, end-users respond to the prices to optimize their energy usages. Comprehensive analysis that covers different aspects of the market effectiveness is provided in [107]: First, existence and uniqueness of the Stackelberg equilibrium (SE) is proved; Second, a distributed solution is derived for the convergence of (SE) without each utility knowing the offer and parameters of other utility companies; Third, stability of the market, under market manipulation, is examined. Given that, a scheme based on shared reserve power is proposed to enhance the stability. Authors in [108] developed a platform simulating the scheduling and bidding processes in a wholesale-local integrated market. Based on the platform, a two-stage stochastic problem was derived, modeling the first stage day-ahead biddings and clearings process, and the second stage near real-time DER dispatch. Operation of the proposed transactive energy system under two different scenarios, when there is a market access to the wholesale market or not, are compared in the case study. Results highlight the potential flexibility and cost-saving benefits provided by local markets. Discussions regarding local energy trading can also be found in [109], [110]

The aforementioned papers discussed various possible interactions among the retailer, end-users and wholesale market. It offers a great beginning for the retail market design and proves the advantage of providing DERs a retail market access. However, these discussions are still limited

at a conceptual stage in the sense that DERs are treated indiscriminately without considering the impact of their physical locations. As a complex network, a distribution system usually comprises of a great number of interconnected nodes sharing correlated physical characteristics, like, voltage phasors, currents and power flows. Due to the complex interdependency, a small amount of disruption at a single node can cause changes in operating conditions over the whole network. Such changes are usually captured by solving a power flow problem. Therefore, to design a practical retail market algorithm, the optimized DERs dispatches should also respect to the power flow equations, which turns the retail pricing into an optimal power flow (OPF) problem.

Recently there has been a strong research interest on the topic of distribution-level LMP (DLMP) which provides a possible retail pricing solution under consideration of the locational distinction. As a borrowed idea from the wholesale market's LMP, DLMP helps quantify the minimized marginal cost for supplying next incremental load at different locations. It can be derived from a distribution-level OPF formula, as it essentially presents the sensitivity of the optimized objective towards incremental loads. The idea of DLMP was first introduced in [111], where it was shown that by applying the nodal pricing, a significant difference among the DLMPs at various locations could be noticed, reflecting high line losses in the distribution system. The DLMP model thus can be utilized to properly reward DGs and incentivize customers to reduce line loses, as it offers high selling/buying prices at the places where are over-loaded. However, the proposed DLMP model assumed an unrealistic distribution network with single-phase structure and no congestion and voltages problems. A relatively more complete DLMP model is proposed in [112] by taking congestion and voltage support into account. In addition, it extends the work in [111] to be applied in a day-ahead manner. Yet, the price difference brought about by three-phase unbalanced loading is still ignored. Further, in [113] a fully-structured three-phase DLMP model is proposed. It derives DLMP within a nonlinear and nonconvex AC-OPF framework, that seeks to minimize the overall generation cost for a given load profile while subjecting to three-phase power balance equations as well as voltage, line flow and unit operation constraints. Case studies are deployed on a 60-bus system, developed based on a real campus grid, to demonstrate functionalities behind the DLMP model. Numerical and analytical results show that difference across locations and phases can be correctly captured. Based on the DLMPs, system operator could easily detect location or phase being stressed. Moreover, it demonstrated that potential market participants in distribution systems could benefit from the DLMP-based markets while driving down the cost of the entire distribution system. Nevertheless, as pointed out in the paper itself, computational efficiency is a concern of the AC-OPF based approach. Linear-OPF, on the other hand, is converted from AC-OPF model through applying approximation or linear relaxation technique. Despite the compromise made on precision, less computational burden, scalability and proof of convergence can be achieved by the linearized model. As mentioned in Section 2.5, linear-OPF models are widely applied in the wholesale market. Nevertheless, in the linear-OPF formulation applied to the transmission system, network losses and three-phase unbalance are generally ignored, which is not the case for the distribution system. Due to the high resistance to reactance ratio and unbalanced nature of the distribution network, linearization of the distribution-level OPF problem becomes very challenging. Only few studies have explored this topic. Based on a linearized power flow model, a linearized OPF model and the distribution system loss factors, a linear DLMP model is proposed in [114]. However, it is only applicable for radial structured network. Moreover, distribution system imbalance is not discussed. A multi-phase linear DLMP model was then proposed in [115] to cover more generalized situations: to support both radial and meshed network structures, wye

and delta connections and balanced and unbalanced loadings. Both of the linear DLMP models proposed in [114] and [115] are decomposable. In other words, the integrated DLMP can be broken down into several parts that represent the cost contributed by power balance, transportation loss, voltage support and congestion support. Overall, current DLMP studies provide another valuable point of view on how the retail pricing should be implemented while taking the physical feasibility into account. However, practicability of the DLMP is being doubted due to the concern of price fairness, as it is hard to justify why the energy unit prices of households within a same neighborhood should be different. In addition, engagement of the price-responsive DERs is not considered in most existing works.

2.7 Knowledge gaps

Based on the literature review presented above, following knowledge gaps are identified.

With regard to the day-ahead load forecasting:

1. The DL-based approach for day-ahead load forecasting has not been fully compared with the traditional time-series approach.
2. There is a great variety of deep neural networks, e.g. RNN and CNN, however, which one performs the best in the application of building-level day-ahead load forecasting has not been comprehensively discussed.
3. Traditionally, the day-ahead load forecasting problem is solved in a recursive multi-step fashion. However, DL-based models have the flexibility to map multi-to-multi relationships and thus can be constructed under both direct and recursive multi-step fashions. How such a feature of DL-based models could benefit the day-ahead load forecasting is not specifically investigated.

With regard to the retail market design:

1. Many papers ignore the impact of price-responsive behaviors back towards the market price, when designing the retail pricing. However, such an assumption will not hold as the penetration of price-responsive DERs keeps increasing.
2. Most papers limit their discussions on the real-time retail market design. However, without a well-functioning day-ahead retail market the real-time retail price volatility issue can hardly be avoid.
3. Despite efforts made on the conceptual design of retail market. A joint consideration of market operating rules and power grid network constraints is generally missing in the literature, which is required for the practical implementation of the pricing tool.
4. Most papers assume the retail market operator (RMO) to be profit-oriented. However, market prices determined by these RMOs tend to benefit themselves rather than end-users. Moreover, aggressive market rulers may even endanger the system security.
5. A single-signal pricing scheme is commonly applied in the literature. To the best of the author’s knowledge, no paper has considered financially distinguishing customers based on their sensitivity to the market price.

3. RETAIL MARKET ENVIRONMENT SETUP AND TEST-BED SYSTEM DEVELOPMENT

This section completes tasks 1 and 2 in Section 1.4. It offers the big picture of how contents presented in following sections are connected. Section 3.1 first explains the retail market model, including market hierarchy, components and interactions among the components. After that, Section 3.2 introduces how the simulated retail market environment is built in the study through Python and OpenDSS programming.

3.1 Retail market model

3.1.1 Market hierarchy

The distribution-system-scaled retail market model applied in this study follows a tri-layer structure, as illustrated by Figure 3-1. It provides the market context of following discussions. Retail market operator (RMO) stays at the middle layer and interacts with the wholesale market, at the top-layer, and market participants, at the bottom-layer, through energy and financial transactions. There are two types of dispatchable energy sources available in the system. First is the supply from the bulk grid. It can be imported from the substation and purchased from the wholesale market at the wholesale price. Second is the additional flexibility provided by elastic customers. It is accessible at distributed locations and can be purchased in forms of financial contracts with elastic customers. The additional flexibility given by elastic customers decreases the dependency of the distribution system on bulk grid leading to a more cost-efficient system operation.

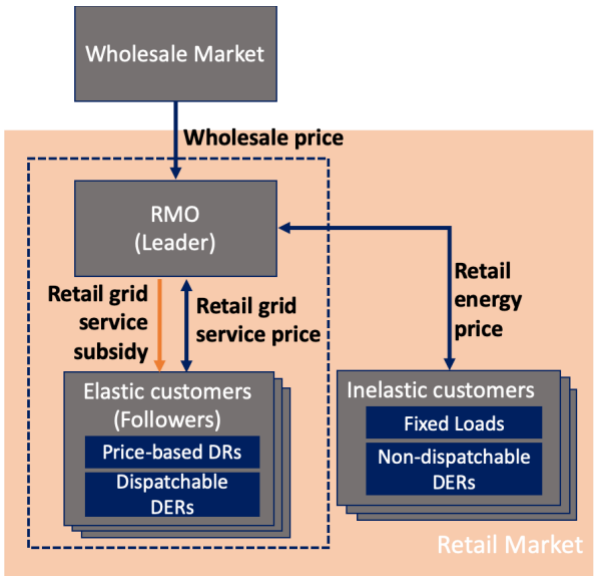


Figure 3-1. Structure of the retail market model

3.1.2 Assumptions made in the market

Several assumptions are made on the underlying retail market setting:

- **Ownership of elastic assets:** It is assumed ownerships of elastic assets belong to private companies or households, who operate independently and have no share of interest with the RMO. It is important to be clarified, since utility-owned elastic assets usually has no incentive to respond to the market price. Although utility-owned elastic assets are more capable of providing the flexibility, since the system operator has a full authority on controlling these devices, the objective in this study is to better utilize elastic assets available from private companies or households, that may sum up to a great amount of flexibility resources.
- **Profitability of the RMO:** In the future, there may be multiple RMOs serving the same region, who compete to attract customers by offering cheaper retail energy price. Therefore, unlike other existing designs, the RMO in this study is assumed to be profit-neutral. In other words, their decisions are not driven by maximizing their own profit but reducing the overall system cost for serving price-insensitive demands in the market.
- **Role of the RMO in the wholesale market:** The RMO is assumed to be a price-taker of the wholesale market. Such an assumption is made to simplify the interaction between the wholesale market and retail market, so as to put a higher focus on interactions within the retail market.
- **Visibility of behind-the-meter resources:** It is assumed the RMO has the full knowledge of key parameters and initial operating status of elastic assets that necessary for understanding their price-responsive behaviors.

3.1.3 Categories of market participants and the double-signal pricing scheme

Unlike other existing studies, market participants in this study have been innovatively classified into two groups: inelastic customers and elastic customers. They are distinguished based on the sensitivity towards the market price. Table 3-1 summarizes major characteristics of them.

Table 3-1. Types of market participants

Type of customer	*Elastic customers*	*Inelastic customers*
Price signal	Retail grid service price (β_{GS})	Retail energy price (β_E)
Sensitivity to the price	Yes	No
Examples	Price-based DRs Dispatchable DERs	Fixed Loads Non-dispatchable DERs
Market Settlement	$\boldsymbol{\beta_E \times Q_{ine}}$	$\boldsymbol{\beta_{GS} \times f(\beta_{GS}) + s_{GS}}$

Inelastic customers manage the part of load/supply that keeps fixed no matter how the market price varies. It is charged/paid at the retail energy price $\boldsymbol{\beta_E}$. Typical inelastic customers include fixed loads and non-dispatchable DERs. Elastic customers, on the other hand, manage the part of

load/supply sensitive to the market price. It is charged/paid at the retail grid service price $\boldsymbol{\beta_{GS}}$. Price-based DRs and dispatchable DERs fall into this category. The market settlement of inelastic customers is presented in the form of gross revenue. It contains two parts: grid service income $\boldsymbol{\beta_{GS} \times f(\beta_{GS})}$ and grid service subsidy $\boldsymbol{s_{GS}}$, where $\boldsymbol{f(\beta_{GS})}$ stands for the elastic load/generation that can be expressed as a function of $\boldsymbol{\beta_{GS}}$. The market settlement of inelastic customers is presented in the form of electricity bill, which can be calculated as $\boldsymbol{\beta_E \times Q_{ine}}$, where $\boldsymbol{Q_{ine}}$ represents the quantity of inelastic load/generation. Correspondingly, a double-signal pricing scheme is proposed given the existence of two price signals. The intuition behind this separation is that instead of treating elastic customers as normal electricity consumers, it is more reasonable to treat them as grid service providers, who adjust their energy consumptions from normal patterns to help improve the system efficiency. Therefore, flexibilities provided by them should be valued differently. In addition to these double price signals, a grid service subsidy ($\boldsymbol{s_{GS}}$) is added in the pricing system to provide supplementary incentive to elastic customers. The reason behind adding this grid service subsidy is explained in Section 3.1.5.

Note that, for those asset owners (e.g. building owner) whose energy consumptions contain both elastic and inelastic parts, their electricity bills will be made up by two parts, as given in Eq.(3-1):

$$\boldsymbol{Electricity\ bill = \beta_E \times Q_{ine} - \beta_{GS} \times f(\beta_{GS}) - s_{GS}} \tag{3-1}$$

For the ease of discussion, battery storages and fixed loads are selected as representative elastic and inelastic customers to be modelled and studied in the following discussion.

3.1.4 Decision makers in the market

The dashed line in Figure 3-1 frames two types of decision makers in the retail market: RMO and elastic customers. Table 3-2 analyzes their decision makings from three different aspects, including decision variables, objective and constraints.

Table 3-2. Decision makers in the retail market

Decision maker	*Elastic customers*	*RMO*
Decision variables	Equipment operating schedule	Retail grid service price
Objective	Maximize its own profit/payoff from the day-ahead retail market	Minimize the total system cost for serving the inelastic customers
Constraints	Equipment operating constraints	Network physical constraints

RMO and elastic customers make their decisions in sequence, following a leader-followers game. RMO acts first as the leader. It determines the retail grid service price to minimize the total system generating cost for serving the inelastic customers. After the retail grid service price is announced, elastic customers then, act as followers, decide their operating plans for the next-day to maximize their profit/payoff from the day-ahead retail market. Although elastic customers take their actions afterwards, their decisions in turn affect effectiveness of the decision made by the RMO. It is because the profit made by elastic customers is part of the system cost. RMO determines the retail

grid service price based on an assumption about how elastic customers will react to the price signal. The closer the RMO is able to capture their actual price-responsive behaviors, the more effective the rendered price signal will be. Such a correlated decision-making process fits into a bi-level optimization framework, where the RMO solves an upper-level day-ahead resource planning problem, elastic customers solve lower-level profit/payoff maximization problems.

3.1.5 Market settlement

Figure 3-2 compares market settlement results obtained under three different market settings to illustrate how the proposed retail market design could benefit different types of market participants. The area of the gray rectangle on the left represents the total electricity bill paid by inelastic customers when they directly purchase their demands from the wholesale market. After they join the retail market, as shown by the case in the middle, the total system cost is expected to reduce resulting in an electricity bill saving benefit for inelastic customers. In the meantime, a certain grid service income is expected to be generated that benefits elastic customers. However, due to a considerable degradation effect of elastic assets, especially for battery storages, the discounted revenue earned by elastic customers could be very limited. Therefore, a grid service subsidy is introduced to offer a supplementary incentive to elastic customers, in addition to the grid service income. Note that the grid service subsidy is basically a part of total system cost-saving being transferred from the electricity bill saving to the gross revenue. A subsidy rate δ_{sub} can be applied by the RMO to define the proportion and control the trade-off between benefits shared to inelastic and elastic customers.

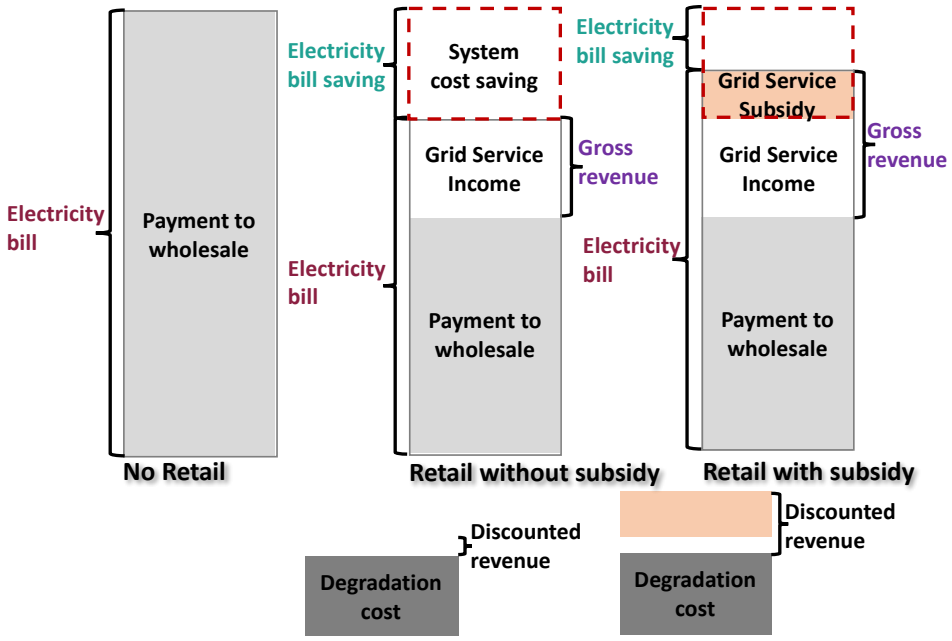


Figure 3-2. Illustration of market settlements under three different market settings

3.2 Test-bed system development

3.2.1 Framework of the test-bed system

For verifying performance of the proposed pricing mechanism, a test-bed system is developed in the study which mimics physical and financial operations of the retail market. Development of the test-bed system contains three parts:

1. Modeling of a three-phase IEEE-34 bus system
2. Development of three functional agents: Elastic customer agent, Inelastic customer agent and RMO agent.
3. Set-up of the communication across agents and the modeled system.

Figure 3-3 illustrates structure of the test-bed system and information exchanges among the components. According to Figure 3-3, it is shown that development of the test-bed system relies on both Python and OpenDSS programming. While the OpenDSS software is applied to model the physical behaviors of a three-phase IEEE-34 bus system and exporting system parameters, Python helps fulfill the forecasting and decision-making features of three functional agents. Arrows in Figure 3-3 indicate data flows within the test-bed system.

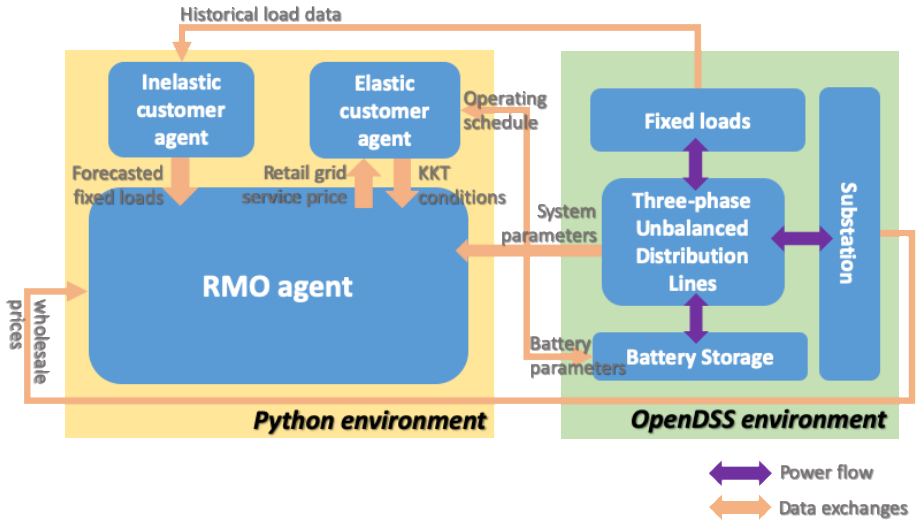


Figure 3-3. Framework of the test-bed platform

To begin with, each inelastic customer agent is associated with a fixed load (selected as the representative inelastic customer in the study) at different locations. Its interactions with the RMO agent and OpenDSS model are unidirectional. It receives the historical fixed demand data from the OpenDSS model and then sends the forecasted fixed load demands to the RMO model. Similarly, each elastic customer agent is associated with a battery storage (selected as the representative elastic customer in the study). Its interactions with the RMO agent and OpenDSS model are bidirectional. On the one hand, it collects equipment parameters and initial operating status from the OpenDSS model battery objects and returns the optimized equipment operating commands to

the OpenDSS model battery objects. On the other hand, it offers the information necessary for formulating the lower-level KKT conditions to the RMO agent and receives the optimized retail grid service price as a return. Finally, the OpenDSS model interacts with the RMO agent unidirectionally by sending the additionally information: wholesale market price and system parameters to the RMO agent to assist its decision making.

3.2.2 Modeling of the three-phase unbalanced distribution system

Among all the electric distribution system simulation tools (e.g. OpenDSS [116], GridLAB-D [117], PowerWorld [118] and CYMDIST [119]), OpenDSS is selected in this study for fulfilling modeling of the test-bed three-phase unbalanced distribution system, due to its free access, easy-to-use component object model (COM) interface and well-managed and detailed documentation. It is a powerful software developed by Electric Power Research Institute (EPRI), primarily for conducting electric distribution system analysis to meet the current and future technology challenges. OpenDSS can be used in many areas ranging from basic topics such as power flow and short-circuit analyses to advanced topics such as quasi-static time-series (QSTS), harmonic, dynamic (electro-mechanical) and probabilistic/scenario analyses. It also supports the modeling of various DERs: PV systems, energy storages, wind systems, demand response, microgrids and smart inverters, and is therefore ideal for conducting researches regarding smart grid, DERs integrations and grid modernization that serves to meet future needs. Moreover, OpenDSS can be implemented in both stand-alone or COM Dynamic Link Library (DLL) modes. The COM interface allows deployment of customized control strategies through interaction with user-defined scripts written in other programming languages, which brings more flexibility to the distribution system studies. This study mainly applies its modeling capability and built-in AC power flow solver to obtain system parameters and validate the power flows under different control strategies. A python-based OpenDSS communicator will be developed to bridge communication between the OpenDSS model and external python-based functional agents (to be discussed in the following sections).

Figure 3-4 shows critical components of the OpenDSS model as well as its interfaces with the OpenDSS communicator. Network infrastructures, nodal loads and batteries are three major parts making up the distribution system model. *"Master.dss"* is the main file that will be invoked by the compile command "Run" to solve the power flow. It defines the network infrastructures, including distribution lines, status of the line switches, capacitors, transformers and possibly regulators. All the other accessary *".dss"* file are appended to the master file through the "redirect" command to build the completed system. *"Linecode.dss"* lists all the available "line codes" that specify impedance characteristics for lines and cables, including number of phases, series resistance matrix (rmatrix), series reactance matrix (xmatrix) and shunt nodal capacitance matrix (cmatrix). Line objects presented in *"Master.dss"* can be sufficiently determined by its line code and length. *"Load.dss"* and *"Battery.dss"* files construct and configure the nodal loads and batteries in the system: including locations, phase numbers, connection approaches (e.g. Wye or Delta connection), types (e.g. constant P-Q node) and rated powers of the load/battery elements. Note that battery is simulated as the generator object in the study due to its flexibility to be specifically controlled by external scripts. *"DailyLoadShape.dss"* and *"DailyBatteryShape.dss"* files define shapes of the consumption or generation curves associated with each nodal load or battery objects, which are provided in hourly resolution and 24-h length. On the bottom of Figure 3 are some mostly applied

functions inside OpenDSS communicator. By calling the *GetYmatrix* function, system admittance matrix (Y matrix) can be returned from the OpenDSS model. It is worthy to mention that Y matrix has to be obtained when only the network impedances are considered, that is to say *GetYmatrix* has to be called when load and battery objects are deactivated in the OpenDSS model. *GetVolt* and *GetSinj* can be used to obtain the real-time/24-h voltage phasors and total power injections values at all the nodes when the master file is executed in a snapshot/daily mode. *AddBattery* function can be used to connect batteries at different locations to satisfy the requirement of flexible case study setting. Finally, *ConfigBattery* functions can be applied to employ optimized battery control strategies yielded from the elastic customer agents.

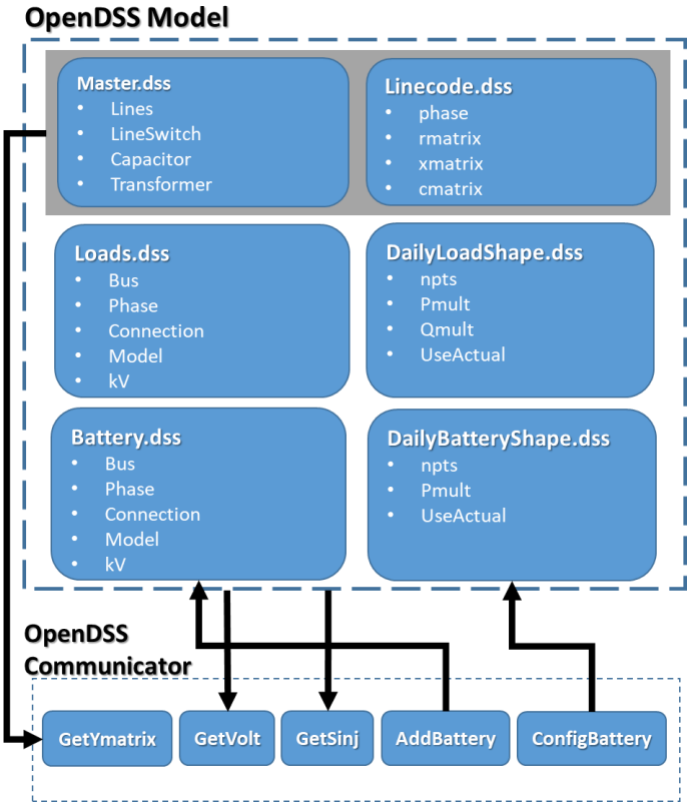


Figure 3-4. Structure of the OpenDSS model and its interactions with the OpenDSS Communicator

3.2.3 Development of three functional agents

All three functional agents are constructed as Python Class. It is a powerful compound data structure and the building block of object-oriented programming. It simplifies the creation of instances owing same attributes and functionalities. owns great properties of attribute reference, instantiation and inheritance. The Python agents developed in this study are made up by two major elements: instance attribute and instance method.

- **Instance attribute:** is a variable associating with a certain object that defines the property or status of the object. Every instance created from the same class owns the same attributes but with unique values.

- **Instance method:** is a Python function used to fulfill a functionality of an object. Every time when the instance method is called, it updates the value and status of the instance attributes following a predefined logic.

Figure 3-5 summarizes attributes and methods inside Python classes created for elastic customer, inelastic customer and RMO agents, with the instance methods responsible for the cross-agent communication highlighted in orange.

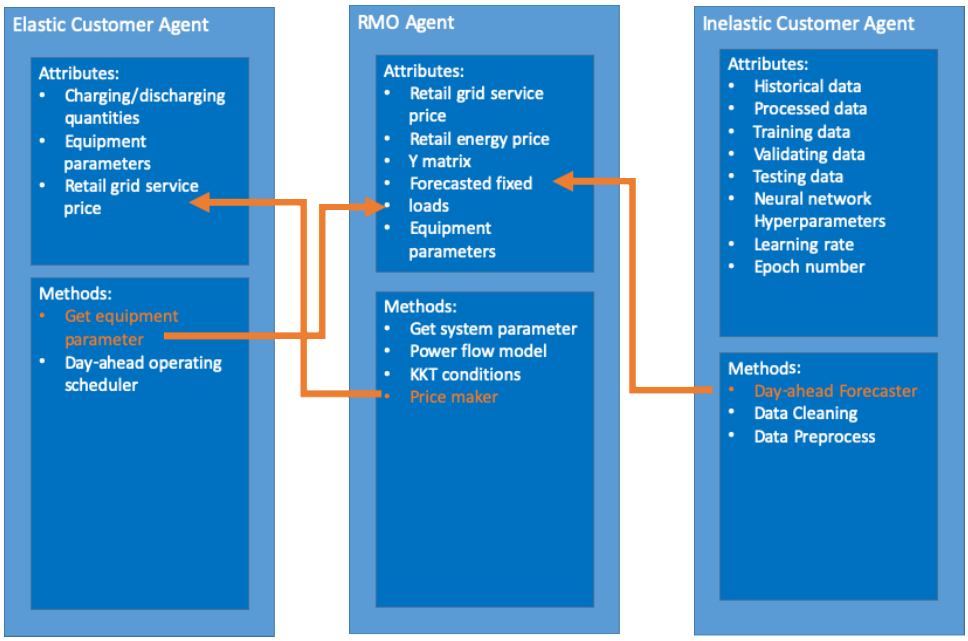


Figure 3-5. Python classes of three functional agents

Among all instance methods, the day-ahead operating scheduler function and retail price maker inside the elastic customer agent and RMO agent are developed using Pyomo, an open-source algebraic modeling language. It is selected in this study due to its user-friendly syntax, modeling flexibility and compatibility with Python. A Pyomo model consists of a collection of modeling components necessary for defining an optimization problem: index sets, parameters, variables, objective, and constraints. Also, it supports the solver interface with commonly applied commercial and free optimization solvers, including GUROBI, GLPK and CPLEX.

4. DEEP LEARNING-BASED LOAD FORECASTING MODELS[1]

All day-ahead load forecasting models compared in this study fall into the time-series category. The fact that time-series methods do not require additional time indexing parameters differentiates them from other supervised learning methods, e.g., MLR, SVR, and ANN, when handling time-series data. As a result, time-series methods are able to detect the time dependency (including hour-of-day and day-of-week) inherently embedded in the input data and avoid potential issues brought about by inappropriate time-index labeling. Candidate models discussed in this paper are the most representative time-series methods from both DL and non-DL scopes: the SARIMAX, the gated RNN and the gated CNN models.

4.1 Problem formulation

To ensure a fair comparison, the three investigated models are formulated into the same supervised learning framework. Where there is an input matrix X that integrates the information of historical load profiles (x^L) and outdoor temperature profiles (x^O), and also an output vector y refers to the predicted load profile with configurable prediction horizon. In this study, only the outdoor temperature is selected as the weather relevant feature as it shows a dominant influence towards building load in comparison with other weather relevant variables (i.e., air pressure, humidity and wind speed). Detailed clarification for such a selection is provided in Section 3.2. The day-ahead building-level load forecast is basically a multi-step forecasting problem. It requires the model to predict the next 24-hour load profile at midnight (12 AM) of each day given the historical load behaviors and prior knowledge about the next day's weather forecast.

The day-ahead building-level load forecasting can be performed in two different ways: recursive and direct.

- For the recursive multi-step, a one-hour prediction model is developed and implemented recursively 24 times to yield the day-ahead load forecast. The predicted value from the previous time step is fed as one of the inputs to the prediction model of the subsequent step.
- For the direct multi-step, on the other hand, a 24-hour prediction model is developed to generate the day-ahead load forecast at once.

To fully test the functionality of different time-series models, both the one-hour prediction models and the 24-hour prediction models are developed and compared.

4.2 Seasonal ARIMAX model

Seasonal ARIMAX (SARIMAX) identifies the time-series patterns inside the series while capturing the linear covariance between target variable and exogenous variables. Such time-series

[1] The work presented in this section has been published in the following journal paper: M. Cai, M. Pipattanasomporn, and S. Rahman, "Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques," *Appl. Energy*, pp. 1078–1088, Feb. 2019 [130].

patterns and covariances are summarized in Table *4-1*, with regard to building-level load profiles. The standard seasonal ARIMAX model follows the notation of ARIMAX$(p,d,q)\times(P,D,Q)S$, where p=non-seasonal auto-regressive (AR) order, d=non-seasonal differencing, q=non-seasonal moving average (MA) order, P=seasonal AR order, D=seasonal differencing, Q=seasonal MA order, S=time span of the repeating seasonal pattern. Let x_t^L denotes the load value at time t, then the ARIMAX model can be mathematically expressed in Eq. (4-1) and Eq. (4-2) with the backshift operator B. Where z_t and x_t^O represent the white noise and the exogenous outdoor temperature covariate at time t, respectively. $\nabla_s x_t^L = x_t^L - x_{t-S}^L$ and $\nabla x_t^L = x_t^L - x_{t-1}^L$ represent differencing operations.

$$\Phi(B^S)\phi(B)\nabla_S^D\nabla^d x_t^L = \beta x_t^O + \Theta(B^S)\theta(B)z_t) \tag{4-1}$$

Where:

$$\begin{aligned}\Phi(B^S) &= 1-\Phi_1 B^S-\cdots-\Phi_P B^{PS}\\ \phi(B) &= 1-\phi_1 B-\cdots-\phi_p B^p\\ \Theta(B^S) &= 1+\Theta_1 B^S+\cdots+\Theta_Q B^{QS}\\ \theta(B) &= 1+\theta_1 B+\cdots+\theta_q B^q\end{aligned} \tag{4-2}$$

Table 4-1. Time-series characteristics of load profiles

Category	***Name***	***Definition***
Patterns	Trend	Long-term trending behaviors describing how the overall power consumption increases or decreases
	Periodicity	Periodical patterns of load profiles that mainly result from the seasonal energy usage preference and daily routine of occupants
	Temporal dependence	Correlation between the lagged load values and future load values
	Uncertainty	Randomness resided in load profiles due to non-routine behaviors of occupants
Covariance	Weather covariance	A part of load pertinent to space cooling/heating needs, which is sensitive to the outdoor temperature variance

4.3 Gated RNN models

RNN is a class of neural networks that are able to recurrently process sequential inputs. Such a capability is enabled by the internal time loops at each hidden layer unit, where the output of the unit at time step t is taken as the input for the next step t+1. Long short-term memory units (LSTMs) have been proposed for addressing the gradient vanishing problem faced by vanilla RNN hidden units [120]. The intuition behind this approach is to maintain the memory property of the vanilla RNN units while being able to filter out redundant or misleading information through a long short-term gating mechanism.

Apart from the original hidden unit output (h_t) that is utilized to restore short-term memory immediately passed from previous time step, an internal memory cell (c_t) is introduced for restoring the long-term memory. Unlike the vanilla RNN unit, which has only one neuron, there are four neurons in the LSTM unit. One neuron works as a tangent function providing the candidate memory cell ($\tilde{c_t}$), computed using Eq. (4-3); the other three neurons work as sigmoid functions, controlling the flow of information, i.e., forget gate (f_t), input gate (i_t), and output gate (o_t), as given in Eq. (4-4). Based on the outputs of three gating functions, the internal memory cell (c_t) and the hidden layer output (h_t) subsequently update themselves according to Eq. (4-5). Where x_t is the input vector at each time step; W_C, W_f, W_i, W_o and b_C, b_f, b_i, b_o represent the weights and biases of the candidate neuron, the forget gate, the input gate, and the output gate.

$$\tilde{c_t} = tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{4-3}$$

$$\begin{aligned} f_t &= \sigma\big(W_f \cdot [h_{t-1}, x_t] + b_f\big) \\ i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\ o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \end{aligned} \tag{4-4}$$

$$\begin{aligned} c_t &= f_t * c_{t-1} + i_t * \tilde{c_t} \\ h_t &= o_t * tanh\,(c_t) \end{aligned} \tag{4-5}$$

Figure 4-1 illustrates the architecture of the gated RNN model after being unrolled along the time axis. Where m and n denote the lengths of lagged values and prediction horizon; $y_1, \dots, y_n$ denote forecasted load values at each time step. Note that weather prediction information is appended at the end of the input series (highlighted in yellow circles in Figure 4-1), which differentiates this study from other existing work.

Unlike the seasonal ARIMAX model which only works as a one-hour prediction model, one advantage of the DL-based models is their flexibility on the lengths of input and output vectors. Parameters m and n in Figure 4-1 can be set to 24 and 1 to obtain the one-hour gated RNN model (i.e., GRNN1). On the other hand, to obtain the 24-hour gated RNN model (i.e., GRNN24), both m and n can be set to 24.

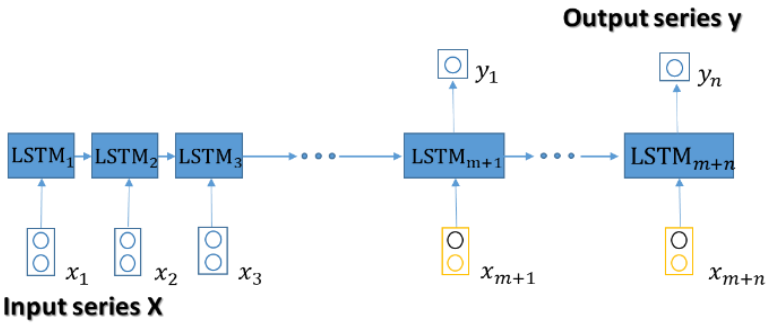


Figure 4-1. Gated RNN model for day-ahead load forecasts

4.4 Gated CNN models

CNN is made up of neurons that apply convolution computation to the inputs hierarchically. It is invented for image processing. The first time when CNN was used to solve sequential problems was in [121]. After introducing a novel gating mechanism, the developed gated CNN model is able to outperform RNN for the language sequential modeling task. Inspired by the autoregressive model and work done in [121], authors in [91] designed a double-channel gated CNN model, called Significance Offset CNN (SOCNN). This model is taken as the prototype for designing the gated CNN model in this study. Following modifications have been made to fit the context of the day-ahead building-level load forecasts:

1. Synchronous input series: The input variables are embedded in a synchronous manner instead of asynchronous manner.
2. Concatenated weather prediction: Weather prediction information is appended to the end of historical series (highlighted in yellow circles in Figure 4-2). It is a $2 \times n$ matrix with the first row filled with all zeros and the second row filled with predicted outdoor temperature values.

Figure 4-2 illustrates the structure of the gated CNN model designed for day-ahead building-level load forecasts. The formats of input variables (X) and the output variable (y) are identical to that of the gated RNN model. According to Figure 4-2, each input variable is processed simultaneously using two different convolutional operations, either operated together with the neighboring time stamps for detecting the temporal dependency or operated alone for identifying the weather correlation at each time stamp. What differentiates these two operations is the kernel size of the filters: They are Conv1D filters with size $2 \times$ k (marked by blue triangles in Figure 4-2) and Conv1D filters with size 2×1 (marked by blue thick arrows in Figure 4-2). On top of the convolutional layers are the linear gates following the gating function in Eq. (4-6), where $O_{Conv1D\_2\times k}$ and $O_{Conv1D\_2\times 1}$ represent the final outputs of the last Conv1D $2 \times$ k layer and the last Conv1D 2×1 layer.

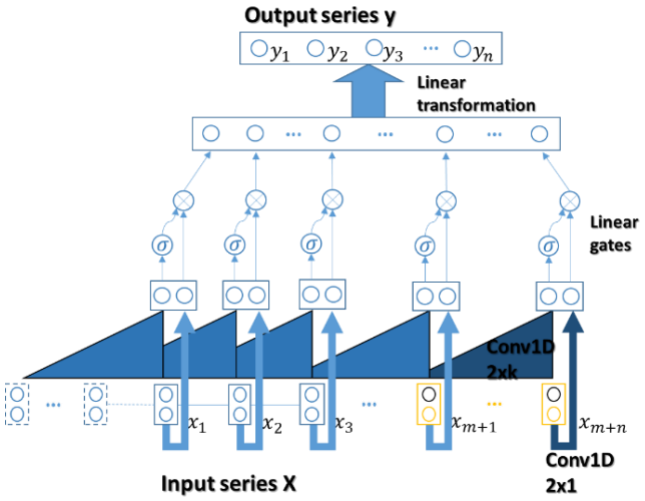


Figure 4-2. Gated CNN model for day-ahead load forecasts

$$O_G = \sigma(O_{Conv1D\_2\times k}) \otimes O_{Conv1D\_2\times 1} \tag{4-6}$$

Then the output from linear gates (O_G) is mapped into the final output series (y) through a fully connected neural network layer, see Eq. (4-7). Where $W_{(m+n)\times n}$ and b represent the weight and bias of the fully connected layer.

$$y = W_{(m+n)\times n} O_G + b \tag{4-7}$$

Note that a $k-1$ length padding (marked by dashed rectangular frames in Figure 4-2) has been applied to the input sequences at each layer in order to make sure that $O_{Conv1D\_2\times k}$ and $O_{Conv1D\_2\times 1}$ share the same dimension.

The whole framework works as a data-dependent autoregressive model, where the Conv1D 2×1 filters provide the regressor at a single time step, and the Conv1D $2 \times k$ filters generate the coefficients for each regressor. Unlike the fixed coefficient applied in the autoregressive model, the output of Conv1D $2 \times k$ filters varies as the input data change. Therefore, the gated CNN model is much more capable of capturing the non-linear temporal relationship exist in the data.

Similar to the gated RNN model, the one-hour gated CNN model (i.e., GCNN1) is obtained by setting $m = 24$ and $n = 1$; The 24-hour gated CNN model (i.e., GCNN24) is obtained by setting both m and $n = 24$.

4.5 Experimental setting

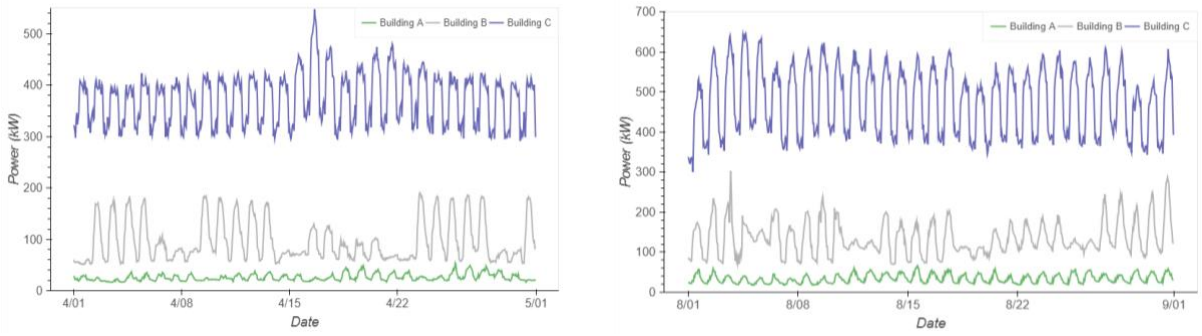
Performances of the five deep learning models (i.e., SARIMAX, GRNN1, GCNN1, GRNN24 and GCNN24) proposed in Section 4.2 - 4.4 are validated in this section. In order to test their ability for handling high uncertain load profiles, building-level load date is selected for testing. Variances on locations and consumption levels of these buildings help substantiate the models'

generalizability. For the purpose of retail market pricing design, nodal node data will be collected in the future to train the validated best performed model, which will then serve to providing the forecasted nodal loads.

4.5.1 Datasets

Three selected building load datasets used in this study were collected in one-hour intervals for a one-year period with 5% of missing data, consisting of five time series: hourly electrical load (kW), outdoor temperature (°F), air pressure (in), humidity (%) and wind speed (mph). Note that all buildings under study are gas-heated. Since hourly natural gas consumption data are not available, this paper focuses on predicting the electricity demand.

Building A, located in Alexandria, VA, is an academic building with the area of around 30,000 square feet. Buildings B, located in Shirley, NY, is a primary/secondary school with the area of around 80,000 square feet. And, Building C, located in Uxbridge, MA, is a grocery store with the area of around 55,000 square feet. Building A's dataset was collected through the BEMOSS project [122]. Buildings B and C' datasets are provided by the public EnerNOC Commercial building dataset [123]. Weather data were fetched from the weather underground website [124]. Due to different building load characteristics when HVAC is set to different modes (e.g., COOL or OFF), datasets of each building were divided into: the COOL period (i.e., HVAC mode is set to COOL, May-October) and the NON-COOL period (i.e., HVAC mode is set to OFF, November-April). Peak electricity demands of buildings A, B and C are approximately 40kW, 180kW and 450kW during the NON-COOL period, and 60kW, 380kW and 650kW during the COOL period. Figure 4-3 depicts load profiles of these three buildings in April (NON-COOL) and August (COOL).



(a) April *(b) August*

Figure 4-3. Load profiles of three buildings

According to Figure 4-3, daily load patterns of each building are relatively constant in April (due to weak weather covariance), whereas vary greatly in August (due to strong weather covariance). There is a trend of decreased share of uncertainty when the building load-scale increases. Despite the relatively constant daily load patterns exhibited by buildings B and C, there still exist some unpredictable irregular patterns.

4.5.2 Weather relevant feature selection

In order to select the most relevant weather indicators as the model features, a feature selection work is conducted prior to model construction. Here the weather variables accessible from weather underground website are considered as candidate features. These are outdoor temperature (x^O), air pressure (x^A), humidity (x^H) and wind speed (x^W). Pearson correlation coefficient is applied as the score function due to its widely recognized ability of measuring the correlation between two continuous variables. It can be computed based on Eq. (4-8), where cov represents the covariance, $\sigma_{x^{weather}}$ and σ_{x^L} are the standard deviations of any weather variable ($x^{weather}$) and building load series (x^L) respectively. Table *4-2* summarizes the Pearson correlation coefficient statistics between each weather variables and building load for buildings A-C. According to Table *4-2*, a strong positive correlation is found between outdoor temperature and building load across all buildings, whereas the correlation between the other weather variables (i.e., air pressure, humidity and wind speed) and building load are either zero or insignificant (with absolute values of less than 0.3). Adding these weakly correlated variables into the feature set will tend to overfit the model. Therefore, in this study, only outdoor temperature (x^O) is selected as the weather relevant feature.

$$\rho_{x^{weather},x^L} = \frac{cov(x^{weather}, x^L)}{\sigma_{x^{weather}}\sigma_{x^L}}, \quad x^{weather} \in \{x^O, x^A, x^H, x^W\}O_G \tag{4-8}$$

Table 4-2. Pearson correlation coefficients between different weather relevant variables and building load

	Building A (40-60kW$_p$)	***Building B (180-380kW$_p$)***	***Building C (450-650kW$_p$)***
Outdoor Temperature	0.74	0.51	0.76
Air Pressure	0	0	0
Humidity	-0.26	-0.07	-0.10
Wind speed	-0.02	-0.04	0.02

4.5.3 Data preprocessing

The following data-preprocessing procedures have been implemented:

Data Cleaning: In order to eliminate the influence of missing data, list-wise deletion was conducted. Raw datasets were then divided into multiple sections bounded by the missing points so as to maintain the continuity of time. Then, weekend data were removed, since the irregular load pattern during weekends has proven to have significant influence on the prediction accuracy on weekdays.

Data Segmentation: After filtering out the missing data and weekend profiles, cleaned datasets for each building were then segmented into: training dataset (90%), validation dataset (5%) and testing dataset (5%).

Time-series to supervised learning dataset: All training, validation and testing datasets were converted into the supervised learning format, where time-series sequences became input-output pairs through the process of sliding windows.

Normalization: Finally, to stabilize the learning process, input variables in each training dataset, together with corresponding validation and testing datasets, were carefully normalized. Normalization helps prevent dramatic changes on the gradient, so as to smoothen the convergence.

4.5.4 Model validation

After all datasets were properly preprocessed, model parameters (i.e., weights and bias in DL-based models) and hyper-parameters (i.e., layers and number of neurons in DL-based models) were tuned using training and validation datasets. Once the optimal model parameters (different for different buildings) and hyper-parameters (same for each building) were obtained, testing datasets of each building were fed into the optimized models for comparing the performance. The Train-Validation-Test split was applied here to verify the capability of each model to be generalized to unseen data.

4.5.5 Implementation

All experiments were carried out in the Python compiling environment using an Intel Core i7-4770 CPU machine. Implementation of the SARIMAX model relied on the Python package: StatsModels. All DL-based models were developed using the TensorFlow [125].

Optimized hyper-parameters used in each model are summarized in Table *4-3*. Where L, NN, r, bs, and ks represent the layer structure, number of neurons in each layer, learning rate, batch size, and kernel size (only available in CNN models), respectively. Adam optimizer was applied to all of the DL-based models.

Table 4-3. Load forecasting model configuration

non-DL model							
Model	p	d	q	P	D	Q	S
SARIMAX	24	1	0	0	1	0	24
DL models							
Model	L		NN	r	bs	ks	
$GRNN\_1$	2		3	0.005	50	-	
$GCNN\_1$	3 conv + 3 conv		8,5,1 + 8,5,1	0.005	50	6,3,3 + 1,1,1	
$GRNN\_24$	3		32	0.005	50	-	
GCNN_24	3 conv + 3 conv		10,8,1 + 10,8,1	0.005	50	6,3,3 + 1,1,1	

4.6 Performance analyses

4.6.1 Accuracy of one-hour prediction models

Accuracy of the one-hour prediction models is presented in Table *4-4* and Table *4-5*, summarizing mean absolute percentage error (MAPE) and coefficient of variance (CV) of the single-step predictions. MAPE and CV were calculated using Eq. (4-9)-(4-10), where Pre_i and Act_i represent the predicted and actual loads. N is the number of observations in the testing dataset. $\overline{Act}$ is the mean of actual values.

$$MAPE = \frac{\sum_{i=1}^{N} \left| \frac{Pre_i - Act_i}{Act_i} \right|}{N} \times 100 \tag{4-9}$$

$$CV = \frac{\sqrt{\frac{\sum_{i=1}^{N} (Pre_i - Act_i)^2}{N-1}}}{\overline{Act}} \times 100 \tag{4-10}$$

Table 4-4. MAPE (%) of one-hour prediction models under different testing cases

	Building A (40-60kW$_p$)		***Building B (180-380kW$_p$)***		***Building C (450-650kW$_p$)***	
	NON-COOL	*COOL*	*NON-COOL*	*COOL*	*NON-COOL*	*COOL*
SARIMAX	13.54	16.13	10.57	8.04	2.95	3.85
GRNN1	8.67	8.99	7.91	5.95	2.36	2.41
GCNN1	8.02	8.53	5.73	5.85	2.23	2.38

Table 4-5. CV (%) of one-hour prediction models under different testing cases

	Building A (40-60kW$_p$)		***Building B (180-380kW$_p$)***		***Building C (450-650kW$_p$)***	
	NON-COOL	*COOL*	*NON-COOL*	*COOL*	*NON-COOL*	*COOL*
SARIMAX	18.71	22.12	13.40	11.14	3.39	5.28
GRNN1	12.20	12.01	8.63	8.10	2.99	3.15
GCNN1	10.97	10.86	6.85	7.68	2.66	3.03

The best performance among all three models is observed in Building C — the building with the largest peak load (450kW/650kW during NON-COOL/COOL period). This building has MAPE values of less than 2.5% and 4.0% for one-hour DL-based models and the SARIMAX model. As the building load-scale become smaller, performances of all three models keep going down (see results for Buildings A and B). It is as expected because larger buildings usually have more occupants. As there are more people in a building, it is more likely that the uncertainty of their aggregated behavior will be averaged, leading to a more regular and predictable building load pattern.

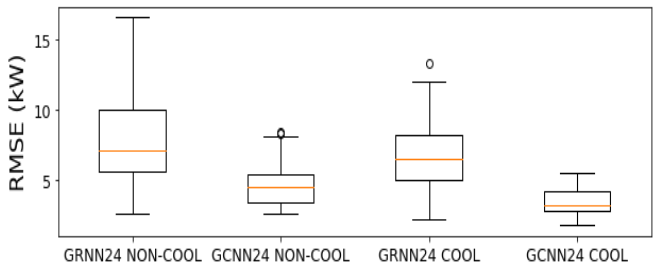
For all three buildings being tested, the SARIMAX model performs significantly poorer than the DL-based models, and its model performance deteriorates significantly as the building peak load goes down. The performance of one-hour DL-based models is not heavily degraded compared to that of the SARIMAX for smaller buildings, as MAPE values remain at around 9% (compared to 17% of the SARIMAX) when dealing with the building with peak load of less than 60kW. This indicates a much better generalizability of the one-hour DL-based models as compared to the SARIMAX model. Within the DL scope, it is also found that the GCNN1 model performs slightly better than the GRNN1 model and contributes the best performance for all cases. It brings down the MAPE errors of the SARIMAX model by 37.0% and 37.5% for NON-COOL and COOL cases on average.

4.6.2 Accuracy of 24-hour prediction models

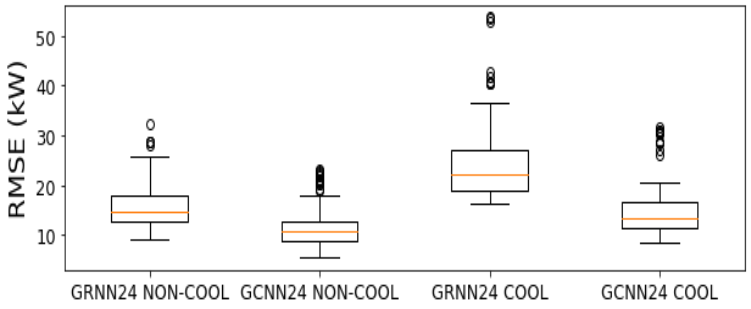
Accuracy of the 24-hour prediction models are analyzed in Figure 4-4, where the boxplots of RMSE tested on buildings A, B and C are shown in subplots Figure 4-4 (a)-(c). Since the SARIMAX model can only be formulated as the one-hour model, it is not considered in this section. Only two 24-hour prediction models (i.e., GRNN24 and GCNN24) are compared. Output of the 24-hour model is a vector instead of a scalar as in the case of the one-hour model. Therefore, RMSE shown in Eq. (4-11) is applied to evaluate their prediction accuracy. Where n=24 represents length of the output vector.

$$RMSE_i = \sqrt{\frac{\sum_{j=1}^{n}(Pre_j - Act_j)^2}{n}} \tag{4-11}$$

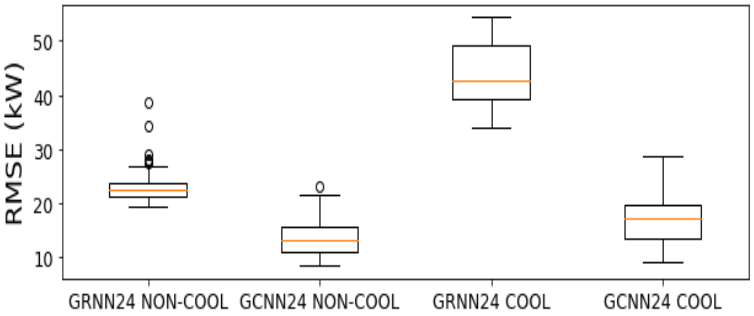
According to Figure 4-4, the accuracy of the GRNN24 model is unsatisfactory, as the mean RMSE values of the GRNN24 model for all COOL cases are 6.6kW, 25.0kW and 44.0kW of buildings A, B and C, all being above 6.6% of the building peaks. With lower RMSE means and variances, the GCNN24 performs much better compared to the GRNN24. Averaged RMSE values of the GCNN24 model are 4.0kW, 13.4kW and 15.2kW for buildings A, B and C, which are around 6.7%, 3.5% and 2.3% of the building peaks, respectively.



(a) Building A-40/60kWp



(b) Building B-180/380kWp

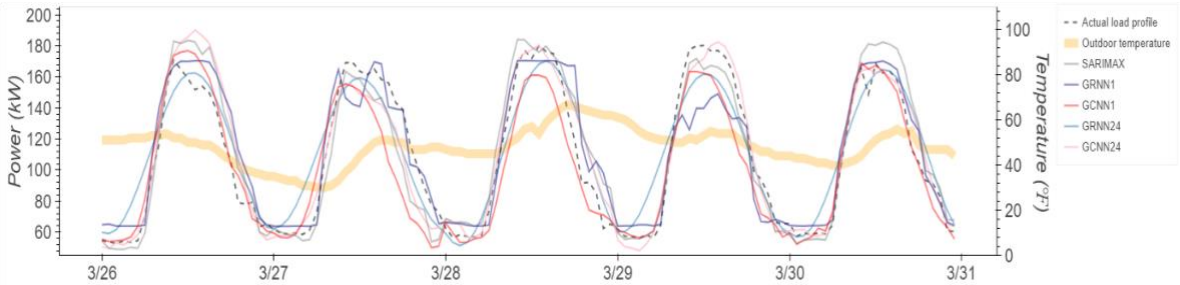


(c) Building C-450/650kWp

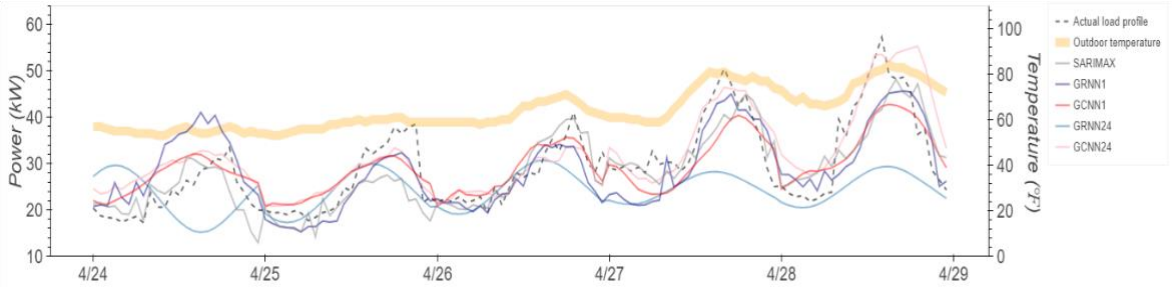
Figure 4-4. Comparison of the RMSE boxplots of 24-hour DL-based models across all scenarios

4.6.3 Performance comparison of day-ahead load forecasts

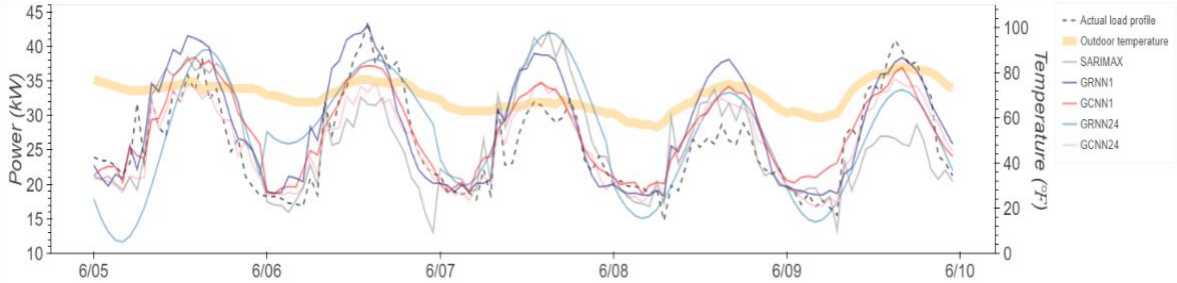
In this section, all developed one-hour/24-hour prediction models are compared when performing the day-ahead building-level load forecasts. Figure 4-5 plots the forecasting results of different models (including SARIMAX, GRNN1, GCNN1, GRNN24 and GCNN24) using one NON-COOL week and one COOL week of buildings A, B and C (only for weekdays). Averaged RMSE values for all five testing weekdays are summarized in Table *4-6*. For each case, the experiment was conducted to perform day-ahead forecasts at midnight for five consecutive weekdays.



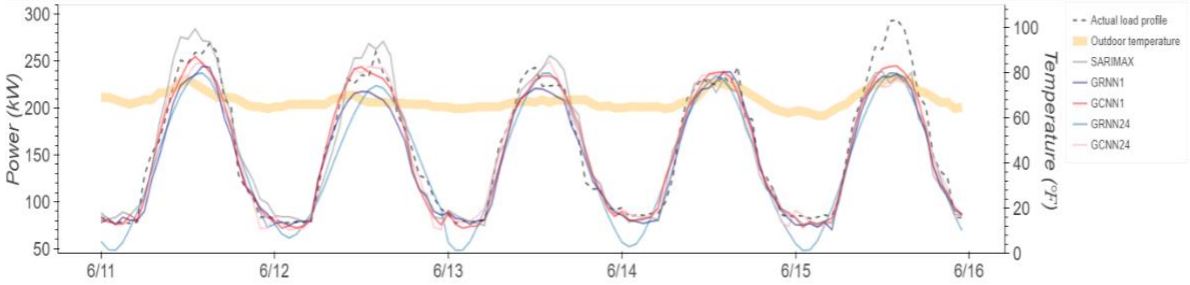
(a) Building A NON-COOL week



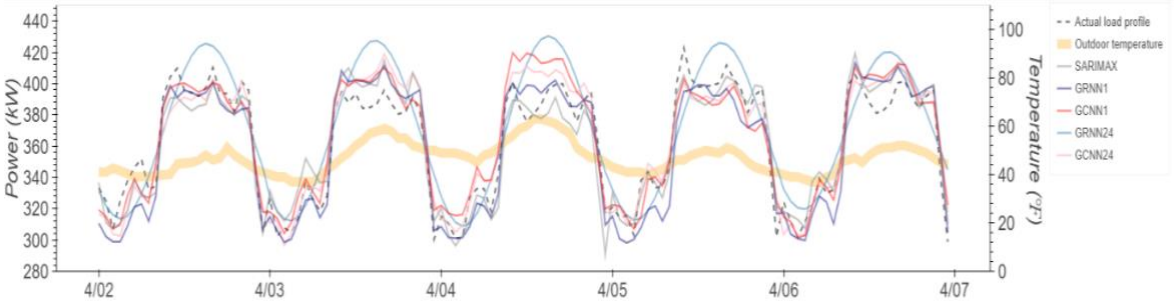
(b) Building A COOL week



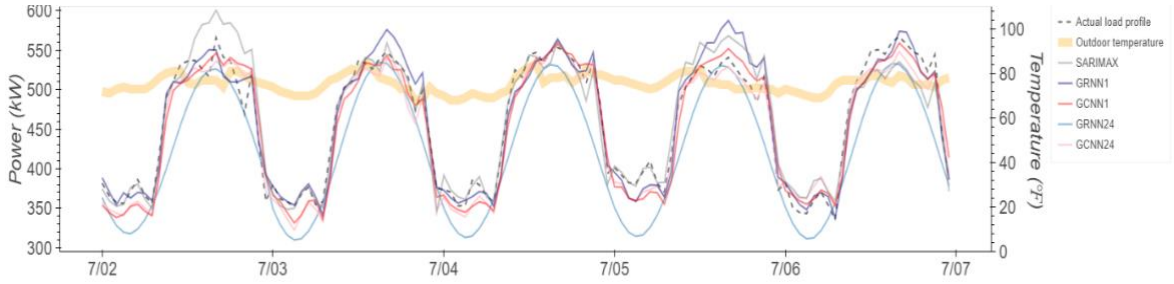
(c) Building B NON-COOL week



(d) Building B COOL week



(e) Building C NON-COOL week



(f) Building C COOL week

Figure 4-5. Comparing day-ahead load forecasting results of different models (SARIMAX, GRNN1, GCNN1, GRNN24 and GCNN24)

According to Figure 4-5, despite the poorer one-hour prediction accuracy of the SARMIAX model compared to that of the one-hour DL-based models, it generates better multi-step forecasting results for all NON-COOL cases than the one-hour DL-based models. However, GCNN24 outperforms it for NON-COOL cases of Buildings A and B. Also, its performance on COOL cases expose its sensitivity to both outdoor temperature covariance and load uncertainty. For outdoor temperature covariance, it is observed that when there is a strong weather covariance (COOL period), the day-ahead load forecasts generated by the SARIMAX model heavily deviates from the actual values. Such an influence, on the other hand, can be successfully handled by the DL-based models. For load uncertainty, the performance gap between the SARIMAX model and DL-based models for both NON-COOL and COOL cases are found to be the largest for Building A, which has the lowest peak load and highest share of uncertainty. This is intuitively understandable, as the SARIMAX model is built on an assumption of linearity, whereas the real temporal relationship and covariance are mostly non-linear. Also, the substantial amount of uncertainty contained in the time-series building load data may heavily degrade the performance of SARIMAX, since regression-based methods assume that both input and output variables follow the Gaussian distribution.

Table 4-6. Comparison of the averaged RMSE (kW) for different testing datasets

Building &Case	***SARIMAX***	***GRNN1***	***GCNN1***	***GRNN24***	***GCNN24***
A NON-COOL	5.2	5.5	5.2	9.0	**5.1**
COOL	5.7	4.7	3.8	5.9	**3.4**
B NON-COOL	14.3	15.3	15.6	15.8	**12.6**
COOL	17.5	17.1	**14.3**	23.1	15.9
C NON-COOL	**11.4**	12.8	13.0	21.4	12.5
COOL	21.2	17.9	19.0	45.0	**17.3**

*Best performances of each case are highlighted in bold

As for the DL models, despite the better one-hour prediction accuracy of the GCNN1 model compared to that of the GRNN1 model, indicated in Table *4-4* and Table *4-5*, the GCNN1 model not necessarily wins the day-ahead load forecasting tasks. Their performances are comparative and it is hard to say which one outperforms the other. As for the 24-hour models, the superior performance of the GCNN24 model over the GRNN24 model is observed as already proved in Figure 4-4. When comparing across all one-hour and 24-hour models, despite the high one-hour prediction accuracy of the one-hour models, their multi-step forecasting results are proven to be slightly worse than that of the GCNN24. This is due to the accumulated errors through recursive operations.

Overall, the GCNN24 model demonstrates the most promising ability of handling the day-ahead building-level load forecasts. The predicted curves provided by it closely follow the load shapes of different weekdays and capture the peak load change against various weather conditions. According to Table *4-6*, GCNN24 outperforms other models for 4 out of 6 cases. It is found that for the GCNN24 model, the RMSE values for the testing cases with peak load of 380kW or above are approximately 4.2% or less of the building peak. However, when the peak load decreases to 180kW or less, the RMSE values are increased to be above 5.7% of the building peak.

In order to analyze how the GCNN24 model provides improvement in day-ahead load forecasting as compared to the other models, an hour-of-day indexed error analysis across SARIMAX, GCNN1 and GCNN24 is conducted, taking Building A, COOL case as an example. Averaged MAPE values at each hour of the day of different models are plotted in Figure 4-6. As illustrated in Figure 4-6, although the hourly prediction errors of the three models are similar in the first hour, these values evolve quite differently across the day. A significant increase in error happens starting from hour 10 when the SARIMAX model is applied. This is mainly due its low single-step prediction accuracy as well as the error accumulation during the recursive operations. By replacing the SARIMAX model with the GCNN1 model, single-step prediction accuracy has been largely improved. However, error accumulation problem still exists as an abrupt error peak is observed during hours 6-8. Finally, by converting the one-hour CNN model into the 24-hour model, the model performance is further improved as the histogram of GCNN24 appears to be the most flattened across the day. Instead of fitting the model hour by hour, GCNN24 is able to learn the daily load shape as a whole so as to avoid abrupt deviation from the regular daily load pattern in its prediction. It largely prevents the negative impact of the error accumulation and contributes to a better multi-step performance.

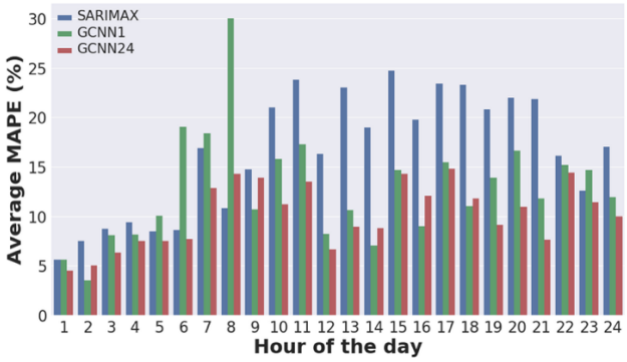


Figure 4-6. Hour-of-day indexed error analysis for SARIMAX, GCNN1 and GCNN24

4.6.4 Computational efficiency

Table *4-7* summarizes the computation time of each model in seconds, based on the experiments conducted using the COOL dataset of Building C. Epoch numbers of each DL-based model were determined based on the convergence time shown on the learning curves (Figure 4-7). According to Table *4-7*, the GCNN24 model demonstrates the best computational efficiency. Although the SARIMAX shows higher computation efficiency compared to most DL-based models, it is less efficient than the GCNN24 model. Compared to the GRNN models, GCNN models have speeded up the computation by 25-42%.

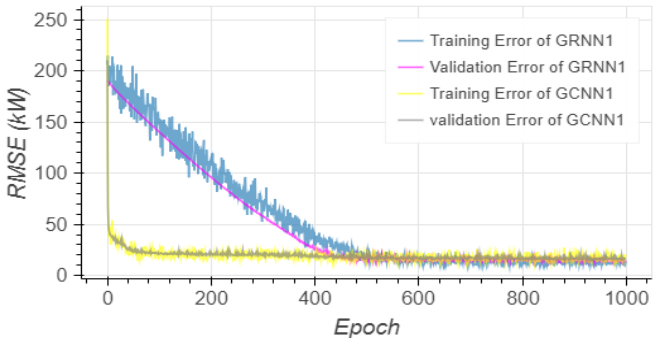
Figure 4-7 (a) and (b) specifically compare the learning curves among different DL-based models, using the COOL dataset of building C.

Based on Figure 4-7, the following observations are made:

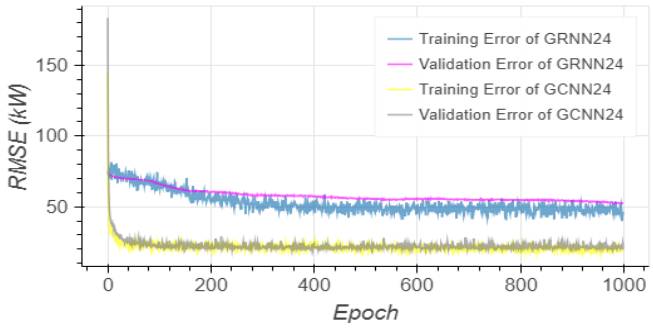
1. With regard to one-hour prediction models, it takes the GCNN1 model less than 400 epochs to converge, while around 850 epochs are required for the GRNN1 model. However, the validation error curve of GRNN1 goes down more smoothly as compared to that of GCNN1.
2. With regard to 24-hour prediction models, the GCNN24 model converges at around 240 epochs, but it takes the GRNN24 around 1000 epochs to converge. Also, the validation error curve of GRNN24 goes down more smoothly as compared to that of GCNN24.

Table 4-7. Computation time for each load forecasting model (in seconds)

One-hour models		***24-hour models***	
SARIMAX	87		
GRNN1 (850 epochs)	123	GRNN24 (1000 epochs)	139
GCNN1 (400 epochs)	92	GCNN24 (240 epochs)	80



(a) Learning curves of one-hour prediction models (GRNN1, GCNN1)



(b) Learning curves of 24-hour prediction models (GRNN24, GCNN24)

Figure 4-7. Learning curves comparison between GRNN and GCNN models

Overall, GCNN24 outperforms all other models, including its traditional time-series counterpart SARIMAX, in terms of the computational efficiency. Such an advantage attributes to its capability of paralleling the computation. Unlike the chain-like structure that the GRNN models apply for processing the time-series data, the GCNN models are able to parallel the sequence and process each section synchronously.

4.6.5 Robustness analysis

So far it has been proven that the GCNN24 and one-hour DL-based models own the top performances among the investigated models for day-ahead building-level load forecasts. However, the predicted weather profiles used in the testing datasets are assumed to be 100% accurate. An additional concern is raised as the accuracy of next-day weather prediction may not always be guaranteed. In order to examine the robustness of GCNN24, GRNN1 and GCNN1 for handling the day-ahead building-level load forecasts in presence of weather prediction error, a further robustness analysis was conducted.

Three sets of manipulated testing datasets were generated by introducing three different levels of noises into the day-ahead weather data. To be specific, for the five-day hourly weather profile (5*24=120 data points), 20%, 40% and 60% of the original data points were added with a Gaussian distributed white noise. The GCNN24, GRNN1 and GCNN1 models were re-run on the COOL dataset of Building C using three noise-introduced weather profiles. Table *4-8* presents the averaged RMSE values for five testing days of GCNN24, GRNN1 and GCNN1 models under different noise levels.

It is shown that with the introduction of up to 60% weather prediction white noise, the averaged RMSE of the GCNN24, GRNN1 and GCNN1 models slightly increase from 17.26 kW to 17.84 kW (by 3.36%), from 17.88 kW to 18.52 kW (by 3.58%) and from 18.99 kW to 19.47 kW (by 2.53%). Therefore, it is verified that all three models: GCNN24, GRNN1 and GCNN1 models are robust against the error in the next-day weather forecasts.

Table 4-8. Averaged RMSE (kW) for five testing days under different noise levels

Percentage of introduced noise	***GCNN24***	***GRNN1***	***GCNN1***
0%	17.26	17.88	18.99
20%	17.62	18.21	19.53
40%	17.55	18.38	19.23
60%	17.84	18.52	19.47

4.6.6 Conclusion

Two popular deep learning networks, RNN and CNN, have been utilized and constructed into the day-ahead building-level load forecasting framework. Their performance under both recursive multi-step and direct multi-step forecasting manners are analyzed. Experimental results indicate that most of the proposed deep learning-based models (except GRNN24) achieve promising results as compared to its traditional counterpart SARIMAX. From the aspect of prediction accuracy, the day-ahead multi-step forecasting errors of the proposed GCNN24 model decrease from 14.8 kW to 12.2 kW on average for cases with strong weather covariance, compared to those of the SARIMAX model. Such a decrement reaches 3.9 kW when the building load patterns are highly uncertain (i.e., for buildings with smaller peak electrical load). From the perspective of computational efficiency, when the GCNN24 model is applied, the operation time is speeded up by 8% compared to that of the SARIMAX model. From the aspect of generalizability, it is proven that the impact of building load uncertainty on the performances of GRNN1, GCNN1 and GCNN24 is less than that of the SARIMAX model.

Within the deep learning scope, the GCNN24 model outperforms all other deep learning-based models investigated in this study. It offers better accuracy compared to the GCNN1 model, and significant superiority to the GRNN models in both accuracy and computational efficiency. From the practical point of view, the multi-thread design of modern hardware well supports the models with parallelized structure. Therefore, it is expected that the GCNN24 model can play a dominant role in future day-ahead load forecasting work, especially when dealing with load forecasts for a large number of entities.

To sum up, the emergence of deep learning techniques provides us the opportunity to bring the performance of day-ahead load forecasts to an even higher level, especially for load dataset with high uncertainty. Compared to the conventional approaches, a well-design hierarchically-structured deep learning network may be more capable of capturing the data-dependent uncertainty and may even increase the computational efficiency for large-scale application. The GCNN24 model proposed in this study demonstrates competitive capabilities. It can be applied as a good start for the deep learning-based network investigation of day-ahead load forecasts.

5. DAY-AHEAD BATTERY ARBITRAGE MODEL

This section investigates how a single profit-oriented battery storage behaves in response to a dynamic price signal and what drives its decisions.

5.1 Modeling of single battery storage arbitrage behavior

As discussed in Section 3, it is assumed the battery storage is owned by a third-party battery company who seeks to maximize its profits in the retail market through battery arbitrage. Such a profit-oriented operation of a single battery storage can be modeled as an optimization problem as given below:

Max

$$\sum_{t=1}^{24} \beta_{GS}^{t} \left(P_{out}^{t} - P_{in}^{t}\right) - D_{kWh} R^{t} \tag{5-1}$$

Subject to:

$$SoC^{1} = SoC^{init} + \frac{P_{in}^{1}(1-l_c)}{C} - \frac{P_{out}^{1}}{(1-l_d)C} \tag{5-2}$$

$$SoC^{t} = SoC^{t-1} + \frac{P_{in}^{t}(1-l_c)}{C} - \frac{P_{out}^{t}}{(1-l_d)C} \quad t \in [2,24] \tag{5-3}$$

$$R^{t} = P_{in}^{t}(1-l_c) + \frac{P_{out}^{t}}{1-l_d} \quad \forall t \tag{5-4}$$

$$SoC^{min} \le SoC^{t} \le SoC^{max} \quad t \in [1,23] \tag{5-5}$$

$$SoC^{end} \le SoC^{24} \le SoC^{max} \tag{5-6}$$

$$R^{min} \le R^{t} \le R^{max} \quad \forall t \tag{5-7}$$

$$P_{in}^{t} \ge 0 \quad \forall t \tag{5-8}$$

$$P_{out}^{t} \ge 0 \quad \forall t \tag{5-9}$$

$$P_{in}^{t} P_{out}^{t} = 0 \quad \forall t \tag{5-10}$$

Where t indexes hours in a 24-h control horizon. β_{GS}^{t} provides the grid service price at time t. P_{in}^{t} (in kWh) and P_{out}^{t} (in kWh) denote control variables: active power purchased from and sold to the retail market at t time interval. Separation of these two variables is necessary since actual active power absorbed and extracted from the battery are calculated in two different ways as given by $P_{in}^{t}(1-l_c)$ and $\frac{P_{out}^{t}}{(1-l_d)}$, where l_c and l_d refer to charging and discharging loss rates of the battery, which are not equal in general cases. SoC^{t} and R^{t} refer to state-of-charge and dis-/charging rate at hour t. They are status variables sufficiently determined by initial system status and sequential control actions. Naming convention of remaining constant parameters are summarized in Table *5-1*:

Table 5-1. Definitions of parameters applied in the battery arbitrage problem

Parameter	**definition**	**Unit**
C	Capacity	kWh
l_c	Charging loss ratio	
l_d	Discharging loss ratio	
SoC^{min}	Minimum state of charge	
SoC^{max}	Maximum state of charge	
SoC^{end}	Minimum state of charge to be maintained at end of the day	
R^{min}	Minimum charging/discharging rate	kW
R^{max}	Maximum charging/discharging rate	kW
D_{kWh}	Degradation cost per kWh	\$/kWh

As stated in Eq. (5-1), the objective function is designed to maximize the net arbitrage income of the battery owner, which is calculated by subtracting the degradation cost from the grid service income. Eq. (5-2) and Eq. (5-3) define the dynamic of SoC, describing how the status variable SoC^t evolves over time. Eq. (5-4) calculates the dis-/charging rate at time t given the control action. Inequality constraints Eq. (5-5)-(5-7) limit the range of SoC^t and R^t. Eq. (5-8) and Eq. (5-9) enforce charging and discharging quantities P_{in}^t and P_{out}^t to be positive. Finally, Eq. (5-10) is introduced to prohibit simultaneous charging and discharging operation.

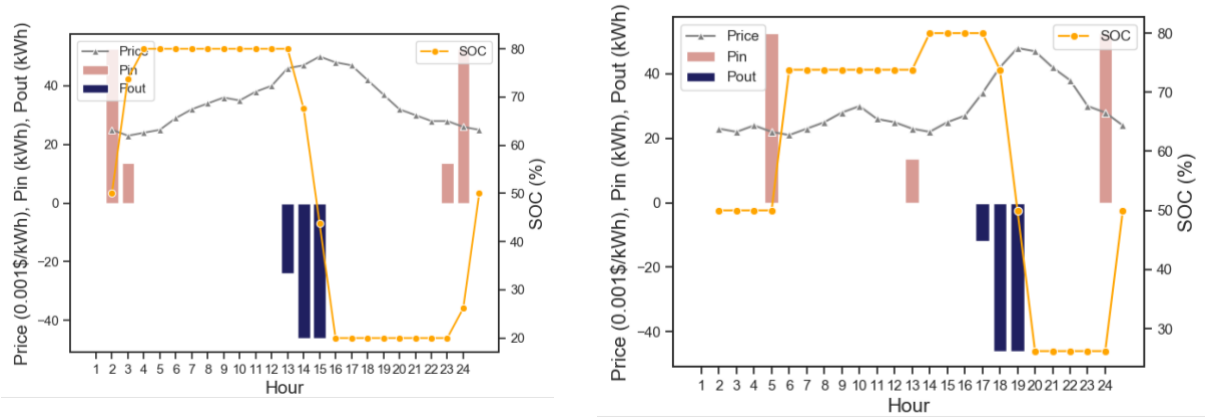
Battery degradation is considered in the study for its significance on influencing the long-term profit of battery owners. All batteries are faced with degradation in terms of capacity fading and increasing resistance. Batteries degrade even when not in operation, called calendar aging. External stress conditions, including high operation temperature, high and low SoC, high Depth of Discharge (DoD) and high current-rate/high power-rate [126], lead to accelerated aging and reduced lifetime. To prevent the battery from serious harms, following settings are considered.

1. High operation temperature is disregard in this study, because battery storages are usually associated with effective temperature control.
2. SoC^{min} and SoC^{max} are set accordingly to avoid high/low SoC and high DoD.
3. The second term is added in the objective function Eq.(5-1) to penalize the high current/power-rate behavior of the battery. By adding it, the degradation cost is approximated to be linearly related to the dis-/charging rate at each time interval. The degradation rate D_{kWh} estimates the degradation cost generated per kWh operation. Such a setting dis-encourages uneconomical cycling of the battery.

Although these stress conditions impact the battery lifespan in a complicated way. Such an impact is approximated to be linearly related the accumulated operating quantity to simplify the analysis, as battery degradation is not the focus in this study. For discussions on more realistic battery degradation modeling, please refer to [126][127] and [128].

According to the problem statement, the presence of Eq. (5-10) makes this problem nonlinear. However, it is an equality constraint which has been inherently taken care of by the problem

formulation. Because the simultaneous charging and discharging actions happened within a single time interval is not favored by the algorithm, as it only increases the degradation cost without bringing other benefits. Therefore, it is safe to remove Eq. (5-10) from the problem formulation without impacting its effectiveness, which turns the day-ahead battery arbitrage problem into a pure linear problem. Given the linear nature of the modified problem formulation (Eq.(5-1)-(5-9)), GNU Linear Programming Kit (GLPK) solver is applied for solving the problem.



(a) Price signal A *(b) Price signal B*

Figure 5-1. Optimized battery scheduling under dynamic price signals

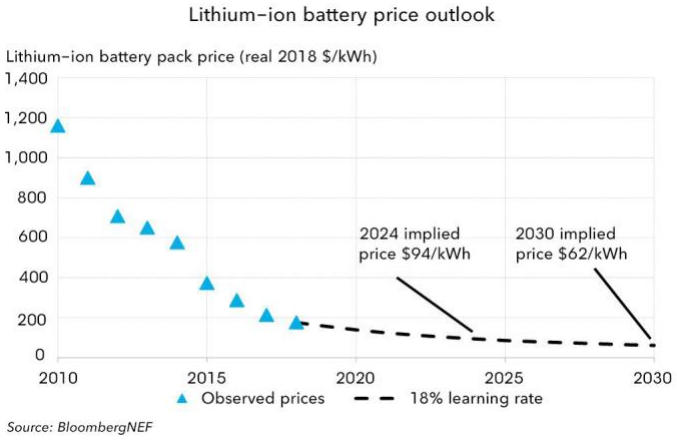


Figure 5-2. Projected Lithium-ion battery pack price fall

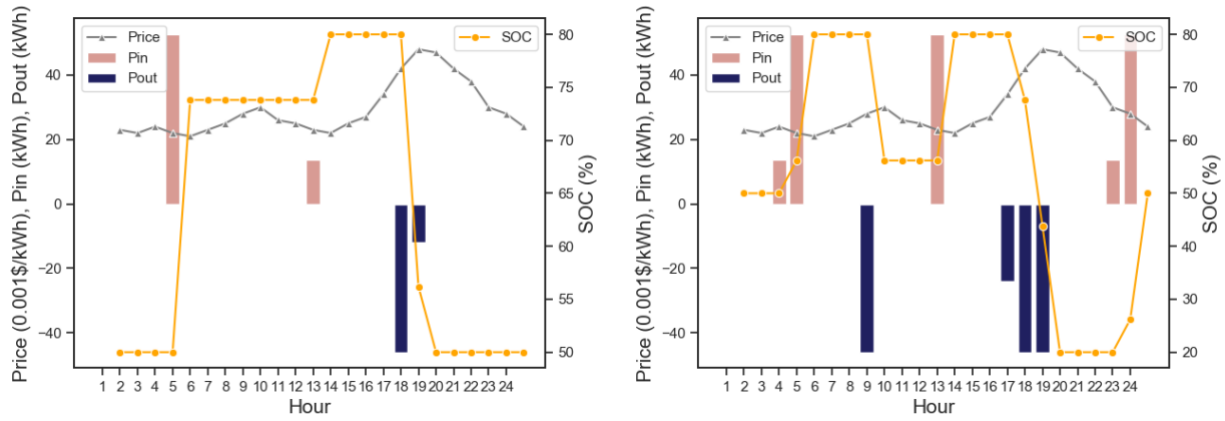
Figure 5-1 demonstrates operating schedules obtained by solving the battery arbitrage problem under two different price signals, ranging between 0.02$/kWh and 0.05$/kWh. Price signal A peaks at 14:00 PM and price signal B peaks at 18:00 PM. Apart from the price signal, all other battery characteristics are the same and configured taking the 210kWh model Tesla Powerpack as the reference (see *Table 5-2*). It is worth mentioning that the initial investment cost of the battery is assumed to be $62/kWh, as projected in 2030 by Bloomberg NEF (see Figure 5-2) [12]. Based on the $13020 investment cost and an assumption of 10000 lifecycles. The per-kWh degradation cost D_{kWh} is set to be 0.005$/kWh, representing an estimated 0.005$ cost per 1kWh operation.

Table 5-2. Parameter configurations

Parameter	**Value**
C	210 kWh
l_c	5 %
l_d	8 %
SoC^{min}	20 %
SoC^{max}	80 %
SoC^{end}	50 %
R^{min}	0 kW
R^{max}	50 kW
D_{kWh}	0.005 $\$/kWh$

The optimized grid service income under two different price signals are $2.42 and $2.20 respectively. However, in the meantime, $1.26 and $1.13 estimated degradation costs are generated. By comparing Figure 5-1 (a) with Figure 5-1 (b), it can be seen that different price signals lead to different battery schedules and SoC trajectories. Nevertheless, there are some common patterns shared between the results: First, the battery tends to charge at the lowest price and to discharge when the price goes up. Second, SoC is maintained to be within the band between 20% and 80% all the time. Third, SoC is set back to 50% at the end of the daily cycle in order to avoid high expense in the following day due to a low initial SoC. This verifies ability of the battery arbitrage tool to correctly respond to the price signal.

5.2 Sensitivity analyses of the price-responsive behavior

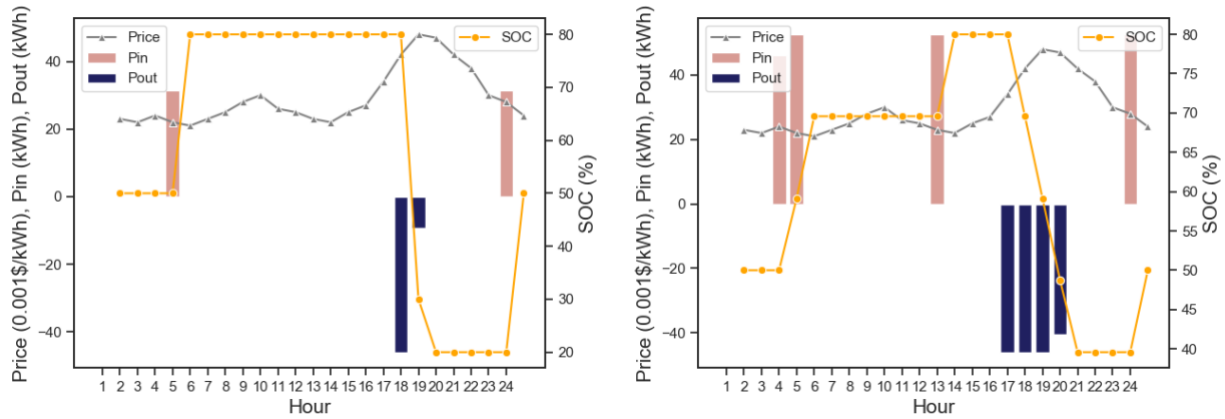


(a) $D_{kWh} = 0.01\ \$/kWh$ (b) $D_{kWh} = 0.001\ \$/kWh$

Figure 5-3. Comparison among different D_{kWh} values

To understand how different parameter impacts the decision making of batteries, three sets of comparative experiments are conducted. Figure 5-3, Figure 5-4 and Figure 5-5 compare the response of battery towards price signal B under different D_{kWh}, C and R^{max} respectively. Table 5-3 summarizes corresponding optimized grid service incomes and degradation costs under different settings.

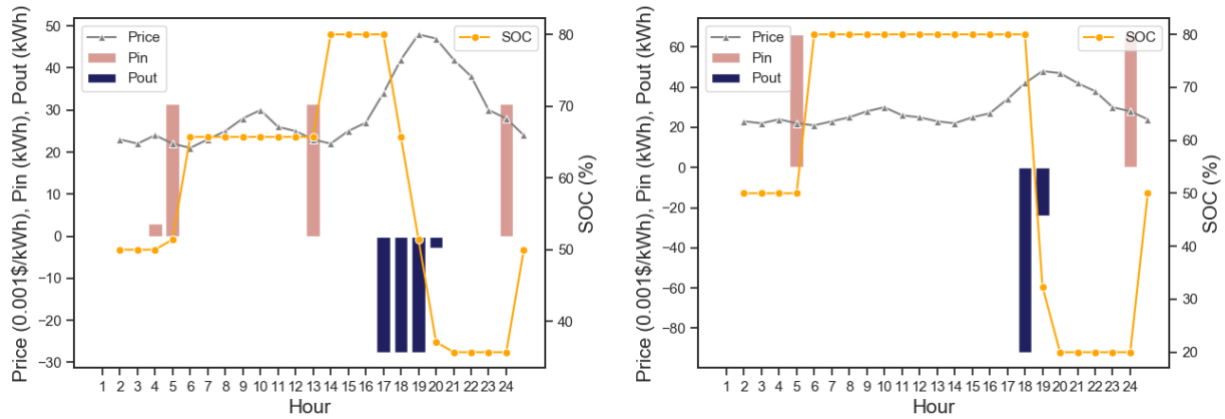
According to Figure 5-3 and Table 5-3, with a high degradation rate, battery tends to operate less frequently to avoid unnecessary degradation, which limits its ability to earn money from the market. For a 0.01 $\$/kWh$ degradation rate, see Figure 5-3 (a), only \$1.36 grid service income is achieved. It is because additional charging/discharging quantity cannot make the incremental income overpass the incremental degradation cost. By contrast, when cost of the battery drops, the battery can gain a higher income by increasing its operation. For a 0.001 $\$/kWh$ degradation rate, see Figure 5-3 (b), \$2.54 grid service income is made while only causing \$0.35 degradation cost.



(a) $C = 100\ kWh$ (b) $C = 480\ kWh$

Figure 5-4. Comparison among different C values

According to Figure 5-4 and Table 5-3, when raising the capacity from 100kWh to 480 kWh, the grid service income made by the battery increases by \$2.24 yet the degradation cost increases only by \$1.34. Therefore, a higher net income can be expected by applying batteries with higher capacity. It is mainly because the battery with higher capacity could better take advantage of gaps between price peaks and price valleys.



(a) $R^{max} = 30\ \$/kWh$ (b) $R^{max} = 100\ \$/kWh$

Figure 5-5. Comparison among different R^{max} values

Similar change is observed by increasing the maximum dis-/charging rate limit. When raising the limit from 30 kW to 100 kW, the grid service income increases by \$0.85 yet the degradation cost

increases only by $0.33. Therefore, it is concluded that a higher maximum dis-/charging rate limit also contributes to a greater profit of the battery. By comparing Figure 5-5 (a) with Figure 5-5 (b), with a stricter limit on dis-/charging rate, restoration and release of the energy are spread out across several hours around the price peak and price valley. In contrast, when the limit is loosed, larger amounts of energy are exchanging right at hours with highest and lowest prices, creating the largest price gap.

Table 5-3. Optimized financial outcomes of the battery under different parameter settings

Parameter	**Value**	**Grid service income**	**Degradation cost**
D_{kWh}	0.001 $\$/kWh$	$ 2.54	$ 0.35
	0.005 $\$/kWh$	$ 2.20	$ 1.13
	0.01 $\$/kWh$	$ 1.36	$ 1.26
C	100 kWh	$ 1.22	$ 0.60
	210 kWh	$ 2.20	$ 1.13
	480 kWh	$ 3.46	$ 1.94
R^{max}	30 kW	$ 1.71	$ 0.93
	50 kW	$ 2.20	$ 1.13
	100 kW	$ 2.56	$ 1.26

Another factor that significantly impacts battery's behavior and worth bringing to readers' attention is the shape of the price signal. Figure 5-6 provides the result generated by the battery arbitrage tool under the incentive of a more fluctuated price signal C. In comparison to price signal B, price signal C varies within the same range, yet has more price peaks and valleys. Note that mean values of price signals B and C are set to be the same to ensure a fair comparison. Under the incentive of price signal C, the battery cycles at a higher frequency and always discharges at price peaks and charges at nearby valleys, according to Figure 5-6. It almost doubles the grid service income, compared to price signal B, from $2.20 to $4.38, however only increases the net income by $0.3.

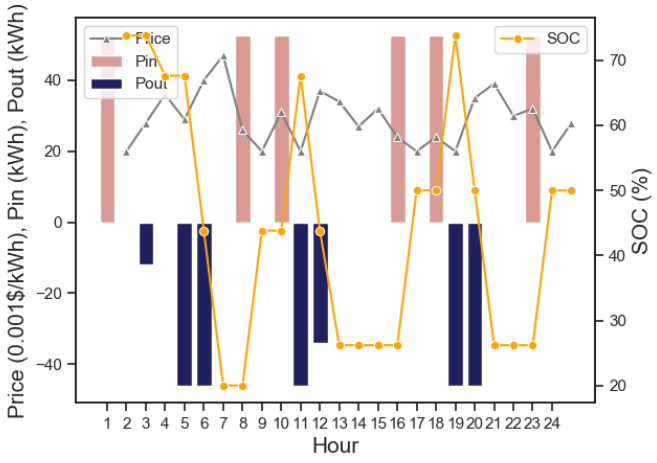


Figure 5-6. Battery scheduling under the price signal C

Figure 5-7 plots the result obtained under the incentive of price signal D, which has the same mean value of signals B and C, but fluctuates with a larger magnitude. Whereas signals B and C range between 0.02$/kWh and 0.05$/kWh, signal D falls into the band between 0.01$/kWh and 0.06$/kWh. According to Figure 5-7, again the battery operates at a higher cycling frequency and alternatively charges and discharges between price valleys and peaks. Table 5-4 compares the grid service income, degradation cost and net income associated with three price signals. According to Table 5-4, Figure 5-6 and Figure 5-7, by replacing the price signal C by price signal D, battery gains a much larger grid service income that is more than doubled from $4.38 to $10.30. However, in the mean tine, the cycling number and degradation cost only increase slightly leading to a 5 times net income.

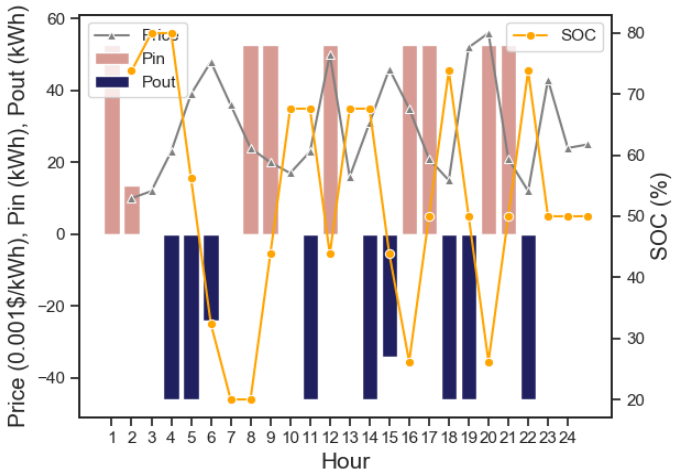


Figure 5-7. Battery scheduling under the price signal D

Above all, it is inferred that price difference drives the action of battery. Due to the limitation on battery capacity and dis-/charging rate, the battery tends to take the advantage of adjacent price peaks and valleys. Frequency and magnitude of the price fluctuation impact profit made by the battery arbitrage from two different aspects. Frequency of the fluctuation controls how often the battery cycles, i.e., the more frequently the price fluctuates the more adjacent price gaps will be created, resulting in greater cycling times. However, a high fluctuation frequency does not guarantee a considerable net income that is proportional to the operation times. It is the magnitude of fluctuation that helps largely improve the net income. Therefore, to encourage engagement of the battery, the retail market operator could consider expanding the fluctuation range of the price signal and letting the price varies steepest at the time when battery support is mostly required.

Table 5-4. Optimized financial outcomes of the battery under different price signals

Price signal	**Grid service income**	**Degradation cost**	**Net income**
B	$ 2.20	$ 1.13	$ 1.07
C	$ 4.38	$ 3.00	$ 1.38
D	$ 10.30	$ 4.13	$ 6.17

6. BI-LEVEL RETAIL PRICING MECHANISM

Based on the market model described in Section 3.1, a retail pricing mechanism is proposed in this section applying the bi-level optimization technique. Section 6.1 provides details of the upper-level and lower-level problem formulations. Section 6.2 linearizes the three-phase power flow model. How upper-level and lower-level problems are merged into a single-level problem is discussed in Section 6.3. Finally, Section 6.4 converts the nonlinear nonconvex single-level problem into a MILP problem.

6.1 Problem statement

Based on the understanding of the market model outlined in Section 3.1, detailed formulations for both lower-level and upper-level problems constituting the retail pricing problem are provided below.

Lower-level payoff maximization problem:

As the battery storage is selected as the representative elastic customer, the lower-level problem becomes the battery arbitrage problem as defined in Section 5. Under the incentive of retail grid service price, battery owners schedule their operations for the next-day to maximize their profits from the day-ahead retail market. Physical characteristics of the battery storage, including capacity (C), maximum dis-/charging rate (R^{max}), dis-/charging loss (l_c/l_d) and boundaries of the SOC (SoC^{min}, SoC^{max}, SoC^{end}), constrain the operation. Refer to Eq. (5-1)-(5-9) for detailed problem formulation.

Upper-level distribution system resources scheduling problem:

The upper-level problem is basically a distribution system resource scheduling problem, within which the RMO seeks to find an optimal dispatch schedule of all available resources by determining the market prices. Based on context of the problem described in Section 3.1, mathematical formulation of the upper-level problem is derived below.

Min

$$\sum_{t=1}^{24}\sum_{i\in\mathcal{N}_S}\alpha^t P_{o,i}^t + \sum_{t=1}^{24}\sum_{j\in\mathcal{N}_B}\beta_{GS}^t(P_{out,j}^t - P_{in,j}^t) \tag{6-1}$$

Subject to:

$$P_{o,k}^t + P_{out,k}^t - P_{in,k}^t - P_{L,k}^t = \Re\{V_k^t \cdot (I_k^t)^*\} \qquad \forall t, k \in \mathcal{N} \tag{6-2}$$

$$Q_{o,k}^t - Q_{L,k}^t = \Im\{V_k^t \cdot (I_k^t)^*\} \qquad \forall t, k \in \mathcal{N} \tag{6-3}$$

$$\boldsymbol{I^t = Y \times V^t} \qquad \forall t \tag{6-4}$$

$$\underline{P_{o,i}} \le P_{o,i}^t \le \overline{P_{o,i}} \qquad \forall t, i \in \mathcal{N}_S \tag{6-5}$$

$$\underline{Q_{o,i}} \le Q_{o,i}^t \le \overline{Q_{o,i}} \qquad \forall t, i \in \mathcal{N}_S \tag{6-6}$$

$$\underline{|V|}^2 \leq |V_k^t|^2 \leq \overline{|V|}^2 \qquad \forall t, k \in \mathcal{N} \qquad (6\text{-}7)$$

$$\boldsymbol{P_{out}}, \boldsymbol{P_{in}} = \mathcal{K}(\boldsymbol{\beta_{GS}}) \qquad (6\text{-}8)$$

Note that boldface letters represent vectors or matrixes, otherwise, indicate scalar values. Superscript $*$ indicate conjugate transformation. $\mathcal{N} \coloneqq \{1,2,\dots,N\}$ indexes all nodes in the system. $\mathcal{N}_S$ and $\mathcal{N}_B$ are subsets of $\mathcal{N}$ collecting indexes of slack nodes (nodes at the substation) and battery nodes (nodes connected with dispatchable batteries) with length of N_S and N_B. α^t and β_{GS}^t represent the wholesale price and retail grid service price at hour t. $P_{o,i}^t/Q_{o,k}^t$ stands for active/reactive power injection at the slack node i hour t. $P_{out,j}^t$ and $P_{in,j}^t$ indicate discharged and charged quantities of electricity from battery node j at hour t. $P_{L,k}^t/Q_{L,k}^t$ represents the nodal inelastic active/reactive demand at node k hour t. $\boldsymbol{V^t}$ and $\boldsymbol{I^t}$ denote voltage phasors and net current injections in the network at time t, in the shape of $N \times 1$. V_k^t and I_k^t specifically refer to the voltage phasor and net current injection at node k hour t. $\boldsymbol{Y}$ denotes the $N \times N$ system admittance matrix. It is a sparse and symmetric matrix with diagonal elements being self-admittances and non-diagonal elements being mutual-admittances.

According to Eq. (6-1), objective function of the upper-level problem contains two parts: costs paid for the bulk grid supply and battery arbitrages. It is targeted to be minimized. There are four types of constraints in the upper-level problem. Eq.(6-2)-(6-4) provide a complete three-phase power flow model. Eq.(6-5)-(6-6) specify limits of active and reactive power injections at the substation. Eq.(6-7) constraints the voltage magnitude at all nodes. Finally, function $\mathcal{K}$ is applied in Eq.(6-8) to abstractly express the multi-to-multi correlation between the retail grid service price signal ($\boldsymbol{\beta_{GS}}$) and the battery dispatch plan ($\boldsymbol{P_{out}}$ and $\boldsymbol{P_{in}}$). Explicit expressions describing such a relationship will be derived in the next subsection. Details about how the three-phase power flow model is constructed are presented below:

With the system admittance matrix $\boldsymbol{Y}$, the branch current flowing from node p to node q can be computed as:

$$I_{pq}^t = Y_{pp}V_p^t + Y_{pq}V_q^t \qquad (6\text{-}9)$$

By summing up all the currents flowing from node p towards adjacent nodes, the total current flowing out of node p can be calculated:

$$I_p^t = \sum_{q \in \mathcal{N}_p} (Y_{pp}V_p^t + Y_{pq}V_q^t) \qquad (6\text{-}10)$$

Where $\mathcal{N}_p$ denotes the set of nodes physically connected with node p. Then the complex power balance at node p can be expressed as:

$$V_p^t \cdot (I_p^t)^* = P_{o,p}^t + P_{out,p}^t - P_{in,p}^t - P_{L,p}^t + j(Q_{o,p}^t - Q_{L,p}^t) \qquad (6\text{-}11)$$

The above equation can be broken down into an active power balance equation and a reactive power balance equation, which gives Eq.(6-2) and (6-3). And Eq.(6-4) provides a compact expression of current flow equations Eq.(6-10) at all nodes using matrix multiplication.

Once the optimum retail grid service price ($\beta_{GS}^{t,*}$) and battery dispatch plans ($P_{out,j}^{t,*}$ and $P_{in,j}^{t,*}$) are obtained, the optimal retail energy price ($\beta_E^{t,*}$), total grid service subsidy (S_{GS}^*) and subsidy paid to individual elastic customer ($s_{GS,j}^*$) can then be determined using Eq.(6-12)-(6-14) based on the customized subsidy rate δ_{sub}. According to Eq. (6-14), the total grid service subsidy is distributed among elastic customers based on their contributions made on decreasing the overall system cost. Such a contribution is proportional to the grid service income each elastic customer makes out of the total grid service income.

$$\beta_E^{t,*} = \delta_{sub}\beta_E^{t,0} + (1-\delta_{sub})\frac{\sum_{i\in\mathcal{N}_S}\alpha^t P_{o,i}^t + \sum_{j\in\mathcal{N}_B}\beta_{GS}^{t,*}\left(P_{out,j}^{t,*} - P_{in,j}^{t,*}\right)}{\sum_{k\in\mathcal{N}}P_{L,k}^t} \tag{6-12}$$

$$S_{GS}^* = \delta_{sub}\sum_{t=1}^{24}[\beta_E^{t,0}\sum_{k\in\mathcal{N}}P_{L,k}^t - \alpha^t\sum_{i\in\mathcal{N}_S}P_{o,i}^t - \beta_{GS}^{t,*}\sum_{j\in\mathcal{N}_B}\left(P_{out,j}^{t,*} - P_{in,j}^{t,*}\right)] \tag{6-13}$$

$$s_{GS,j}^* = \frac{\sum_{t=1}^{24}\beta_{GS}^{t,*}\left(P_{out,j}^{t,*} - P_{in,j}^{t,*}\right)}{\sum_{t=1}^{24}[\sum_{j\in\mathcal{N}_B}\beta_{GS}^{t,*}\left(P_{out,j}^{t,*} - P_{in,j}^{t,*}\right)]}S_{GS}^* \tag{6-14}$$

Where $\beta_E^{t,0}$ denotes the retail energy price calculated when no elastic customer is in the market to provide the flexibility. Note that the problem statement provide in Eq. (6-1)-(6-7) are expressed with complex variables $\boldsymbol{V^t}$, $\boldsymbol{I^t}$ and $\boldsymbol{Y}$. However, those are not ready to be recognized by the algebraic modeling languages (AML). To facilitate implementation of the AML which supports only real numbers computations. Eq.(6-2)-(6-4) are further converted into Eq.(6-15)-(6-17). Where subscripts "real" and "imag" denote the real part and imaginary part of a complex variable. As the power balance at each node not only relates to its own voltage phasor but also dependents on voltage phasors at other nodes, compact matrix computations are adopted to express the active and reactive power balances in the whole system, as given in Eq.(6-15) and Eq.(6-16).

$$\begin{aligned}\boldsymbol{P_o^t + P_{out}^t - P_{in}^t - P_L^t} = \boldsymbol{V_{real}^t}\cdot\left(\boldsymbol{Y_{real}^t}\times\boldsymbol{V_{real}^t} - \boldsymbol{Y_{imag}^t}\times\boldsymbol{V_{imag}^t}\right)\\ + \boldsymbol{V_{imag}^t}\cdot\left(\boldsymbol{Y_{imag}^t}\times\boldsymbol{V_{real}^t} + \boldsymbol{Y_{real}^t}\times\boldsymbol{V_{imag}^t}\right)\end{aligned} \quad \forall t \tag{6-15}$$

$$\begin{aligned}\boldsymbol{Q_o^t - Q_L^t} = \boldsymbol{V_{imag}^t}\cdot\left(\boldsymbol{Y_{real}^t}\times\boldsymbol{V_{real}^t} - \boldsymbol{Y_{imag}^t}\times\boldsymbol{V_{imag}^t}\right)\\ - \boldsymbol{V_{real}^t}\cdot\left(\boldsymbol{Y_{imag}^t}\times\boldsymbol{V_{real}^t} + \boldsymbol{Y_{real}^t}\times\boldsymbol{V_{imag}^t}\right)\end{aligned} \quad \forall t \tag{6-16}$$

$$|\boldsymbol{V^t}|^2 = |\boldsymbol{V_{real}^t}|^2 + |\boldsymbol{V_{imag}^t}|^2 \quad \forall t \tag{6-17}$$

6.2 Linearization of the upper-level problem

Given the quadratic terms present in equality constraints Eq.(6-15)-(6-17) and inequality constraint Eq.(6-7), the upper-level problem is identified as a nonlinear and nonconvex problem. And it is extremely hard to solve. Therefore, a linearization technique proposed in [129] is adopted to further linearize the upper-level day-ahead resources scheduling problem.

By interpreting the fixed-point function derived from the full-structured AC power flow equations Eq.(6-2)-(6-4) using newton method, the nonlinear power flow model can be converted into several linear power flow models as given in Eq. (6-18)-(6-21). Therefore, system voltage and voltage magnitude profiles $\boldsymbol{V^t}$ and $|\boldsymbol{V^t}|$, required active and reactive power injections at substations $\boldsymbol{P_o^t}$ and $\boldsymbol{Q_o^t}$ to balance the system flow can be directly expressed as linear functions of the net active and reactive power injections at all the non-slack nodes.

$$\boldsymbol{V^t = A^t(P_{out}^t - P_{in}^t - P_L^t) - jA^t(-Q_L^t) + a^t} \qquad \forall t \qquad (6\text{-}18)$$

$$\boldsymbol{P_o^t = B_{real}^t(P_{out}^t - P_{in}^t - P_L^t) - B_{imag}^t(-Q_L^t) + b_{real}} \qquad \forall t \qquad (6\text{-}19)$$

$$\boldsymbol{Q_o^t = B_{imag}^t(P_{out}^t - P_{in}^t - P_L^t) + B_{real}^t(-Q_L^t) + b_{imag}} \qquad \forall t \qquad (6\text{-}20)$$

$$\boldsymbol{|V^t| = C_P^t(P_{out}^t - P_{in}^t - P_L^t) + C_Q^t(-Q_L^t) + c} \qquad \forall t \qquad (6\text{-}21)$$

Linear coefficients of the linear power flow models can be sufficiently determined by the system Y matrix, substation reference voltage V_0 and system voltage profile $\widehat{V^t}$ at a given feasible operating point:

$$\boldsymbol{w = -{Y_{LL}}^{-1}Y_{L0}V_0} \qquad (6\text{-}22)$$

$$\boldsymbol{W = diag(w)} \qquad (6\text{-}23)$$

$$\boldsymbol{A^t = {Y_{LL}}^{-1} diag(\widehat{V^t}^*)^{-1}} \qquad \forall t \qquad (6\text{-}24)$$

$$\boldsymbol{a = w} \qquad \forall t \qquad (6\text{-}25)$$

$$\boldsymbol{B^t = V_0 {Y_{0L}}^* (A^t)^*} \qquad \forall t \qquad (6\text{-}26)$$

$$\boldsymbol{b = V_0({Y_{00}}^* {V_0}^* + {Y_{0L}}^* w^*)} \qquad \forall t \qquad (6\text{-}27)$$

$$\boldsymbol{C_P^t = \big||W|\big| \Re\{W^{-1}A^t\}} \qquad \forall t \qquad (6\text{-}28)$$

$$\boldsymbol{C_Q^t = ||W||\, \Re\{W^{-1}(-jA^t)\}} \qquad \forall t \qquad (6\text{-}29)$$

$$\boldsymbol{c = ||w||} \qquad \forall t \qquad (6\text{-}30)$$

Where w represents zero-load voltage at all non-slack nodes. $\boldsymbol{Y_{LL}}$, $\boldsymbol{Y_{L0}}$, $\boldsymbol{Y_{0L}}$ and $\boldsymbol{Y_{00}}$ are fractions of the system admittance matrix $\boldsymbol{Y}$. $||\cdot||$ refers to element-wise norm-2 calculation. $diag(\cdot)$ yields square matrixes with the input vector being the diagonal elements. Since $\widehat{V^t}^*$ changes across time, first order coefficient matrixes $\boldsymbol{A^t}$, $\boldsymbol{B^t}$, $\boldsymbol{C_P^t}$ and $\boldsymbol{C_Q^t}$ are time dependent, yet intercepts vectors $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$ are fixed across time. Therefore, by replacing Eq.(6-19)-(6-30) with Eq.(6-15)-(6-17), and modifying the original voltage constraint Eq.(6-7) into Eq.(6-31), linearized power flow model and voltage constraints can be obtained.

$$\underline{|V|} \le V_k^t \le \overline{|V|} \qquad \forall t, k \in \mathcal{N} \qquad (6\text{-}31)$$

Accuracy of the linear power flow against its fully-constructed nonlinear counterpart has been compared in [129]. It is shown that the linear model behaves very well with relative errors for voltage phasor and voltage magnitude less than 0.2% when the power injections range between $-1s^{ref}$ and $2s^{ref}$ in a 13-bus feeder system. Where s^{ref} represents the reference power injections profile. Despite the compromise made on a small amount loss of accuracy, there are two major advantages of applying such a linearization. Firstly, solvability of the nonlinear model is pretty

sensitive to the selection of the initial point. It is hard to make a fair guess of the initial point that lies in the feasible region of the nonlinear model. However, by simply selecting the non-zero load operating point to initialize coefficients of the linear model can always guarantee finding the optimum of the linear model through several runs of iterations. Secondly, as the scale of the retail market and the number of market participants grow up, the computational complexity for solving the retail market pricing problem increases exponentially. Implementing the linear approximation instead of the fully-constructed nonlinear model could largely improve the computation efficiency so as to make the application of the retail pricing tool more practical.

6.3 Transformation of the bi-level optimization problem

The upper-level and lower-level problems discussed in Section 6.1 are independently formulated, and cannot be directly solved using off-the-shelf solvers. Therefore, in this study, the bi-level problem is converted into a single-level problem through plugging the Karush–Kuhn–Tucker (KKT) conditions of lower-level problems into the upper-layer constraints. Assume x^* and u^*, v^* being the primal and dual solutions of an optimization problem, KKT conditions can sufficiently determine the relationship between x^* and u^*, v^*. Under a strong duality, x^* and u^*, v^* necessarily satisfy the KKT conditions. Since the lower-level problem is a pure linear problem which satisfies the strong duality, its KKT conditions are applied in this study to alternatively express lower-level decision makings in an explicit way. Eq.(6-32)-(6-40) reformulate the lower-level problem as a standard minimization problem for the battery located at battery node j. Dual variables associated with each constraint are introduced aside to facilitate derivation of the KKT conditions.

Min

$$\sum_{t=1}^{24} -\beta_{GS}^{t}\left(P_{out,j}^{t} - P_{in,j}^{t}\right) + D_{kWh,j}R_{j}^{t} \tag{6-32}$$

Subject to:

$$SoC_j^1 = SoC_j^{init} + \frac{P_{in,j}^1(1-l_{c,j})}{C_j} - \frac{P_{out,j}^1}{(1-l_{d,j})C_j} \leftrightarrow u_{soc,j}^1 \qquad j \in \mathcal{N}_B \tag{6-33}$$

$$SoC_j^t = SoC_j^{t-1} + \frac{P_{in,j}^t(1-l_{c,j})}{C_j} - \frac{P_{out,j}^t}{(1-l_{d,j})C_j} \leftrightarrow u_{soc,j}^t \qquad t \in [2,24],\ j \in \mathcal{N}_B \tag{6-34}$$

$$R_j^t = P_{in,j}^t(1-l_{c,j}) + \frac{P_{out,j}^t}{1-l_{d,j}} \leftrightarrow u_{R,j}^t \qquad \forall t, j \in \mathcal{N}_B \tag{6-35}$$

$$SoC_j^{min} \leq SoC_j^t \leq SoC_j^{max} \leftrightarrow \underline{v_{soc,j}^t}, \overline{v_{soc,j}^t} \qquad t \in [1,23],\ j \in \mathcal{N}_B \tag{6-36}$$

$$SoC_j^{end} \leq SoC_j^{24} \leq SoC_j^{max} \leftrightarrow \underline{v_{soc,j}^{24}}, \overline{v_{soc,j}^{24}} \qquad j \in \mathcal{N}_B \tag{6-37}$$

$$R_j^{min} \leq R_j^t \leq R_j^{max} \leftrightarrow \underline{v_{R,j}^t}, \overline{v_{R,j}^t} \qquad \forall t, j \in \mathcal{N}_B \tag{6-38}$$

$$P_{in,j}^t \geq 0 \leftrightarrow \underline{v_{Pin,j}^t} \qquad \forall t, j \in \mathcal{N}_B \tag{6-39}$$

$$P_{out,j}^t \geq 0 \leftrightarrow \underline{v_{Pout,j}^t} \qquad \forall t, j \in \mathcal{N}_B \tag{6-40}$$

The Lagrangian function of the minimization problem is then rendered below:

$$\mathcal{L}\left(\boldsymbol{P_{out,j}}, \boldsymbol{P_{in,j}}, \boldsymbol{SoC_j}, \boldsymbol{R_j}, \boldsymbol{u_{soc,j}}, \boldsymbol{u_{R,j}}, \underline{\boldsymbol{v_{soc,j}}}, \overline{\boldsymbol{v_{soc,j}}}, \underline{\boldsymbol{v_{R,j}}}, \overline{\boldsymbol{v_{R,j}}}, \underline{\boldsymbol{v_{Pin,j}}}, \underline{\boldsymbol{v_{Pout,j}}}\right) =$$
$$\sum_{t=1}^{24}[-\beta_{GS,j}^{t}\left(P_{out,j}^{t}-P_{in,j}^{t}\right)+D_{kWh,j}R_j^t]-u_{soc,j}^{1}[SoC_j^1-SoC_j^{init}-\frac{P_{in,j}^{1}\left(1-l_{c,j}\right)}{C_j}+\frac{P_{out,j}^{1}}{\left(1-l_{d,j}\right)C_j}$$
$$-\sum_{t=2}^{24}u_{soc,j}^{t}\left[SoC_j^t-SoC_j^{t-1}-\frac{P_{in,j}^{t}\left(1-l_{c,j}\right)}{C_j}+\frac{P_{out,j}^{t}}{\left(1-l_{d,j}\right)C_j}\right]$$
$$-\sum_{t=1}^{24}u_{R,j}^{t}\left[R_j^t-P_{in,j}^{t}\left(1-l_{c,j}\right)-\frac{P_{out,j}^{t}}{1-l_{d,j}}\right]$$
$$+\sum_{t=1}^{23}[\underline{v_{soc,j}^{t}}\left(SoC_j^{min}-SoC_j^t\right)+\overline{v_{soc,j}^{t}}\left(SoC_j^t-SoC_j^{max}\right)]$$
$$+\underline{v_{soc,j}^{24}}\left(SoC_j^{end}-SoC_j^{24}\right)+\overline{v_{soc,j}^{24}}\left(SoC_j^{24}-SoC_j^{max}\right)$$
$$+\sum_{t=1}^{24}\left[\underline{v_{R,j}^{t}}\left(R_j^{min}-R_j^t\right)+\overline{v_{R,j}^{t}}\left(R_j^t-R_j^{max}\right)\right]$$
$$-\sum_{t=1}^{24}\underline{v_{Pin,j}^{t}}P_{in,j}^{t}-\sum_{t=1}^{24}\underline{v_{Pout,j}^{t}}P_{out,j}^{t} \tag{6-41}$$

Stationarity:

Stationarity conditions of the lower-level problem associated with battery node j can be obtained by setting the first-order partial derivatives of primal variables to be zero:

$$-\beta_{GS}^{t}-\frac{u_{soc,j}^{t}}{\left(1-l_{d,j}\right)C_j}+\frac{u_{R,j}^{t}}{1-l_{d,j}}-\underline{v_{Pout,j}^{t}}=0 \qquad \forall t,j\in\mathcal{N}_B \tag{6-42}$$

$$\beta_{GS}^{t}+\frac{u_{soc,j}^{t}(1-l_{c,j})}{C_j}+u_{R,j}^{t}(1-l_{c,j})-\underline{v_{Pin,j}^{t}}=0 \qquad \forall t,j\in\mathcal{N}_B \tag{6-43}$$

$$u_{soc,j}^{t+1}-u_{soc,j}^{t}+\overline{v_{soc,j}^{t}}-\underline{v_{soc,j}^{t}}=0 \qquad t\in[1,23],\ j\in\mathcal{N}_B \tag{6-44}$$

$$-u_{soc,j}^{24}+\overline{v_{soc,j}^{24}}-\underline{v_{soc,j}^{24}}=0 \qquad j\in\mathcal{N}_B \tag{6-45}$$

$$D_{kWh,j}-u_{R,j}^{t}+\overline{v_{R,j}^{t}}-\underline{v_{R,j}^{t}}=0 \qquad \forall t,j\in\mathcal{N}_B \tag{6-46}$$

Complementary slackness:

Complementary slackness describes a condition: when the dual variable is positive, the objective function could be increased if the corresponding primal constraint is relaxed. If not, the primal objective function value could still be improved.

$$\underline{v_{soc,j}^{t}}\left(SoC_j^{min} - SoC_j^{t}\right) = 0 \qquad t \in [1,23],\ j \in \mathcal{N}_B \qquad (6\text{-}47)$$

$$\underline{v_{soc,j}^{24}}\left(SoC_j^{end} - SoC_j^{24}\right) = 0 \qquad j \in \mathcal{N}_B \qquad (6\text{-}48)$$

$$\overline{v_{soc,j}^{t}}\left(SoC_j^{t} - SoC_j^{max}\right) = 0 \qquad \forall t, j \in \mathcal{N}_B \qquad (6\text{-}49)$$

$$\underline{v_{R,j}^{t}}\left(R_j^{min} - R_j^{t}\right) = 0 \qquad \forall t, j \in \mathcal{N}_B \qquad (6\text{-}50)$$

$$\overline{v_{R,j}^{t}}\left(R_j^{t} - R_j^{max}\right) = 0 \qquad \forall t, j \in \mathcal{N}_B \qquad (6\text{-}51)$$

$$\underline{v_{Pin,j}^{t}} P_{in,j}^{t} = 0 \qquad \forall t, j \in \mathcal{N}_B \qquad (6\text{-}52)$$

$$\underline{v_{Pout,j}^{t}} P_{out,j}^{t} = 0 \qquad \forall t, j \in \mathcal{N}_B \qquad (6\text{-}53)$$

Primal feasibility:

Primal feasibility refers to constraints of the primal problem as given in Eq. (6-33)-(6-40).

Dual feasibility:

Dual feasibility enforces dual variables associated with inequality constraints to be non-negative.

$$\underline{v_{soc,j}^{t}}, \overline{v_{soc,j}^{t}} \geq 0 \qquad \forall t, j \in \mathcal{N}_B \qquad (6\text{-}54)$$

$$\underline{v_{R,j}^{t}}, \overline{v_{R,j}^{t}} \geq 0 \qquad \forall t, j \in \mathcal{N}_B \qquad (6\text{-}55)$$

$$\underline{v_{Pin,j}^{t}} \geq 0 \qquad \forall t, j \in \mathcal{N}_B \qquad (6\text{-}56)$$

$$\underline{v_{Pout,j}^{t}} \geq 0 \qquad \forall t, j \in \mathcal{N}_B \qquad (6\text{-}57)$$

By replacing the stationarity, complementary slackness, primal feasibility and dual feasibility conditions listed above with Eq.(6-8) in the upper-level problem, the bi-level problem can then be fully converted into a single-level minimization problem with Eq.(6-1) being the objective function, and Eq.(6-5), Eq.(6-6), Eq.(6-19)-(6-31), Eq.(6-33)-(6-40) and Eq.(6-42)-(6-57) being constraints.

6.4 Linearization of the merged single-level problem

Although both upper-level and lower-level problems are linear programming problems in terms of decision variables $P_{out,j}^{t}$, $P_{in,j}^{t}$ and β_{GS}^{t} respectively, when these two linear problems merge together, additional nonlinearity and non-convexity are introduced. In the single-level problem, both $P_{out,j}^{t}$, $P_{in,j}^{t}$ and β_{GS}^{t} now become decision variables. Therefore, cross-product terms in the objective function and complementary slackness conditions turn this problem into a nonlinear nonconvex problem, which is still hard to solve. Therefore, two additional linearization techniques have been applied in this study to fully convert the nonlinear nonconvex single-level problem into a mix-integer linear programming (MILP) problem.

Linearization of the objective function:

The strong duality property of lower-level battery arbitrage problems is utilized to linearize the objective function Eq.(6-1). According to the strong duality theorem, because the lower-level battery arbitrage problem is linear, objective functions of the primal and dual problems equal to each other at their optimum. Therefore, Eq. (6-58) stands when the lower-level problem reaches its optimum, according to the problem formulation given in Eq.(6-32)-(6-40):

$$\begin{aligned}\sum_{t=1}^{24} \beta_{GS}^{t}\left(P_{out,j}^{t}-P_{in,j}^{t}\right)-D_{kWh,j}R_{j}^{t} \\ =-u_{soc,j}^{1}SoC_{j}^{init}+\sum_{t=1}^{24}\overline{v_{soc,j}^{t}}SoC_{j}^{max}-\sum_{t=1}^{23}\underline{v_{soc,j}^{t}}SoC_{j}^{min} \\ -\underline{v_{soc,j}^{24}}SoC_{j}^{end}+\sum_{t=1}^{24}\overline{v_{R,j}^{t}}\,R_{j}^{max}-\sum_{t=1}^{24}\underline{v_{R,j}^{t}}R_{j}^{min}\end{aligned} \tag{6-58}$$

Hereby, $\sum_{t=1}^{24}\underline{\beta_{GS}^{t}}\left(P_{out,j}^{t}-P_{in,j}^{t}\right)$ can be expressed as a linear combination of variables: R_{j}^{t}, $u_{soc,j}^{1}$, $\overline{v_{soc,j}^{t}}$, $\underline{v_{soc,j}^{t}}$, $\overline{v_{R,j}^{t}}$ and $\underline{v_{R,j}^{t}}$. Replacing the cross-product term $\sum_{t=1}^{24}\beta_{GS}^{t}\left(P_{out,j}^{t}-P_{in,j}^{t}\right)$ in Eq.(6-1) using Eq.(6-58) renders the linearized objective function:

$$\begin{aligned}\sum_{t=1}^{24}\sum_{i\in\mathcal{N}_S}\alpha^{t}P_{o,i}^{t}-\sum_{i\in\mathcal{N}_B}u_{soc,j}^{1}SoC_{j}^{init}+\sum_{t=1}^{24}\sum_{i\in\mathcal{N}_B}\overline{v_{soc,j}^{t}}SoC_{j}^{max} \\ -\sum_{t=1}^{23}\sum_{i\in\mathcal{N}_B}\underline{v_{soc,j}^{t}}SoC_{j}^{min}-\sum_{i\in\mathcal{N}_B}\underline{v_{soc,j}^{24}}SoC_{j}^{end}+\sum_{t=1}^{24}\sum_{i\in\mathcal{N}_B}\overline{v_{R,j}^{t}}R_{j}^{max} \\ -\sum_{t=1}^{24}\sum_{i\in\mathcal{N}_B}\underline{v_{R,j}^{t}}R_{j}^{min}+\sum_{i\in\mathcal{N}_B}D_{kWh,j}R_{j}^{t}\end{aligned} \tag{6-59}$$

Linearization the complementary slackness conditions:

Linearization of complementary slackness conditions Eq.(6-47)-(6-53) is fulfilled applying the "big-M" method. By introducing one binary variable for each complementary condition, original equalities constraints Eq.(6-47)-(6-53) are then converted into following inequality constraints.

$$0 \le \underline{v_{soc,j}^t} \le \underline{\omega_{soc,j}^t} M \qquad t \in [1,23],\ j \in \mathcal{N}_B \qquad (6\text{-}60)$$

$$0 \le SoC_j^t - SoC_j^{min} \le (1 - \underline{\omega_{soc,j}^t}) M \qquad t \in [1,23],\ j \in \mathcal{N}_B \qquad (6\text{-}61)$$

$$0 \le \underline{v_{soc,j}^{24}} \le \underline{\omega_{soc,j}^{24}} M \qquad j \in \mathcal{N}_B \qquad (6\text{-}62)$$

$$0 \le SoC_j^{24} - SoC_j^{end} \le (1 - \underline{\omega_{soc,j}^{24}}) M \qquad j \in \mathcal{N}_B \qquad (6\text{-}63)$$

$$0 \le \overline{v_{soc,j}^t} \le \overline{\omega_{soc,J}^t} M \qquad \forall t, j \in \mathcal{N}_B \qquad (6\text{-}64)$$

$$0 \le SoC_j^{max} - SoC_j^t \le (1 - \overline{\omega_{soc,J}^t}) M \qquad \forall t, j \in \mathcal{N}_B \qquad (6\text{-}65)$$

$$0 \le \underline{v_{R,j}^t} \le \underline{\omega_{R,j}^t} M \qquad \forall t, j \in \mathcal{N}_B \qquad (6\text{-}66)$$

$$0 \le R_j^t - R_j^{min} \le (1 - \underline{\omega_{R,j}^t}) M \qquad \forall t, j \in \mathcal{N}_B \qquad (6\text{-}67)$$

$$0 \le \overline{v_{R,j}^t} \le \overline{\omega_{R,J}^t} M \qquad \forall t, j \in \mathcal{N}_B \qquad (6\text{-}68)$$

$$0 \le R_j^{max} - R_j^t \le (1 - \overline{\omega_{R,J}^t}) M \qquad \forall t, j \in \mathcal{N}_B \qquad (6\text{-}69)$$

$$0 \le \underline{v_{Pin,j}^t} \le \underline{\omega_{Pin,j}^t} M \qquad \forall t, j \in \mathcal{N}_B \qquad (6\text{-}70)$$

$$0 \le P_{in,j}^t \le (1 - \underline{\omega_{Pin,j}^t}) M \qquad \forall t, j \in \mathcal{N}_B \qquad (6\text{-}71)$$

$$0 \le \overline{v_{Pout,j}^t} \le \overline{\omega_{Pout,j}^t} M \qquad \forall t, j \in \mathcal{N}_B \qquad (6\text{-}72)$$

$$0 \le P_{out,j}^t \le (1 - \overline{\omega_{Pout,j}^t}) M \qquad \forall t, j \in \mathcal{N}_B \qquad (6\text{-}73)$$

$$\underline{\omega_{soc,j}^t}, \overline{\omega_{soc,J}^t}, \underline{\omega_{R,j}^t}, \overline{\omega_{R,J}^t}, \underline{\omega_{Pin,j}^t}, \underline{\omega_{Pout,j}^t} \in \{0,1\} \qquad \forall t, j \in \mathcal{N}_B \qquad (6\text{-}74)$$

Where M is a sufficiently large number. The binary variable controls whether the upper bound of the inequality constraint is hit or not. Each pair of the complementary multiplier are controlled by the same binary variable such that at least one of the multipliers is forced to be zero. For example, when $\underline{\omega_{soc,j}^t} = 0$, $\underline{v_{soc,j}^t} = 0$, otherwise, $SoC_j^{min} - SoC_j^t = 0$. Such a design provides equivalent mix-integer linear expressions of original complementary slackness conditions.

Given above transformations, the single-level retail pricing problem is now converted into a mixed integer linear programming (MILP) problem with Eq.(6-59) being the objective function, and Eq.(6-5), Eq.(6-6), Eq.(6-19)-(6-31), Eq.(6-33)-(6-40), Eq.(6-42)-(6-46), Eq.(6-54)-(6-57) and Eq.(6-60)-(6-74) being constraints. It can be efficiently solved by off-the-shelf solvers, e.g. GLPK, GUROBI, CPLEX.

7. RETAIL MARKET SIMULATION

Sections 3-6 lay the theoretical foundation of the retail market design proposed in this study. A comprehensive case study is carried out in this section to evaluate the performance of the proposed retail market solution. Section 7.1 lists questions to be answered in the case study and designs the experiments accordingly. Section 7.2 summarizes the experimental setting of the simulated retail market. Detailed result analyses can be found in Sections 7.3-7.7.

7.1 Experimental design

To better understand the value and limitation of the proposed retail market solution, it is important to address following questions in the case study:

1. Will market participants, either elastic or inelastic customers, have the incentive to participate in the retail market?
2. Is the retail grid service subsidy necessary to be added in the pricing system?
3. Will the impact of network constraints on decision making be reasonably captured by the pricing scheme?
4. How do equipment and market parameters influence the market settlement result?
5. What is the advantage of the proposed retail market solution against other alternatives?

To answer above questions, following numerical experiments are designed accordingly:

- **Rationality Validation:** to verify the rationality of both elastic and inelastic customers to participate in the market.
- **Necessity of the Subsidy**: to compare market settlement results obtained under two different retail market settings before and after adding the grid service subsidy.
- **Network Constraints:** to compare market solutions obtained under three different levels of network constraints. Detailed descriptions and mathematical formulations for three levels of network constraints settings are presented in Table 7-1.
- **Parameter Sensitivity:** to analyze how battery degradation rate, capacity, maximum dis-/charging rate and retail grid service price mean value impact the market settlement result.
- **Proposed vs. Alterative Pricing Schemes:** to compare the performance of the proposed profit-neutral double-price-signal retail market solution against other possible alternatives.

7.2 Parameter configuration

Consider a retail market, scaled at an IEEE 34-bus system, with up to 20 batteries (representing elastic customers) and multiple fixed electricity demands (representing inelastic customers) distributed at different nodes. The day-ahead fixed system load varies between 594.22 kWh (17:00 PM-18:00 PM) and 1009.37 kWh (3:00 AM – 4:00 AM), and averaged at 840.58 kWh. The day-ahead wholesale market ranges between 0.060$/kWh (10:00 AM-11:00 AM) and 0.025$/kWh (2:00 AM-3:00AM), and averaged at 0.041$/kWh.

Table 7-1. Three levels of network constraints

	DESCRIPTION	OBJECTIVE	CONSTRAINTS
NO NETWORK	Without considering any network constraints, but only the active demand and supply balance.	*Min* Eq. (6-59)	*Subject to:* $\sum_{i\in\mathcal{N}_S} P_{o,i}^t + \sum_{j\in\mathcal{N}_B}(P_{out,j}^t - P_{in,j}^t) = \sum_{k\in\mathcal{N}} P_{L,k}^t \quad \forall t$ *And* Eq. (6-5), (6-6), (6-33)-(6-40),(6-42)-(6-46), (6-54)-(6-57), (6-60)-(6-74)
PARTIAL NETWORK	Added partial network constraints by considering the distribution system line losses.	*Min* Eq. (6-59)	*Subject to:* Eq. (6-5), (6-6), (6-33)-(6-40), (6-42)-(6-46), (6-54)-(6-57), (6-60)-(6-74), (6-19), (6-20), (6-22)-(6-27)
FULL NETWORK	Added full network constraints by considering both line losses and voltage constraints.	*Min* Eq. (6-59)	*Subject to:* Eq. (6-5), (6-6), (6-33)-(6-40), (6-42)-(6-46), (6-54)-(6-57), (6-60)-(6-74), (6-19)-(6-21), (6-22)-(6-31)

Batteries are configured with identical parameters, as summarized in Table 7-2, taking Tesla Powerpack 210kWh model as the reference. The battery degradation rate is calculated from an optimistic perspective. It is roughly estimated based on the projected per kWh battery investment cost ($62\$/\text{kWh}$) in 2030 (see Figure 5-2) and a 10000-times charging cycles life-span.

Table 7-2. Common parameters shared by 20 batteries

Parameter	Definition
$\boldsymbol{C}$	210 kWh
$\boldsymbol{l_c}$	4 %
$\boldsymbol{l_d}$	4.5 %
$\boldsymbol{R^{min}}$	0 kW
$\boldsymbol{R^{max}}$	50 kW
$\boldsymbol{D_{kWh}}$	0.005 $\$/kW$h

However, SoCs of batteries are initialized differently at the beginning of the operating day. And batteries are distributed at different locations. Detailed settings can be found in Table 7-3. In addition to that, elastic customers who own batteries are with different threshold subsidy rate δ_{sub}, which means their willingness to participate in the retail market varies against the system subsidy rate. For instance, when $\delta_{sub} = 34\%$, there are 11 batteries willing to participate in the retail market. When δ_{sub} raises to 50%, all 20 batteries will join the market.

Table 7-3. Unique parameters of different batteries

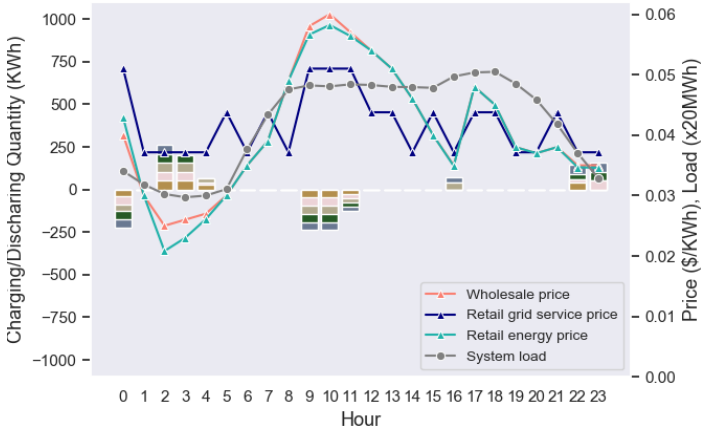
Battery No.	SoC^{init}	SoC^{end}	Location	Threshold δ_{sub}
1	80%	80%	850.1	12%
2	60%	60%	850.2	14%
3	40%	40%	850.3	16%
4	50%	50%	832.2	18%
5	40%	40%	858.2	20%
6	30%	30%	888.3	22%
7	50%	50%	834.3	24%
8	65%	65%	860.3	26%
9	60%	60%	842.3	28%
10	75%	75%	836.1	30%
11	30%	30%	836.2	32%
12	45%	45%	836.3	34%
13	50%	50%	840.2	36%
14	70%	70%	862.1	38%
15	50%	50%	862.2	40%
16	20%	20%	862.3	42%
17	50%	50%	844.3	44%
18	60%	60%	846.3	46%
19	20%	20%	848.1	48%
20	30%	30%	852.3	50%

M and C_β are set to be 10000 and 1. The retail pricing problem is programmed and solved in Python environment, by implementing Pyomo, an extensible Python-based open-source optimization modeling language, and GLPK (GNU Linear Programming Kit), a stand-alone LP/MIP solver.

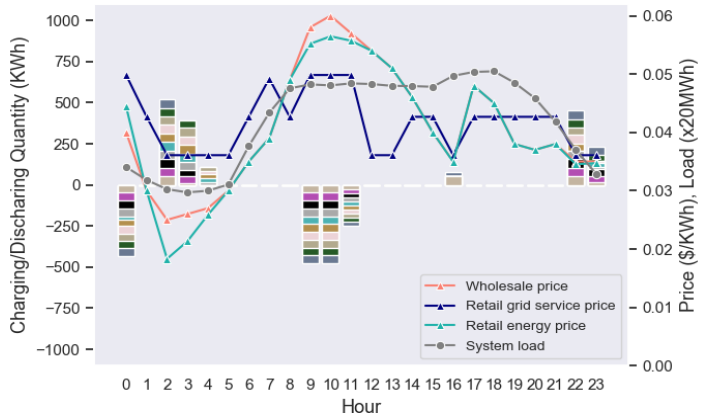
7.3 Rationality of customers to participate in the retail market

Figure 7-1 (a)-(d) plot market settlement results obtained under different subsidy rate, respectively when δ_{sub} equals to 20%, 30%, 40% and 50%, and correspondingly 5, 10, 15 and 20 batteries participate in the retail market. The gray and orange curves illustrate fluctuation of the system-level fixed load and wholesale price in the operating day; The navy and blue curves indicate the optimized retail grid service price and retail energy price; Bar blocks in different colors represent the optimized charging (above zero) and discharging (below zero) operations of different batteries. According to Figure 7-1, different market settlement solutions are generated when the number of elastic participants changes. However, batteries all tend to discharge at hours when the wholesale

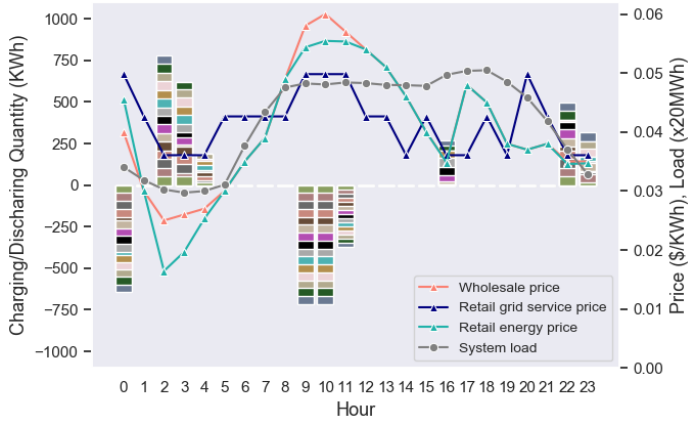
price goes high and charge when the wholesale price becomes low. This reflects the profit-driven nature of elastic customers.



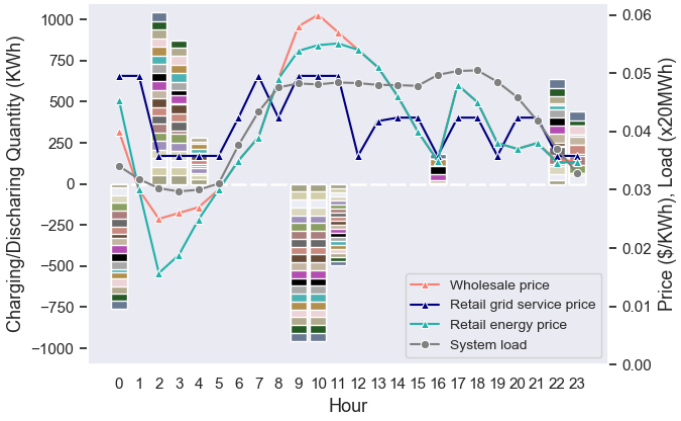
(a) δ_{sub}=20% (5 batteries)



(b) δ_{sub}=30% (10 batteries)



(c) δ_{sub}=40% (15 batteries)



(d) δ_{sub}=50% (20 batteries)

Figure 7-1. Optimized retail market solutions under different subsidy rates (No Network)

By comparing above market solutions, it is shown that the retail energy price is lower than the wholesale price for most of the time in the day. On the one hand, the retail energy price near wholesale price peak hours drops due to the cheaper replacement of supplies from batteries rather than from the substation. On the other hand, the retail energy price near wholesale price valley hours decreases as batteries purchase their charging quantities at a price higher than the wholesale price. In addition to that, the retail energy price reduces more significantly as the number of elastic participants grows, indicated by a wider gap between the wholesale price and the retail energy price. This is because a larger number of batteries could contribute to a greater amount of operational flexibility. Therefore, by participating in the proposed retail market, inelastic customers could be benefited by a cheaper electricity price compared to directly purchasing their demands from the wholesale market. However, this verifies only the rationality of inelastic customers. To verify the rationality of elastic customers, it is necessary to look at revenues made by them from the retail market. Table 7-4 and Table 7-5 calculate financial indicators relevant to inelastic and elastic customers correspondingly when δ_{sub} equal to 0%, 20%, 30%, 40% and 50%. According to Table 7-5, both gross revenue and discounted revenue made by elastic customers are above zero when at least one battery is in the retail market. It means the retail energy price is not lowered at the expense of elastic customers' sacrifices. Moreover, as the subsidy rate raises, total electricity bill saving, gross revenue and discounted revenue all increase. It tells that a higher subsidy rate benefits not only elastic customers but also inelastic customers.

Above all, it is proven that both inelastic and elastic customers are under the incentive to participate in the proposed retail market, as both electricity bill savings and discounted revenues can be achieved by inelastic and elastic customers.

Table 7-4. Financial analyses for inelastic customers under different subsidy rates

	Total system cost ($)	*Total electricity bill ($)*	*Total electricity bill saving ($)*
$\delta_{sub} = 0\%$ *(no battery)*	850.76	850.76	0.00
$\delta_{sub} = 20\%$ *(5 batteries)*	841.77	843.57	7.20
$\delta_{sub} = 30\%$ *(10 batteries)*	832.96	838.30	12.46
$\delta_{sub} = 40\%$ *(15 batteries)*	823.50	834.40	16.36
$\delta_{sub} = 50\%$ *(20 batteries)*	812.88	831.82	18.94

Table 7-5. Financial analyses for elastic customers under different subsidy rates

	Total grid service income ($)	*Total grid service subsidy ($)*	*Total gross revenue ($)*	*Total discounted revenue ($)*
$\delta_{sub} = 0\%$ *(no battery)*	0.00	0.00	0.00	0.00
$\delta_{sub} = 20\%$ *(5 batteries)*	8.64	1.80	10.44	1.80
$\delta_{sub} = 30\%$ *(10 batteries)*	17.15	5.34	22.49	5.34
$\delta_{sub} = 40\%$ *(15 batteries)*	25.66	10.90	36.56	10.90
$\delta_{sub} = 50\%$ *(20 batteries)*	33.17	18.94	52.11	18.94

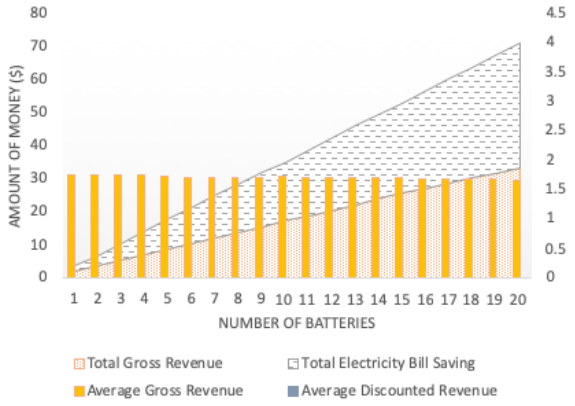
7.4 Impacts of the retail grid service subsidy

To better understand how the grid service subsidy makes a difference on incentivizing market participants, a comparative experiment is conducted the next. It compares market settlement results against the number of engaged elastic customers before and after adding the grid service subsidy, as presented in Figure 7-2. Stacked area plots in Figure 7-2 illustrate how the total electricity bill saving and total gross revenue vary against the number of batteries. Bar plots in Figure 7-2, on the other hand, indicate averaged gross and discounted revenues earned per elastic customer.

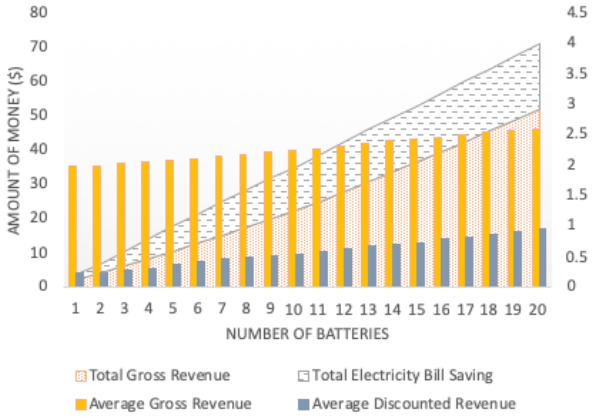
By comparing Figure 7-2 (a) with Figure 7-2 (b), it is found that the total area of stacked plots in both subfigures keeps the same. It is just a part of total electricity bill saving has been transferred into total gross revenue in form of grid service subsidy. Before adding the subsidy, the average gross revenue slightly declines as the number of elastic customers grows, attributing to a more intensive competition among a greater number of batteries. In addition to that, the grid service income is almost balanced out by the degradation cost resulting in a near-zero average discounted revenue. Therefore, rational elastic customers will not choose to join the retail market without subsidy as they cannot make enough profit in long term. However, after adding the subsidy, the average discounted revenue becomes non-zero. Also, as more batteries join the market, both gross and discounted revenues earned per elastic customer are expected to increase regardless of the growing competition concern. Although inelastic customers have to give up a part of their benefit on saving the electricity bill, such a compromise ensures the participation of elastic customers, which in turn secures the benefit of themselves. Under the additional incentive given by grid service subsidy, more elastic customers would be encouraged to participate in the market as it benefits not only the inelastic customer but also the elastic customer itself.

Moreover, according to Figure 7-2, under the threshold subsidy preference assumed in Table 7-3,

it is most beneficial to set the subsidy rate to be 50% as it creates highest averaged benefits to both elastic customers and inelastic customers in the simulated case. The RMO could also take use of this financial analysis graph to determine the optimal subsidy rate given different threshold subsidy preference.



(a) Without subsidy

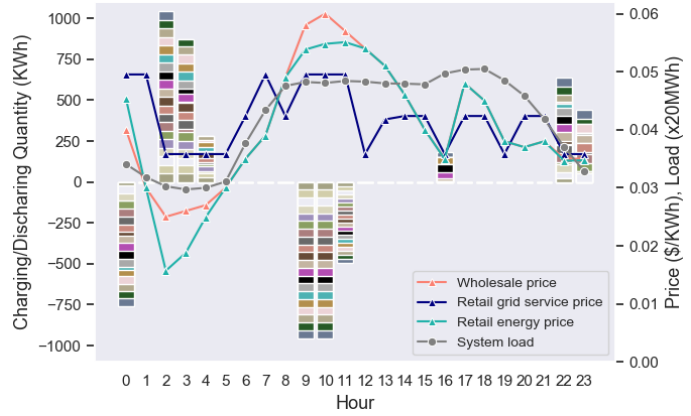


(a) With subsidy

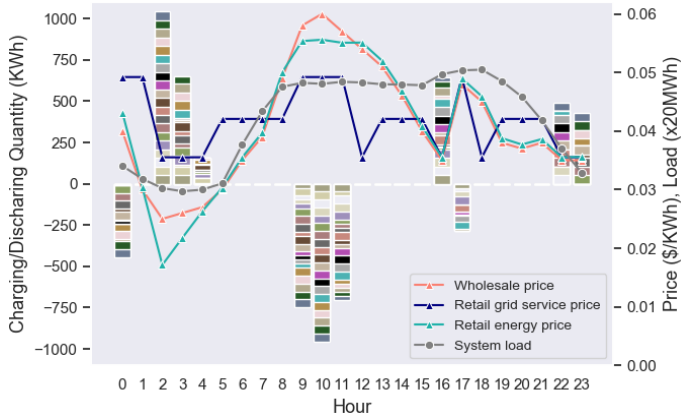
Figure 7-2. Financial analyses against the number of elastic customers before and after adding the grid service subsidy

7.5 Impacts of network constraints

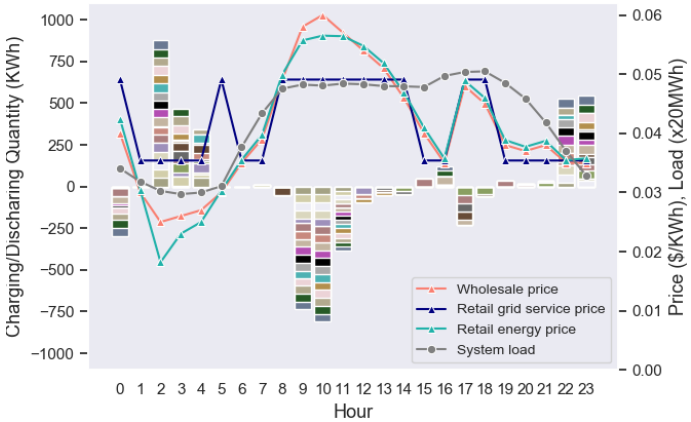
To capture impacts of network constraints on the retail market solution, three experiments are conducted by gradually adding network constraints into the problem formulation (refer to Table 7-1 for more details). The comparative results are visualized in Figure 7-3. Corresponding financial analyses, including total electricity bill saving and total gross/discounted revenue are computed in Table 7-6. Note that, all three cases are simulated with 20 elastic customers' participation. Results combined shown in Figure 7-3 and Table 7-6 imply that various network constraints influence decisions made by the RMO in different ways. Decision making of the RMO is pretty straightforward under the "No Network" condition, however, becomes more complicated as more influential factors are taken into accounts. Detailed observations are summarized below:



(b) No Network



(c) Partial Network



(d) Full Network

Figure 7-3. Optimized retail market solutions under different levels of network constraints

1. According to Figure 7-3 (a), in the "No Network" setting, where only the active power supply and demand balance is considered, the optimal pricing solution is fully determined by the wholesale price signal and has nothing to do with the system load profile. It is

because shifting the system load from wholesale peak price hours to valley price hours is the only way to effectively reduce the system cost, when the line loss is ignored.

2. Line losses contribute to a higher system cost, because the substation needs to provide more supply than demand to compensate line losses. That's why in Table 7-6 original system costs (i.e. the system cost generated when no flexibility is provided from elastic customers) of "Partial Network" and "Full Network" cases are higher than that of the "No Network" case. Such a reason also explains why retail energy price curves in Figure 7-3 (b) and (c) are shifted up compared to that in Figure 7-3 (a). According to Table 7-6 the total cost spent on line losses is \$15.64 (=\$866.40-\$850.76) when there are only inelastic customers in the system.
3. By comparing Figure 7-3 (b) with Figure 7-3 (a), it is observed that not only the wholesale price signal but also the system load profile drive the decision making of retail pricing. The retail grid service price obtained from the "Partial Network" case tends to put a higher incentive at the peak load hour (4:00 PM-5:00 PM), compared to that obtained from the "No Network" case. The reason behind this is that a higher system load usually leads to a higher line loss. Hereby, shifting the demand from peak load hours to valley load hours can help reduce the system loss, and thus decrease the overall system cost. However, it is worth to mention that since the line losses is comparatively small, therefore, wholesale price profile is still the main driver for determining the retail price in the "Partial Network" case.
4. By comparing Figure 7-3 (c) with Figure 7-3 (a)(b), it is found that the pricing solution obtained in a "Full Network" setting tends to spread out the battery discharging operation for a longer period compared to that of "Not Network" and "Partial Network" cases. It is the voltage constraint that leads to such a difference. Due to the limitation on voltage magnitude, batteries cannot be freely charged or discharged at the same hour, even though that gives the highest financial benefit to minimize the system cost. The concurrent charging and discharging behavior of multiple batteries could easily result in voltage violation and endanger the system stability.
5. Based on Table 7-6, because of the addition of network constraints, both electricity bill saving and gross/discounted revenue made under the retail pricing incentive cannot actually reach the ideal amount as calculated in the "No Network" setting. The stricter the constraints are, the less electricity bill saving and gross/discounted revenue can be achieved.

Based on observations made above, it is obvious that network constraints play a significant role on shaping the retail pricing solution. For real-world applications, it is not fair to apply a retail pricing mechanism developed under no network consideration or partial network consideration, as the resulted pricing solution will mislead the operation of price-responsive DERs and result in unacceptable system instability or failure. However, analyses based on the retail market conceptual design in a "No Network" setting, can still be helpful on verifying basic functionalities of the market, e.g. whether the retail market can provide proper incentives to market participants.

Table 7-6. Financial analyses under three different network constraint conditions

	ORIGINAL SYSTEM COST ($)	TOTAL ELECTRCITY BILL SAVING ($)	TOTAL GROSS REVENUE ($)	TOTAL DISCOUNTED REVENUE ($)
NO NETWORK	850.76	18.94	52.11	18.94
PARTIAL NETWORK	866.40	18.65	51.45	18.65
FULL NETWORK	866.40	15.98	44.93	15.98

7.6 Sensitivity analysis

The next step in the case study is to analyze how different equipment and market parameters influence the electricity bill saving and revenue gaining potentials in the retail market. Four financial indicators are of the interest to be displayed. They are total electricity bill saving, total grid service income, total grid service subsidy and total social welfare (i.e. the sum of first three variables, see Eq.(7-1)), representing benefits brought to inelastic customers, elastic customers as well as the whole system.

$$\begin{aligned} \boldsymbol{Social\ welfare} \\ &= \boldsymbol{Electricity\ bill\ saving + Grid\ service\ income} \\ &+ \boldsymbol{Grid\ service\ subsidy} \end{aligned} \tag{7-1}$$

7.6.1 Sensitivity of battery parameters

The first parameter under investigation is the battery degradation rate D_{kWh}. Figure 7-4 is obtained by varying D_{kWh} from 0.0005 $/kWh to 0.013 $/kWh, where bars in yellow, green and gray blue indicate total electricity bill saving, total grid service subsidy and total grid service income respectively. The gray line in Figure 7-4 is added to demonstrate variation of their sum (denoted as total social welfare).

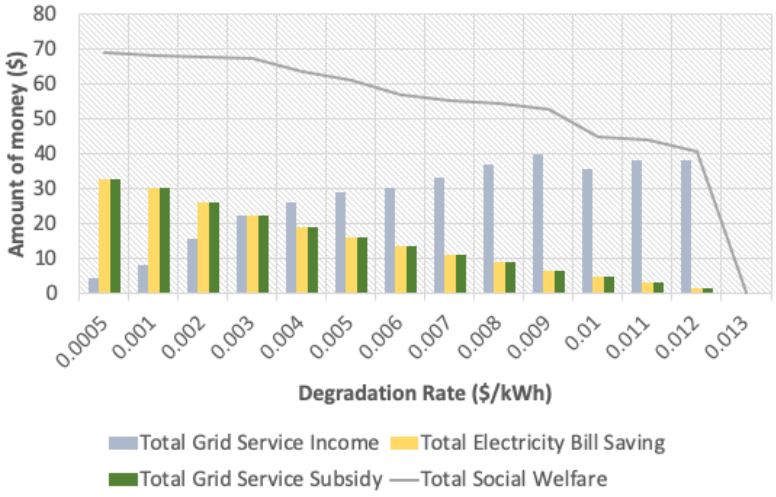
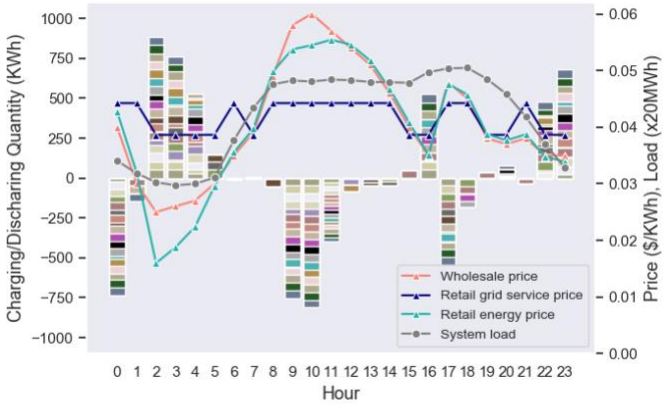
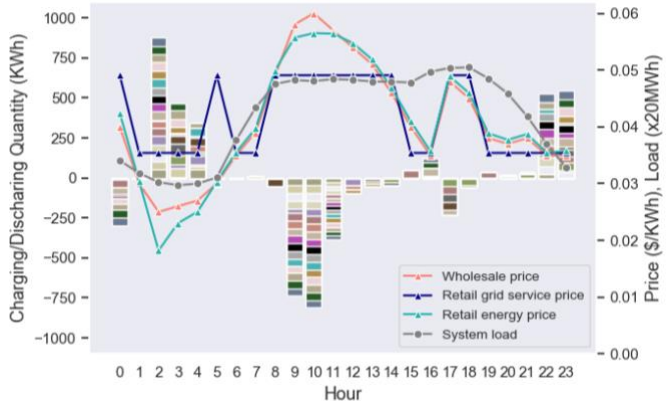


Figure 7-4. Sensitivity analysis of D_{kWh}

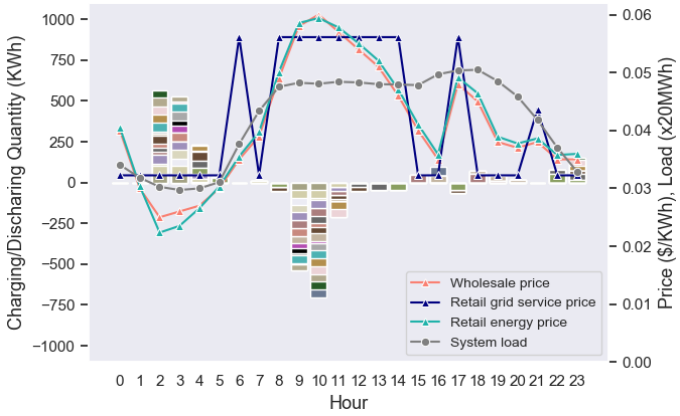
According to Figure 7-4, overall, a higher degradation rate slightly increases the grid service income made by batteries. Surprisingly, the grid service income reaches its maximum when D_{kWh} raises to 0.012 $/kWh. This seems counter-intuitive, because a higher degradation rate should usually dis-encourage the operation of batteries, as what is observed in Table 5-3. The reason behind this can be found in Figure 7-5, where retail market solutions at three selected degradation rates are plotted for details.



(a) $D_{kWh} = 0.001\ \$/kWh$



(b) $D_{kWh} = 0.005\ \$/kWh$



(c) $D_{kWh} = 0.01\ \$/kWh$

Figure 7-5. Retail market solutions under different D_{kWh}

By comparing navy curves in Figure 7-5, it is shown that as the degradation rate increases, the optimized retail grid service price fluctuates more significantly to better incentivize batteries. Therefore, despite the higher degradation cost, elastic assets owners can still expect a higher short-term income made from the retail market, thanks to the adaptive adjustment of the pricing strategy. However, due the considerable gradation effect of batteries, the grid service income is almost balanced out by the degradation cost. In the meantime, by viewing Figure 7-4 from right to left, as the degradation rate reduces, both total electricity bill saving and grid service subsidy increase. They equal to each other due to the 50% subsidy rate setting. Similarly, the total social welfare improves as the degradation rate reduces. It implies that technology improvements on reducing the battery investment cost is encouraged by the retail pricing mechanism, as batteries with lower degradation cost are more profitable in long term when joining the retail market. There is another observation made from Figure 7-4 worth to mention. For this particular retail market and system setting, when the degradation rate reaches above 0.012$/kWh, no incentive will be given to batteries.

Similarly, impacts of capacity C and maximum dis-/charging rate R^{max} on the retail market solution are studied the next. Financial analyses under different values of C and R^{max} are presented in Figure 7-6 and Figure 7-7, respectively.

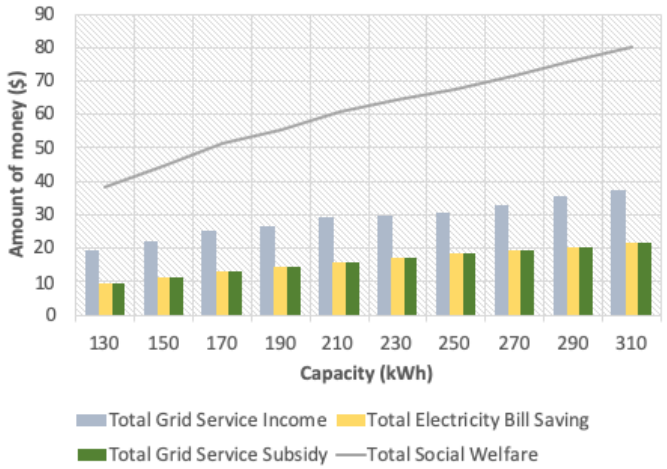


Figure 7-6. Sensitivity analysis of C

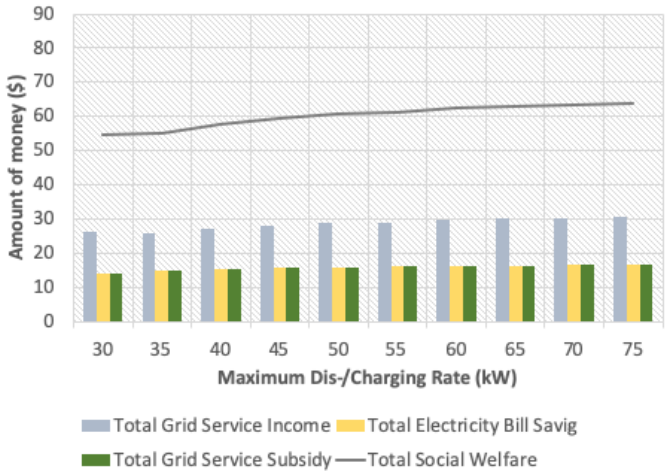


Figure 7-7. Sensitivity analysis of R^{max}

According to Figure 7-6 and Figure 7-7, total electricity saving bill, total grid service income/subsidy are less sensitive to battery capacity and maximum dis-/charging rate versus to the degradation rate. When the battery capacity and maximum dis-/charging rate increase, the electricity bill saving and grid service income/subsidy vary in a same direction, unlike what happens to the degradation rate. It is observed from Figure 7-6 and Figure 7-7 that batteries with higher capacity and maximum dis-/charging rate are more profitable in the retail market, and they also bring higher benefit to the inelastic customer. Therefore, it is concluded that the proposed retail market model also encourages the technology innovation made on improving the battery capacity and dis-/charging rate limit.

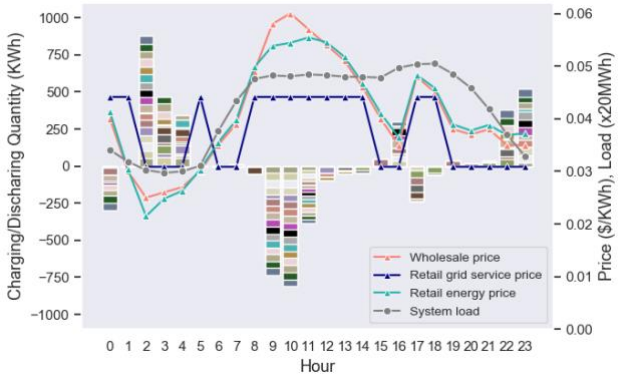
7.6.2 Sensitivity of the market parameter

Apart from battery relevant parameters mentioned above, another market relevant parameter C_β is also of the interest to be analyzed. C_β determines the sum of grid service prices over the 24-hour period. The RMO could apply C_β to control the average level of the retail grid service price signal. It impacts the market solution in a quite different way than battery-relevant parameters. Table 7-7 and Figure 7-8 present financial analyses and market solutions under different values of C_β.

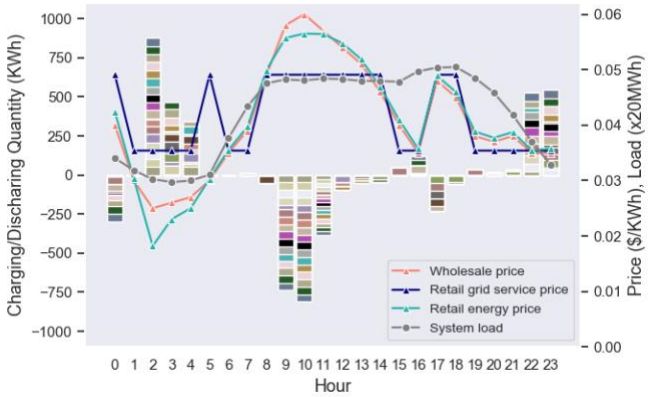
Table 7-7. Financial analyses under three different values of C_β

	ORIGINAL SYSTEM COST ($)	TOTAL ELECTRCITY BILL SAVING ($)	TOTAL GROSS REVENUE ($)	TOTAL DISCOUNTED REVENUE ($)
$C_\beta = 0.9$ $	866.40	15.98	44.93	15.98
$C_\beta = 1.0$ $	866.40	15.98	44.93	15.98
$C_\beta = 1.1$ $	866.40	15.98	44.93	15.98

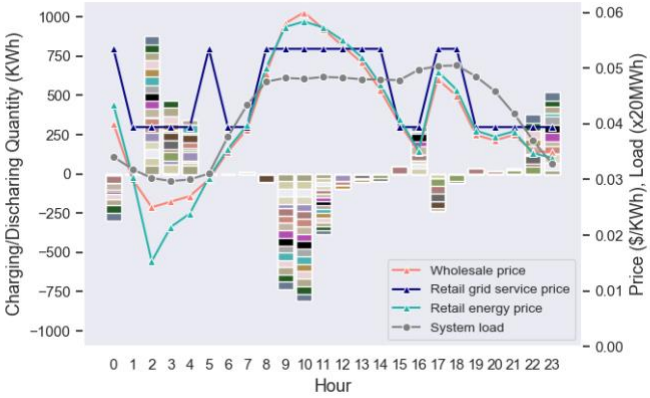
Based on the previous discussion, battery-relevant parameters affect the financial outcome of the retail market. However, identical financial settlements are achieved regardless of the value of C_β, as shown in Table 7-7. It is because retail grid service price signals are generated in the same shape in all three cases, yet with different average levels. It results in identical battery operation schedules. Nevertheless, by comparing light blue curves in Figure 7-8 (a)(b)(c), it is found that C_β does have an influence on shaping the retail energy price signal. While a smaller C_β tends to lower the retail energy price near peak wholesale price hours more significantly (because it saves the system payment used to compensate batteries' supply), a larger C_β tends to put more efforts on reducing the retail energy price near valley wholesale price hours (since a higher payment is given by batteries for purchasing the charged electricity). Consequently, a fattener retail energy price curve is generated with the incentive of a lower-averaged retail grid service price.



(a) $C_\beta = 0.9\ \$$



(b) $C_\beta = 1.0\ \$$



(c) $C_\beta = 1.1\ \$$

Figure 7-8. Sensitivity analysis of C_β

The above discussion concludes sensitivity analyses of key battery and market parameters, showing the advance of the proposed retail market model form following aspects:

1. The proposed retail pricing scheme is flexible and adaptive. It could actively adjust its pricing strategy to better incentivize batteries with unfavorable conditions. And the total social welfare created for the whole system will be maximized as a result of such an adjustment.
2. Technology improvements made on reducing the degradation rate, enlarging the battery capacity and increasing the maximum dis-/charging rate limit are encouraged by the proposed retail pricing scheme, as it brings more social welfare outcomes to the retail market.
3. The proposed retail pricing mechanism provides the flexibility to the RMO on adjusting the flatness of the retail energy price signal via customizing the C_{β}.

7.7 Proposed vs. alterative pricing schemes

Finally, the last part of the case study compares the proposed retail pricing mechanism with other possible pricing scheme alternatives. Based on the profitability of RMO and the number of price signals, retail market pricing schemes are classified into four categories. Such a division can be illustrated by a four-quadrant drawing presented in Figure 7-9, where the horizontal axis separates double-signal markets and single-signal markets, and the vertical axis separates markets managed by profit-oriented RMO and profit-neutral RMO. Plots located inside each quadrant illustrate the corresponding market solutions obtained under different retail pricing schemes. Table 7-8 summarizes corresponding financial outcomes, including total electricity bill saving, total gross/discounted revenue and profit made by the RMO. Note that quadrant I indicates the retail pricing mechanism proposed in this study. The light green curve in quadrant III represents the only retail price signal in the market. The result associated with profit-neutral and single-signal market pricing mechanism is not presented in Figure 7-9 and Table 7-8, because the embedded demand & supply balance and revenue & cost balance requirements in the market may be conflict for most cases.

By switching the profit-neutral RMO to profit-oriented RMO, the objective function in the pricing problem formulation changes from minimizing the system cost to maximizing the difference between system revenue and system cost. Instead of reducing the total system cost, which could be very inefficient due to the additional expense paid for distributed DERs, increasing the total system revenue by posing higher retail market prices on peak load hours would be a much easier choice for the profit-oriented RMO. This explains why the light blue and light green curves in quadrant II and III in Figure 7-9 reach the price cap (set as 0.06 $/kWh) during hours with highest loads and stays at 0.00 $/kWh during hours with lowest loads. According to Table 7-8, with regard to profit-oriented cases, although elastic customers can make certain amounts of revenue from the retail market, inelastic customers are actually sacrificed. They need to pay much higher electricity bills than the money they paid when directly purchasing demands from the wholesale market. It implies that when RMO's decisions are driven by its own profit, it will usually hurt the interest of inelastic customers and lower the system efficiency. Only in the proposed retail market inelastic and elastic customers are benefited at the same time. By further comparing profit-oriented retail pricing schemes with single and double price signals, it is discovered that when there are two price signals separately charge/pay elastic and inelastic customers, an even higher profit can be achieved by the RMO compared to the single signal case. It is because these two price signals are not financially binding, such that the RMO has a larger flexibility to specifically control two price

signals and maximize its benefit in each group. Such a flexibility, however, will be constrained when single price signal is applied.

In comparison with other alternative retail pricing schemes, the profit-neutral single-signal retail pricing scheme proposed in this study demonstrates the most promising properties, as it ultimately benefits both types of market participants and the system efficiency. Additionally, the retail energy price signal generated by the proposed retail pricing scheme is more dynamic, and is therefore better at representing the real system operating status across hours. It is suggested from the comparison that retail market designer should be careful about the profit-driven behavior of the RMO, as it may endanger the system efficiency and hurt market participants' interests. The RMO should not be given a full freedom on determining market prices. Regulations have to be enforced to limit the profitability of the RMO. When evaluating the performance of a retail market design, system efficiency promotion should be treated as a critical standard. The profit-neutral double-signal retail pricing scheme proposed in this study could serve as a great start for redefining the retail electricity market development.

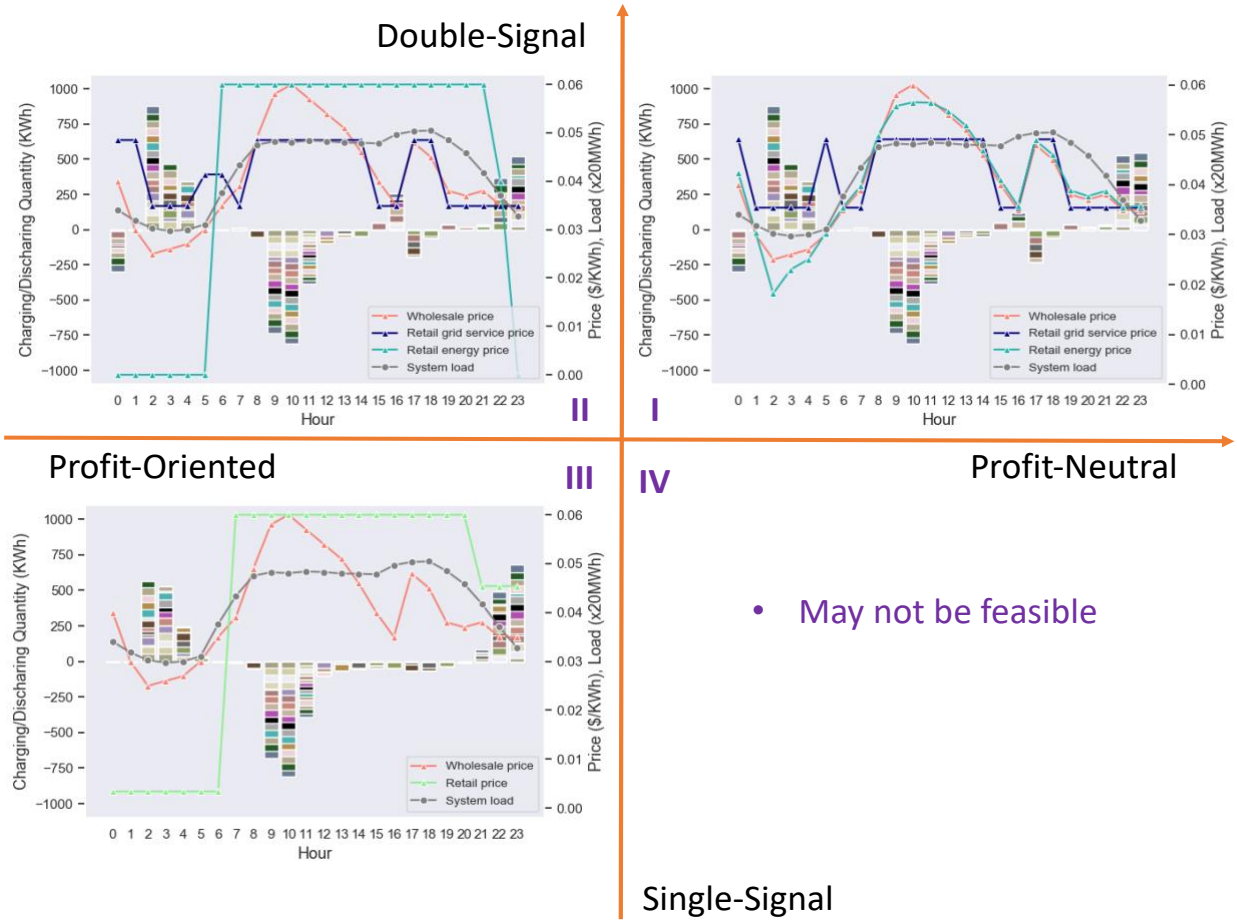


Figure 7-9. Four types of retail market and corresponding market solutions

Table 7-8. Financial analyses of four retail pricing mechanisms

	TOTAL ELECTRICITY BILL SAVING ($)	TOTAL GROSS/DISCOUNTED REVENUE ($)	RMO PROFIT ($)
PROFIT-NEUTRAL & DOUBLE-SIGNAL	15.98	44.93/15.98	0.00
PROFIT-ORIENTED & DOUBLE-SIGNAL	-65.55	28.48/0.00	97.38
PROFIT-ORIENTED & SINGLE-SIGNAL	-57.23	82.64/57.44	29.35
PROFIT-NEUTRAL& SINGLE-SIGNAL	None	None	None

8. SUMMARY, CONCLUSIONS AND FUTURE WORK

8.1 Summary

The objective of this dissertation is to provide a reliable and computational affordable retail market solution in the vision of future smart grid with a high proportion of price-responsive system behavior. To achieve such a goal, a systematic retail market design is carried out covering aspects of theoretical study, test-bed platform development and experimental analyses.

To begin with, a retail market model is proposed, which builds the foundation of following discussions. It follows a commonly applied three-layer hierarchy, with the RMO interacts with the wholesale market and retail market participants through physical and financial transactions. However, unlike existing works, market participants are divided into two groups: inelastic and elastic participants, based on their sensitivity to the market price. They are paid/charged by two different price signals: retail energy price and retail grid service price. Fixed loads and battery storages are selected in the study as representative inelastic and elastic participants. Load forecasting technique and price-responsive behavior modeling are investigated the next, as they offer necessary information for the RMO to understand how the fixed loads and battery storages behave under different market incentives. On the one hand, deep learning-based forecasting models are developed to predict day-ahead loads insensitive to the market price. Both recurrent and convolutional neural networks are utilized, and they are constructed under both recursive and direct multi-step manners. On the other hand, a battery arbitrage optimization problem is formulated to model battery's behaviors in response to the market price. It targets at maximizing battery's profit from the retail market while respecting equipment operational constraints. Based on the underlying market framework and dynamic behaviors of market participants, a retail pricing scheme is further proposed. It enables the RMO to provide optimized pricing incentive that ultimately minimizes the system operational cost. Under the incentive, battery storages are coordinated to schedule their day-ahead operations and maximize their payoff. Due to the interdependency between decisions made by the RMO and elastic participants, the retail pricing problem is formulated using a bi-level optimization technique. KKT conditions are utilized to merge the upper-level day-ahead system resources scheduling problem and lower-level battery arbitrage problems. After that, three techniques: respectively, a linearized three-phase power flow model, the strong duality theorem and the big-M method are deployed to linearize power flow constraints, the objective function as well as complimentary slackness conditions. As a result of efforts put on reformulating the problem, the retail pricing problem is finally converted into a mixed integer linear programming problem, which can be efficiently solved by off-the-shelf optimizers.

In addition to the theoretical study, a multi-agent test-bed system is developed using OpenDSS and Python, simulating both physical and financial operations of the grid and market. Three functional agents, including inelastic agent, elastic agent and RMO agent, are built to serve the load forecasting, battery arbitrage and retail pricing purposes. Communications are enabled among them to support necessary data exchanges. Finally, a comprehensive case study is conducted on the established test-bed system. The case study throughout analyzes the performance of the proposed retail market solution, and compares it with other possible alternatives.

8.2 Conclusions

Based on experimental results obtained across the dissertation, important conclusions are drawn from three aspects, together ensuring an efficient operation of the retail electricity market.

8.2.1 Deep learning models for day ahead load forecasting

According to performance analyses presented in Section 4.6, DL-based models proposed in this dissertation show an overall greater performance on predicting the day-ahead load portfolio compared to the traditional time-series model SARIMAX. They are able to properly capture the nonlinear time-dependency and autocorrelation between and within the load and outdoor temperature series. Among all DL-based models, the GCNN24 model performs the best. With regard to the prediction accuracy, compared to the SARIMAX model, the GCNN24 model helps decrease the day-ahead load forecasting error from 14.8 kW to 12.2 kW on average for datasets with strong weather covariance. With regard to the computational efficiency, the GCNN24 model is able to be trained 8% faster than the SARIMAX model to achieve an acceptable forecasting accuracy, due to the parallel data processing capability of CNN models. To sum up, the GCNN24 model can serve as a great start for deep learning-based network design tailored for day-ahead load forecasting.

8.2.2 Optimization-based battery arbitrage model for modeling the price-responsive behavior of batteries

According to experimental results illustrated in Section 5, the battery arbitrage model could smartly respond to different market prices to maximize its payoff. Under the arbitrage control, the battery charges at a lower price and discharges at a higher price to make profit. It is found that under the incentive of same price signal, batteries with lower degradation rate, larger capacity and higher dis-/charging rate limit could earn more grid service income. It is the price gap that drives the arbitrage behavior of batteries. Frequency and magnitude of the price fluctuation influence the profit of battery arbitrage differently. Whereas frequency impacts how often the battery cycles, a higher magnitude could more significantly improves the net profit of the battery. Therefore, the market operator could guide the battery to provide the most needed flexibility resource through properly controlling magnitude, frequency and location of the price gaps.

8.2.3 Profit-neutral double-signal retail market model and pricing scheme for optimal system operation

The retail market simulation conducted in Section 7 comprehensively evaluates the performance of proposed retail market model and pricing scheme. Several important conclusions are drawn from the simulation:

1. To begin with, motivations of both inelastic and elastic customers to participate in the retail market are verified. It is shown that flexibilities provided by elastic customers result in a lower retail energy price, saving the electricity bills for inelastic customers. Also, such a

price reduction is not at the expense of elastic customers' sacrifices, because as a return, elastic customers receive a revenue that combines grid service income and subsidy. As the number of elastic customers grows, higher benefits can be achieved by both inelastic and elastic customers.

2. Addition of the grid service subsidy is proven to be necessary. Firstly, it ensures the profitability of elastic customers despite the high degradation cost associated with batteries. Secondly, it is able to compensate the decrease on average revenue shared by individual elastic customer, when a severer competition is brought up by a larger number of elastic customers.
3. The ability of the proposed retail pricing scheme to optimize its pricing strategy while respecting network constraints is also shown in the study. Two tendencies of the retail market solution obtained under a full network consideration are observed. First is to reward discharging behaviors near peak load hours so as to better reduce line losses. Second is to discourage concurrent battery operations so as to avoid voltage violations.
4. Last but not least, sensitivity analyses are conducted to compare retail market solutions obtained under different battery conditions. The proposed retail market model is able to adaptively adjust its pricing strategy to better incentivize batteries with worse conditions, such that the system efficiency can still be guaranteed. Furthermore, the proposed retail market model encourages improvement of the battery technology, as batteries with lower degradation rate, larger capacity and higher dis-/charging rate limit are able to make more profits from the retail market.

To sum up, superiority of the proposed retail market model and retail pricing scheme has been proven. It provides a beneficial, practical, adaptive and supportive retail market solution that can serve as a promising start of future retail electricity market reconstruction.

8.3 Future work

Following future work are considered to further complete the retail market design that ready for real-world implementation.

1. This dissertation simulates an ideal retail market setting, where only homogeneous batteries (with identical parameters but various initial SOC and location) exist in the market as elastic participants. Therefore, the next step is to conduct simulations with heterogeneous batteries and to analyze how the retail pricing scheme incentivizes them differently.
2. Battery storage is modeled in this study as the representative price-responsive DER. It is worth to study and model other types of price-responsive DERs, e.g. shift-able loads and thermostatically controlled loads, to build a more realistic retail market environment. Note that, this will require extra works on formulating different groups of lower-level problems in addition to the battery arbitrage problem.
3. Current, the bi-level retail pricing problem is solved in a centralized fashion, which requires a full knowledge of RMO on parameters and initial operational status of batteries. Therefore, additional efforts are suggested to convert the centralized decision-making problem into a distributed decision marking problem to avoid unnecessary privacy exchanges.

4. Currently, price-responsive behaviors of elastic customers are modelled using an optimization-based approach. However, information required for modeling the optimization model may not always be available to the RMO. Also, models tailored for other price-responsive DERs other than the battery storage may not be linear and convex. Therefore, data-driven approaches for learning such price-responsive behaviors from observable historical energy and financial transaction data will be investigated in the future.
5. In this study, the RMO is assumed to be a price-taker in the wholesale market to emphasize interactions within the retail market. However, when the scale of the system goes up, the influence of retail market's operation towards wholesale market price cannot be ignored. Therefore, the RMO will be modeled as a price-maker of the wholesale market in the future to cover a more realistic application.
6. The proposed retail pricing mechanism is now formulated and solved as a deterministic optimization problem. However, many uncertainties exist in the retail market, including load forecasting errors, price forecasting errors, and deviations between the actual and modelled price-responsive behaviors. Existence of these uncertainties invalidates effectiveness of the deterministic approach. Therefore, a stochastic retail pricing mechanism will be studied in the future for handling these system uncertainties.

REFERENCES

[1] Bloomberg New Energy Finance, "Sustainable Energy in America 2019 Factbook," 2019.

[2] E. Litvinov, F. Zhao, and T. Zheng, "Electricity Markets in the United States: Power Industry Restructuring Processes for the Present and Future," *IEEE Power Energy Mag.*, vol. 17, no. 1, pp. 32–42, 2019.

[3] R. Adhikari, X. Zhang, M. Pipattanasomporn, M. Kuzlu, and S. Rahman, "A Data-driven Approach for Quantifying Energy Savings in a Smart Building," in *2017 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, 2017, pp. 1–5.

[4] X. Zhang, R. Adhikari, M. Pipattanasomporn, M. Kuzlu, and S. R. Bradley, "Deploying IoT devices to make buildings smart: Performance evaluation and deployment experience," *2016 IEEE 3rd World Forum Internet Things, WF-IoT 2016*, pp. 530–535, 2017.

[5] X. Zhang *et al.*, "Identification of key transmission lines in power grid using modified K-core decomposition," *2013 3rd Int. Conf. Electr. Power Energy Convers. Syst. EPECS 2013*, pp. 16–21, 2013.

[6] X. Zhang, M. Pipattanasomporn, T. Chen, and S. Rahman, "An IoT-Based Thermal Model Learning Framework for Smart Buildings," *IEEE Internet Things J.*, vol. 7, no. 1, pp. 518–527, 2020.

[7] H. Jiayi, J. Chuanwen, and X. Rong, "A review on distributed energy resources and MicroGrid," *Renew. Sustain. Energy Rev.*, vol. 12, no. 9, pp. 2472–2483, 2008.

[8] "U.S. Energy Information Administration." [Online]. Available: https://www.eia.gov/.

[9] U.S. EIA, "Annual Energy Outlook 2019 with projections to 2050," *Annu. Energy Outlook 2019 with Proj. to 2050*, vol. 44, no. 8, pp. 1–64, 2019.

[10] D. Distributed, E. Resources, and M. Sites, "Workshop Summary Report : R & D for Dispatchable Distributed Energy Resources at Manufacturing Sites."

[11] EIA, "U.S. Battery Storage Market Trends," *U.S. Dep. Energy*, no. May, pp. 1–26, 2018.

[12] "Bloomberg New Energy Finance." [Online]. Available: https://about.bnef.com/.

[13] "International Energy Agency." [Online]. Available: https://www.iea.org.

[14] M. Behl, F. Smarra, and R. Mangharam, "DR-Advisor: A data-driven demand response recommender system," *Appl. Energy*, vol. 170, pp. 30–46, 2016.

[15] T. Peffer, M. Pritoni, A. Meier, C. Aragon, and D. Perry, "How people use thermostats in homes: A review," *Build. Environ.*, vol. 46, no. 12, pp. 2529–2541, 2011.

[16] X. Zhang, M. Pipattanasomporn, and S. Rahman, "A self-learning algorithm for coordinated control of rooftop units in small- and medium-sized commercial buildings," *Appl. Energy*, vol. 205, no. June, pp. 1034–1049, 2017.

[17] M. Cai, S. Ramdaspalli, M. Pipattanasomporn, S. Rahman, A. Malekpour, and S. R. Kothandaraman, *Impact of HVAC Set Point Adjustment on Energy Savings and Peak Load Reductions in Buildings*. 2018.

[18] X. Zhang, M. Pipattanasomporn, M. Kuzlu, and S. R. Bradley, “Conceptual framework for a multi-building peak load management system,” in *IEEE PES Innovative Smart Grid Technologies Conference Europe*, 2016.

[19] X. Zhang, “A Data-driven Approach for Coordinating Air Conditioning Units in Buildings during Demand Response Events,” *Virginia Tech*, 2019. [Online]. Available: https://vtechworks.lib.vt.edu/handle/10919/87517.

[20] X. Zhang, M. Cai, M. Pipattanasomporn, and S. Rahman, “A Power Disaggregation Approach to Identify Power-Temperature Models of HVAC Units,” *2018 IEEE Int. Smart Cities Conf. ISC2 2018*, 2019.

[21] Z. Xu, R. Diao, S. Lu, J. Lian, and Y. Zhang, “Modeling of electric water heaters for demand response: A baseline PDE model,” *IEEE Trans. Smart Grid*, vol. 5, no. 5, pp. 2203–2210, 2014.

[22] B. Bhattarai, M. Maharjan, S. Hanif, M. Cai, and R. Pratt, “Transactive Electric Water Heater Agent: Design and Performance Evaluation,” pp. 1–5, 2020.

[23] P. Finn, M. O’Connell, and C. Fitzpatrick, “Demand side management of a domestic dishwasher: Wind energy gains, financial savings and peak-time load reduction,” *Appl. Energy*, vol. 101, no. 2013, pp. 678–685, 2013.

[24] D. Zhang, S. Li, M. Sun, and Z. O’Neill, “An Optimal and Learning-Based Demand Response and Home Energy Management System,” *IEEE Trans. Smart Grid*, vol. 7, no. 4, pp. 1790–1801, 2016.

[25] X. Guan, Z. Xu, and Q. S. Jia, “Energy-efficient buildings facilitated by microgrid,” *IEEE Trans. Smart Grid*, vol. 1, no. 3, pp. 243–252, 2010.

[26] N. Liu *et al.*, “A Heuristic Operation Strategy for Commercial Building Microgrids Containing EVs and PV System,” *IEEE Trans. Ind. Electron.*, vol. 62, no. 4, pp. 2560–2570, 2015.

[27] “Green Tech Media.”

[28] J. Zhong *et al.*, “Coordinated control for large-scale EV charging facilities and energy storage devices participating in frequency regulation,” *Appl. Energy*, vol. 123, pp. 253–262, 2014.

[29] H. Hui, Y. Ding, and M. Zheng, “Equivalent Modeling of Inverter Air Conditioners for Providing Frequency Regulation Service,” *IEEE Trans. Ind. Electron.*, vol. 66, no. 2, pp. 1413–1423, 2019.

[30] Q. Shi, F. Li, G. Liu, D. Shi, Z. Yi, and Z. Wang, “Thermostatic Load Control for System Frequency Regulation Considering Daily Demand Profile and Progressive Recovery,” *IEEE Trans. Smart Grid*, vol. 10, no. 6, pp. 6259–6270, 2019.

[31] H. Hui, Y. Ding, Y. Song, and S. Rahman, “Modeling and control of flexible loads for frequency regulation services considering compensation of communication latency and detection error,” *Appl. Energy*, vol. 250, no. April, pp. 161–174, 2019.

[32] Q. Shi, F. Li, Q. Hu, and Z. Wang, “Dynamic demand control for system frequency regulation : Concept review , algorithm comparison , and future vision,” *Electr. Power Syst. Res.*, vol. 154, pp. 75–87, 2018.

[33] B. M. Sanandaji, H. Hao, K. Poolla, and T. L. Vincent, “Improved battery models of an aggregation of Thermostatically Controlled Loads for frequency regulation,” *Proc. Am. Control Conf.*, pp. 38–45, 2014.

[34] J. Hughes, A. Dominguez-Garcia, and K. Poolla, “Coordinating Heterogeneous Distributed Energy Resources for Provision of Frequency Regulation Services,” *Proc. 50th Hawaii Int. Conf. Syst. Sci.*, pp. 2983–2992, 2017.

[35] S. Khan, M. Shahzad, U. Habib, W. Gawlik, and P. Palensky, “Stochastic battery model for aggregation of thermostatically controlled loads,” *Proc. IEEE Int. Conf. Ind. Technol.*, vol. 2016-May, pp. 570–575, 2016.

[36] O. Borne, M. Petit, and Y. Perez, “Provision of frequency-regulation reserves by distributed energy resources: Best practices and barriers to entry,” *Int. Conf. Eur. Energy Mark. EEM*, vol. 2016-July, pp. 1–7, 2016.

[37] E. D. Anese, S. S. Guggilam, A. Simonetto, Y. C. Chen, and S. V. Dhople, “Optimal Regulation of Virtual Power Plants,” *IEEE Trans. Power Syst.*, vol. 33, no. 2, pp. 1868–1881, 2018.

[38] P. Srivastava, C. Y. Chang, and J. Cortés, “Participation of Microgrids in Frequency Regulation Markets,” *Proc. Am. Control Conf.*, vol. 2018-June, pp. 3834–3839, 2018.

[39] Y. Zhu and K. Tomsovic, “Real-time control of distributed energy resources,” *IEEE PES Gen. Meet. PES 2010*, pp. 1–6, 2010.

[40] S. Li, K. Tomsovic, and T. Hiyama, “Load following functions using distributed energy resources,” *Proc. IEEE Power Eng. Soc. Transm. Distrib. Conf.*, vol. 3, no. c, pp. 1756–1761, 2000.

[41] X. Zhang, D. Biagioni, P. Graf, and J. King, “Cooperative Load Scheduling for Multiple Aggregators Using Hierarchical ADMM,” in *2020 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, 2020, no. 1, pp. 1–5.

[42] M. C. Bozchalui and R. Sharma, “Operation strategies for energy storage systems in distribution networks,” *IEEE Power Energy Soc. Gen. Meet.*, vol. 2014-Octob, no. October, pp. 1–5, 2014.

[43] F. S. Yanikara, P. Andrianesis, and M. Caramanis, “Strategic Behavior of Distributed Energy Resources in Energy and Reserves Co-Optimizing Markets,” *Proc. IEEE Conf. Decis. Control*, vol. 2018-Decem, no. Cdc, pp. 4875–4881, 2019.

[44] S. Bahramara, M. Yazdani-Damavandi, J. Contreras, M. Shafie-Khah, and J. P. S. Catalão, “Modeling the Strategic Behavior of a Distribution Company in Wholesale Energy and Reserve Markets,” *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3857–3870, 2018.

[45] S. R. Dabbagh and M. K. Sheikh-El-Eslami, “Risk Assessment of Virtual Power Plants Offering in Energy and Reserve Markets,” *IEEE Trans. Power Syst.*, vol. 31, no. 5, pp. 3572–3582, 2016.

[46] H. Hui, Y. Ding, W. Liu, Y. Lin, and Y. Song, “Operating reserve evaluation of aggregated air conditioners,” *Appl. Energy*, vol. 196, pp. 218–228, 2017.

[47] M. Caramanis, E. Ntakou, W. W. Hogan, A. Chakrabortty, and J. Schoene, “Co-optimization of power and reserves in dynamic T&D power markets with nondispatchable

renewable generation and distributed energy resources," *Proc. IEEE*, vol. 104, no. 4, pp. 807–836, 2016.

[48] H. Xu, X. Li, X. Zhang, and J. Zhang, "Arbitrage of Energy Storage in Electricity Markets with Deep Reinforcement Learning," vol. 1, pp. 1–3, 2019.

[49] H. Xu and A. D. Dom, "Optimal Tap Setting of Voltage Regulation Transformers Using Batch Reinforcement Learning," vol. 8950, no. c, pp. 1–12, 2020.

[50] C. Wu, G. Hug, and S. Kar, "Smart inverter for voltage regulation: Physical and market implementation," *IEEE Trans. Power Syst.*, vol. 33, no. 6, pp. 6181–6192, 2018.

[51] M. Ammar and A. M. Sharaf, "Optimized Use of PV Distributed Generation in Voltage Regulation: A Probabilistic Formulation," *IEEE Trans. Ind. Informatics*, vol. 15, no. 1, pp. 247–256, 2019.

[52] M. Zeraati, M. E. Hamedani Golshan, and J. M. Guerrero, "A Consensus-Based Cooperative Control of PEV Battery and PV Active Power Curtailment for Voltage Regulation in Distribution Networks," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 670–680, 2019.

[53] X. Wang *et al.*, "Optimal voltage regulation for distribution networks with multi-microgrids," *Appl. Energy*, vol. 210, no. August 2017, pp. 1027–1036, 2018.

[54] X. Zhou, Z. Liu, E. Dall'Anese, and L. Chen, "Stochastic dual algorithm for voltage regulation in distribution networks with discrete loads," *2017 IEEE Int. Conf. Smart Grid Commun. SmartGridComm 2017*, vol. 2018-Janua, pp. 405–410, 2018.

[55] Y. Zhang, A. Melin, M. Olama, S. Djouadi, J. Dong, and K. Tomsovic, "Battery energy storage scheduling for optimal load variance minimization," *2018 IEEE Power Energy Soc. Innov. Smart Grid Technol. Conf. ISGT 2018*, pp. 1–5, 2018.

[56] G. T. Costanzo *et al.*, "A coordination scheme for distributed model predictive control: Integration of flexible DERs," *2013 4th IEEE/PES Innov. Smart Grid Technol. Eur. ISGT Eur. 2013*, pp. 1–5, 2013.

[57] K. Mahmud, S. Morsalin, M. J. Hossain, and G. E. Town, "Domestic peak-load management including vehicle-to-grid and battery storage unit using an artificial neural network," *Proc. IEEE Int. Conf. Ind. Technol.*, pp. 586–591, 2017.

[58] F. Shen, S. Huang, Q. Wu, S. Repo, Y. Xu, and J. Ostergaard, "Comprehensive Congestion Management for Distribution Networks based on Dynamic Tariff, Reconfiguration and Re-profiling Product," *IEEE Trans. Smart Grid*, vol. 10, no. 5, pp. 4795–4805, 2018.

[59] W. Liu, Q. Wu, F. Wen, and J. Ostergaard, "Day-ahead congestion management in distribution systems through household demand response and distribution congestion prices," *IEEE Trans. Smart Grid*, vol. 5, no. 6, pp. 2739–2747, 2014.

[60] Z. Xu, P. Yang, Z. Zeng, J. Peng, and Z. Zhao, "Black start strategy for pv-ess multi-microgrids with three-phase/single-phase architecture," *Energies*, vol. 9, no. 5, p. 372, 2016.

[61] B. Zhang, P. Dehghanian, and M. Kezunovic, "Optimal Allocation of PV Generation and Battery Storage for Enhanced Resilience," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp.

535–545, 2019.

[62] I. K. Song, W. W. Jung, J. Y. Kim, S. Y. Yun, J. H. Choi, and S. J. Ahn, “Operation schemes of smart distribution networks with distributed energy resources for loss reduction and service restoration,” *IEEE Trans. Smart Grid*, vol. 4, no. 1, pp. 367–374, 2013.

[63] B. Chen, C. Chen, J. Wang, and K. L. Butler-Purry, “Sequential Service Restoration for Unbalanced Distribution Systems and Microgrids,” *IEEE Trans. Power Syst.*, vol. 33, no. 2, pp. 1507–1520, 2018.

[64] L. I. Minchala-Avila, L. Garza-Castanon, Y. Zhang, and H. J. A. Ferrer, “Optimal Energy Management for Stable Operation of an Islanded Microgrid,” *IEEE Trans. Ind. Informatics*, vol. 12, no. 4, pp. 1361–1370, 2016.

[65] W. Tushar *et al.*, “Three-Party Energy Management with Distributed Energy Resources in Smart Grid,” *IEEE Trans. Ind. Electron.*, vol. 62, no. 4, pp. 2487–2498, 2015.

[66] W. J. Ma, J. Wang, V. Gupta, and C. Chen, “Distributed energy management for networked microgrids using online ADMM with regret,” *IEEE Trans. Smart Grid*, vol. 9, no. 2, pp. 847–856, 2018.

[67] E. Foruzan, L. K. Soh, and S. Asgarpoor, “Reinforcement Learning Approach for Optimal Distributed Energy Management in a Microgrid,” *IEEE Trans. Power Syst.*, vol. 33, no. 5, pp. 5749–5758, 2018.

[68] M. B. C. Salles, J. Huang, M. J. Aziz, and W. W. Hogan, “Potential arbitrage revenue of energy storage systems in PJM,” *Energies*, vol. 10, no. 8, 2017.

[69] R. Sioshansi, P. Denholm, T. Jenkin, and J. Weiss, “Estimating the value of electricity storage in PJM: Arbitrage and some welfare effects,” *Energy Econ.*, vol. 31, no. 2, pp. 269–277, 2009.

[70] B. Cheng and W. B. Powell, “Co-Optimizing Battery Storage for the Frequency Regulation and Energy Arbitrage Using Multi-Scale Dynamic Programming,” *IEEE Trans. Smart Grid*, vol. 9, no. 3, pp. 1997–2005, 2018.

[71] F. Wankmüller, P. R. Thimmapuram, K. G. Gallagher, and A. Botterud, “Impact of battery degradation on energy arbitrage revenue of grid-level energy storage,” *J. Energy Storage*, vol. 10, pp. 56–66, 2017.

[72] A. Singh and P. W. Lehn, “Nonlinear reactive power control scheme to maximize penetration of distributed generation in distribution networks,” *2017 IEEE Electr. Power Energy Conf. EPEC 2017*, vol. 2017-Octob, pp. 1–6, 2018.

[73] V. Kekatos, L. Zhang, G. B. Giannakis, and R. Baldick, “Voltage regulation algorithms for multiphase power distribution grids,” *IEEE Trans. Power Syst.*, vol. 31, no. 5, pp. 3913–3923, 2016.

[74] L. Wang, F. Bai, R. Yan, and T. K. Saha, “Real-Time Coordinated Voltage Control of PV Inverters and Energy Storage for Weak Networks with High PV Penetration,” *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 3383–3395, 2018.

[75] L. Gutierrez-Lagos and L. F. Ochoa, “OPF-Based CVR Operation in PV-Rich MV-LV Distribution Networks,” *IEEE Trans. Power Syst.*, vol. 34, no. 4, pp. 2778–2789, 2019.

[76] Y. Zhang, N. Gatsis, and G. B. Giannakis, “Robust energy management for microgrids with high-penetration renewables,” *IEEE Trans. Sustain. Energy*, vol. 4, no. 4, pp. 944–953, 2013.

[77] A. D. Papalexopoulos and T. C. Hesterberg, “A regression-based approach to short-term system load forecasting,” *IEEE Trans. Power Syst.*, vol. 5, no. 4, pp. 1535–1547, 1990.

[78] D. H. Vu, K. M. Muttaqi, and A. P. Agalgaonkar, “Short-term load forecasting using regression based moving windows with adjustable window-sizes,” *Ind. Appl. Soc. Annu. Meet.*, pp. 1–8, 2014.

[79] C. M. Lee and C. N. Ko, “Short-term load forecasting using lifting scheme and ARIMA models,” *Expert Syst. Appl.*, vol. 38, no. 5, pp. 5902–5911, 2011.

[80] K. Yun, R. Luck, P. J. Mago, and H. Cho, “Building hourly thermal load prediction using an indexed ARX model,” *Energy Build.*, vol. 54, pp. 225–233, 2012.

[81] R. K. Jain, K. M. Smith, P. J. Culligan, and J. E. Taylor, “Forecasting energy consumption of multi-family residential buildings using support vector regression : Investigating the impact of temporal and spatial monitoring granularity on performance accuracy,” *Appl. Energy*, vol. 123, pp. 168–178, 2014.

[82] C. Fan, F. Xiao, and S. Wang, “Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques,” *Appl. Energy*, vol. 127, pp. 1–10, 2014.

[83] F. Yu and X. Xu, “A short-term load forecasting model of natural gas based on optimized genetic algorithm and improved BP neural network,” *Appl. Energy*, vol. 134, pp. 102–113, 2014.

[84] F. Javed, N. Arshad, F. Wallin, I. Vassileva, and E. Dahlquist, “Forecasting for demand response in smart grids : An analysis on use of anthropologic and structural data and short term multiple loads forecasting,” *Appl. Energy*, vol. 96, pp. 150–160, 2012.

[85] Z. Jing *et al.*, “Commercial Building Load Forecasts with Artificial Neural Network,” *2019 IEEE Power Energy Soc. Innov. Smart Grid Technol. Conf. ISGT 2019*, pp. 1–5, 2019.

[86] M. Duan, A. Darvishan, R. Mohammaditab, K. Wakil, and O. Abedinia, “A novel hybrid prediction model for aggregated loads of buildings by considering the electric vehicles,” *Sustain. Cities Soc.*, vol. 41, no. April, pp. 205–219, 2018.

[87] G. Suryanarayana, J. Lago, D. Geysen, P. Aleksiejuk, and C. Johansson, “Thermal load forecasting in district heating networks using deep learning and advanced feature selection methods,” *Energy*, vol. 157, pp. 141–149, 2018.

[88] X. Qiu, Y. Ren, P. N. Suganthan, and G. A. J. Amaratunga, “Empirical Mode Decomposition based ensemble deep learning for load demand time series forecasting,” *Appl. Soft Comput. J.*, vol. 54, pp. 246–255, 2017.

[89] A. Rahman, V. Srikumar, and A. D. Smith, “Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks,” *Appl. Energy*, vol. 212, no. October 2017, pp. 372–385, 2018.

[90] H. Shi, M. Xu, and R. Li, “Deep Learning for Household Load Forecasting – A Novel

Pooling Deep RNN," *IEEE Trans. Smart Grid*, vol. 3053, no. c, pp. 1–1, 2017.

[91] M. Binkowski, G. Marti, and P. Donnat, "Autoregressive Convolutional Neural Networks for Asynchronous Time Series," 2017.

[92] J. W. Taylor and R. Buizza, "Neural Network Load Forecasting With Weather Ensemble Predictions," *Trans. power Syst.*, vol. 17, no. 3, pp. 626–632, 2002.

[93] C. Fan, F. Xiao, and Y. Zhao, "A short-term building cooling load prediction method using deep learning algorithms," *Appl. Energy*, vol. 195, pp. 222–233, 2017.

[94] "Database of State Incentives for Renewables & Efficiency." [Online]. Available: https://www.dsireusa.org/.

[95] T. Roy and Z. Ni, "A Game Theoretic Approach for Distributed Electricity Providers in Deregulated Power Market," *IEEE Power Energy Soc. Gen. Meet.*, vol. 2018-Augus, pp. 1–5, 2018.

[96] W. Wei, F. Liu, and S. Mei, "Energy Pricing and Dispatch for Smart Grid Retailers Under Demand Response and Market Price Uncertainty," *IEEE Trans. Smart Grid*, vol. 6, no. 3, pp. 1364–1374, 2015.

[97] Y. Liu, J. Li, and L. Wu, "Coordinated Optimal Network Reconfiguration and Voltage Regulator/DER Control for Unbalanced Distribution Systems," *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 2912–2922, 2019.

[98] J. Campos Do Prado and W. Qiao, "A vision of the next-generation retail electricity market in the context of distributed energy resources," *2018 IEEE Power Energy Soc. Innov. Smart Grid Technol. Conf. ISGT 2018*, vol. 5, pp. 1–5, 2018.

[99] T. Chen, Q. Alsafasfeh, H. Pourbabak, and W. Su, "The next-generation U.S. retail electricity market with customers and prosumers-A bibliographical survey," *Energies*, vol. 11, no. 1, 2018.

[100] K. Subbarao, J. Fuller, K. Kalsi, R. Pratt, S. Widergren, and D. Chassin, "Transactive Control and Coordination of Distributed Assets for Ancillary Services," *Pacific Northwest Natl. Lab.*, no. September, 2013.

[101] J. Lian, L. Marinovici, W. Zhang, K. Kalsi, Y. Sun, and S. Widergren, "Transactive System Part I: Theoretical Underpinnings of Payoff Functions, Control Decisions, Information Privacy, and Solution Concepts," 2017.

[102] Q. Huang *et al.*, "Simulation-Based Valuation of Transactive Energy Systems," *IEEE Trans. Power Syst.*, vol. 34, no. 5, pp. 4138–4147, 2019.

[103] O. F. Power, "the Stability of Power System Markets," *IEEE Trans. Power Syst.*, vol. 14, no. 2, pp. 505–511, 1999.

[104] Z. Zhao, L. Wu, and G. Song, "Convergence of volatile power markets with price-based demand response," *IEEE Trans. Power Syst.*, vol. 29, no. 5, pp. 2107–2118, 2014.

[105] G. E. Asimakopoulou, A. L. Dimeas, and N. D. Hatziargyriou, "Leader-follower strategies for energy management of multi-microgrids," *IEEE Trans. Smart Grid*, vol. 4, no. 4, pp. 1909–1916, 2013.

[106] H. Xu, K. Zhang, and J. Zhang, "Optimal joint bidding and pricing of Profit-Seeking load serving entity," *IEEE Trans. Power Syst.*, vol. 33, no. 5, pp. 5427–5436, 2018.

[107] S. Maharjan, Q. Zhu, Y. Zhang, S. Gjessing, and T. Başsar, "Dependable demand response management in the smart grid: A stackelberg game approach," *IEEE Trans. Smart Grid*, vol. 4, no. 1, pp. 120–132, 2013.

[108] F. Lezama, J. Soares, P. Hernandez-Leal, M. Kaisers, T. Pinto, and Z. M. Almeida do Vale, "Local Energy Markets: Paving the Path Towards Fully Transactive Energy Systems," *IEEE Trans. Power Syst.*, vol. 34, no. 5, pp. 4081–4088, 2018.

[109] T. Chen and W. Su, "Local energy trading behavior modeling with deep reinforcement learning," *IEEE Access*, vol. 6, pp. 62806–62814, 2018.

[110] T. Chen and W. Su, "Indirect Customer-to-Customer Energy Trading with Reinforcement Learning," *IEEE Trans. Smart Grid*, vol. 10, no. 4, pp. 4338–4348, 2019.

[111] P. M. Sotkiewicz and J. M. Vignolo, "Nodal Pricing for Distribution Networks : Efficient," *Power*, vol. 21, no. 2, pp. 1013–1014, 2006.

[112] L. Bai, J. Wang, C. Wang, C. Chen, and F. Li, "Distribution Locational Marginal Pricing (DLMP) for Congestion Management and Voltage Support," *IEEE Trans. Power Syst.*, vol. 33, no. 4, pp. 4061–4073, 2018.

[113] R. Yang and Y. Zhang, "Three-phase AC optimal power flow based distribution locational marginal price," *2017 IEEE Power Energy Soc. Innov. Smart Grid Technol. Conf. ISGT 2017*, pp. 1–5, 2017.

[114] H. Yuan, F. Li, Y. Wei, and J. Zhu, "Novel linearized power flow and linearized OPF models for active distribution networks with application in distribution LMP," *IEEE Trans. Smart Grid*, vol. 9, no. 1, pp. 438–448, 2018.

[115] S. Hanif, M. Barati, A. Kargarian, H. B. Gooi, and T. Hamacher, "Multiphase Distribution Locational Marginal Prices: Approximation and Decomposition," *IEEE Power Energy Soc. Gen. Meet.*, vol. 2018-Augus, pp. 1–5, 2018.

[116] "OpenDSS." [Online]. Available: https://smartgrid.epri.com/SimulationTool.aspx.

[117] "GridLAB-D." [Online]. Available: https://www.gridlabd.org/.

[118] "PowerWorld." [Online]. Available: https://www.powerworld.com/.

[119] "CYMDIST." [Online]. Available: https://www.eaton.com/us/en-us/products/utility-grid-solutions/software-modules/distribution-system-analysis-package-CYMDIST.html.

[120] S. Hochreiter and J. Urgen Schmidhuber, "Long Short-Term Memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[121] Y. N. Dauphin, A. Fan, M. Auli, and D. Grangier, "Language Modeling with Gated Convolutional Networks," 2016.

[122] "BEMOSS (An open source platform for building energy management)." [Online]. Available: www.bemoss.org.

[123] "EnerNOC commerical building dataset." [Online]. Available: https://open-enernoc-data.s3.amazonaws.com/anon/index.html. [Accessed: 09-Oct-2018].

[124] "Weather underground." [Online]. Available: https://www.wunderground.com/. [Accessed: 09-Oct-2018].

[125] M. Abadi *et al.*, "TensorFlow: A system for large-scale machine learning," 2016.

[126] M. Koller, T. Borsche, A. Ulbig, and G. Andersson, “Defining a degradation cost function for optimal control of a battery energy storage system,” *2013 IEEE Grenoble Conf. PowerTech, POWERTECH 2013*, 2013.

[127] E. V. Thomas, I. Bloom, J. P. Christophersen, and V. S. Battaglia, “Rate-based degradation modeling of lithium-ion cells,” *J. Power Sources*, vol. 206, pp. 378–382, 2012.

[128] D. Wang, J. Coignard, T. Zeng, C. Zhang, and S. Saxena, “Quantifying electric vehicle battery degradation from driving vs. vehicle-to-grid services,” *J. Power Sources*, vol. 332, pp. 193–203, 2016.

[129] A. Bernstein, C. Wang, and E. D. Anese, “Load Flow in Multiphase Distribution Networks : Linear Models,” vol. 33, no. 6, pp. 5832–5843, 2018.

[130] M. Cai, M. Pipattanasomporn, and S. Rahman, “Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques,” *Appl. Energy*, pp. 1078–1088, Feb. 2019.