

# UC Riverside

## UC Riverside Electronic Theses and Dissertations

### Title

Data-Driven Modeling and Algorithmic Trading in Electricity Market

### Permalink

<https://escholarship.org/uc/item/4qm909wm>

### Author

Li, Yinglun

### Publication Date

2024

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA  
RIVERSIDE

Data-Driven Modeling and Algorithmic Trading in Electricity Market

A Dissertation submitted in partial satisfaction  
of the requirements for the degree of

Doctor of Philosophy

in

Electrical Engineering

by

Yinglun Li

March 2024

Dissertation Committee:

Dr. Nanpeng Yu, Chairperson  
Dr. Daniel Wong  
Dr. Weixin Yao

Copyright by  
Yinglun Li  
2024

The Dissertation of Yinglun Li is approved:

---

---

---

Committee Chairperson

University of California, Riverside

## Acknowledgments

First, I express my sincere gratitude to my Ph.D. advisor, Dr. Nanpeng Yu, for his excellent supervision, long-term support, and considerate patience. I am extremely grateful for his guidance, insight, and encouragement, which makes my research trip a wonderful adventure. His expertise, dedication, and persistence in scientific research and education facilitate me to reach a higher level in both work and life.

I greatly appreciate my committee members, Dr. Daniel Wong and Dr. Weixin Yao for their attendance and service. They have provided invaluable suggestions and comments that help me greatly in shaping my research work.

We thank all the collaborators from academia and industry for providing applicable data, consultations, and comments. I also thank all my labmates and friends for abundant brainstorm, helpful advice, and mutual support.

I am grateful to all professors and educators who have instructed me during my Ph.D. study. Without their instructions, I can hardly make progress in learning new knowledge.

Finally and most importantly, I would like to express my deep gratitude to my family, especially my parents, Liwen Li and Yi Li. They not only have raised me with all the best they have, but also respect my own decision and encourage me to chase my dream. I would not have been able to complete this work without their love, understanding, and unconditional support.

The content of this dissertation is a reprint of the materials that appeared in the following publications:

- Y. Li, N. Yu, and W. Wang, “Machine learning-driven virtual bidding with electricity market efficiency analysis,” in *IEEE Transactions on Power Systems*, vol. 37, no. 1, pp. 354-364, Jan. 2022. (Chapter 2)
- Y. Li and N. Yu, “Learning to arbitrage congestion in electricity market with virtual bids,” in *2021 IEEE ISGT Europe, 2021*, pp. 01-06. (Chapter 3)
- W. Wang, Y. Li, and N. Yu, “Predict locational marginal greenhouse gas emission factors of electricity with spatial-temporal graph convolutional networks,” in *2023 IEEE ISGT Europe*. (Chapter 4)

To my parents for all the support and love.

To my grandparents, in loving memory.

## ABSTRACT OF THE DISSERTATION

Data-Driven Modeling and Algorithmic Trading in Electricity Market

by

Yinglun Li

Doctor of Philosophy, Graduate Program in Electrical Engineering  
University of California, Riverside, March 2024  
Dr. Nanpeng Yu, Chairperson

The wholesale electricity markets in the United States operate under the two-settlement system, comprising the day-ahead (DA) market and the real-time (RT) market. The DA market clears bid-in supply against bid-in demand and determines DA physical schedules for generators, virtual awards, and DA locational marginal prices (LMPs), which are defined as the marginal costs of serving the next increment of demand at pricing nodes consistent with existing transmission constraints and performance characteristics of generation resources. The RT market procures "balancing" energy to meet the forecast RT grid energy demand and determines RT dispatch signals for resources and RT LMPs. The introduction of virtual bids to electricity markets is to mitigate the discrepancy between the DA market and the RT market in LMP spreads. In this dissertation, we focus on developing a machine learning-driven portfolio optimization framework for virtual bidding in electricity markets considering both risk constraint and price sensitivity. The algorithmic trading strategy is developed from the perspective of a proprietary trading firm to maximize profit. A recurrent neural network-based LMP spread forecast model is developed by leveraging the inter-hour dependencies of the market clearing algorithm. The LMP spread sensitivity with respect to net virtual bids is



modeled as a monotonic function with the proposed constrained gradient boosting tree. We perform a comprehensive empirical analysis on PJM, ISO-NE, and CAISO with the proposed framework. We further introduce a similar framework to arbitrage congestion with virtual bids. A deep neural network is designed to estimate the difference between congestion spreads in DA and RT markets. A clustering algorithm is adopted to separate pricing nodes into a few groups, between which the congestion spreads can be exploited. We validate the proposed algorithmic trading strategy using publicly available data from CAISO.

The electric power system is a major contributor to greenhouse gas (GHG) emissions. To reduce GHG emissions, accurate emission predictions are essential. The marginal emission factor (MEF) is a useful signal for distributed energy resource aggregators and end-use customers to mitigate GHG emissions by scheduling the flexible loads accordingly. We propose a hybrid machine learning framework to predict GHG emissions and locational MEF, which integrates feed-forward neural networks (FNNs) with spatio-temporal graph convolutional networks (STGCNs). A comprehensive case study in CAISO shows that the proposed approach outperforms the existing techniques in prediction accuracy. The proposed model provides short-term locational MEF predictions with high time granularity using only publicly available data.

# Contents

<b>List of Figures</b>	<b>xii</b>
<b>List of Tables</b>	<b>xiii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Background . . . . .	1
1.1.1 Wholesale Electricity Market . . . . .	1
1.1.2 Virtual Bidding . . . . .	5
1.1.3 Greenhouse Gas Emission . . . . .	6
1.2 Technical Challenges and Research Opportunities . . . . .	7
1.2.1 Machine Learning-Driven Virtual Bidding Strategy and Market Efficiency Analysis in Wholesale Electricity Market . . . . .	7
1.2.2 Prediction of Marginal Greenhouse Gas Emission Factors of Electricity with Graph Neural Networks . . . . .	8
1.3 Contributions . . . . .	9
1.4 Thesis Organization . . . . .	10
<b>2 Machine Learning-Driven Virtual Bidding with Market Efficiency Analysis in Electricity Market</b>	<b>11</b>
2.1 Introduction . . . . .	11
2.2 Virtual Bid Portfolio Optimization Problem with Price Sensitivity . . . . .	15
2.2.1 Modeling the Net Profits of Virtual Bids . . . . .	16
2.2.2 Virtual Bid Portfolio Optimization with Budget and Risk Constraints Considering Price Sensitivity . . . . .	16
2.2.3 Sensitivity of LMP Spreads with Respect to Virtual Bid Trading Quantities	18
2.2.4 Reformulation of Portfolio Optimization Problem . . . . .	20
2.2.5 Summary of the Portfolio Optimization Formulation . . . . .	21
2.2.6 Formulation of Risk Constrained Portfolio Optimization . . . . .	22
2.3 Data-Driven Forecasting for Virtual Bidding . . . . .	23
2.3.1 Neural Network-based LMP Spread Forecast . . . . .	23
2.3.2 Neural Network-based Virtual Bid Trading Quantity Forecast . . . . .	25
2.3.3 Constrained Gradient Boosting Tree-based Price Sensitivity Modeling . . . . .	25

2.4	Numerical Study . . . . .	29
2.4.1	Setup for Numerical Study . . . . .	29
2.4.2	Performance Comparison of LMP Spread Forecasting Algorithms . . . . .	31
2.4.3	Profitability of Algorithmic Trading Strategies Without Considering Price Sensitivity . . . . .	32
2.4.4	Impact of Portfolio Risk Limit on the Profitability of the Algorithmic Trading Strategy . . . . .	34
2.4.5	Profitability of Algorithmic Trading Strategies Considering Price Sensitivity . . . . .	35
2.4.6	Efficiency Analysis of Two-settlement Power Markets . . . . .	38
2.5	Conclusion . . . . .	41
2.6	List of Symbols in Chapter 2 . . . . .	42
<b>3</b>	<b>Learning to Arbitrage Congestion with Virtual Bids in Electricity Market</b>	<b>44</b>
3.1	Introduction . . . . .	44
3.2	Portfolio Optimization with Pairs of Virtual Supply Offers and Virtual Demand Bids	47
3.2.1	Modeling the Profits of Virtual Bid Pairs . . . . .	47
3.2.2	Virtual Bid Portfolio Optimization with Budget and Risk Constraints for Congestion Arbitrage . . . . .	48
3.2.3	Reformulation of Risk Constrained Portfolio Optimization . . . . .	49
3.3	Data-Driven Methodology for Congestion Arbitrage . . . . .	51
3.3.1	Cluster Analysis for Nodes Eligible for Virtual Bids . . . . .	51
3.3.2	DNN-based LMP Congestion Spread Forecast . . . . .	52
3.3.3	Allocation of Virtual Bid Quantity within Each Group . . . . .	53
3.4	Numerical Study . . . . .	54
3.4.1	Setup of Numerical Study . . . . .	54
3.4.2	Results of K-Means Clustering for Bidding Nodes . . . . .	55
3.4.3	Profitability of Algorithmic Virtual Bid Trading Strategy for Congestion Arbitrage . . . . .	55
3.5	Conclusion . . . . .	57
3.6	List of Symbols in Chapter 3 . . . . .	58
<b>4</b>	<b>Predict Locational Marginal Greenhouse Gas Emission Factors of Electricity with Spatial-Temporal Graph Convolutional Networks</b>	<b>60</b>
4.1	Introduction . . . . .	60
4.2	Problem Setup and Dataset Description . . . . .	64
4.2.1	Problem Setup . . . . .	64
4.2.2	Description of the Dataset . . . . .	64
4.3	Technical Methodology . . . . .	68
4.3.1	Overall Framework of the GHG Emission Prediction Model . . . . .	68
4.3.2	Data Preprocessing . . . . .	69
4.3.2.1	Preprocessing of Time Data . . . . .	69
4.3.2.2	Transformation of Data . . . . .	70
4.3.3	Brief Introduction of STGCN . . . . .	70
4.3.4	Load Graph Block . . . . .	72
4.3.5	Renewable Generation Graph Block . . . . .	73

4.3.6	Parallel Block . . . . .	73
4.3.7	Output Block . . . . .	74
4.3.8	Data Split and Hyperparameter Tuning . . . . .	74
4.4	GHG Emission Prediction Performance and Analysis of Marginal GHG Emission .	75
4.4.1	GHG Emission Prediction Performance . . . . .	75
4.4.2	Analysis of Locational Marginal GHG Emission . . . . .	76
4.4.2.1	Locational MEF on a Day with High Renewable Energy Output .	77
4.4.2.2	Weekday/Weekend Effect on Marginal GHG Emission . . . . .	78
4.5	Conclusion . . . . .	78
4.6	List of Symbols in Chapter 4 . . . . .	79
<b>5</b>	<b>Conclusions</b>	<b>80</b>
	<b>Bibliography</b>	<b>82</b>
<b>A</b>	<b>Appendix for Chapter 2</b>	<b>87</b>
A.1	Proof of (2.17) . . . . .	87

# List of Figures

1.1	Electricity Markets in the U.S. (Source: Federal Energy Regulatory Commission) . . . . .	2
2.1	Illustration of the piece-wise linear LMP sensitivity function. . . . .	19
2.2	Cumulative net profits of algorithmic trading strategy without considering price sensitivity. . . . .	33
2.3	Cumulative net profits of virtual bid trading strategy under different risk limits without considering price sensitivity. . . . .	35
2.4	Cumulative net profits of algorithmic trading strategy considering the impact of virtual bids on LMP when reporting profit and loss. “Full PS” refers to the full price sensitivity scenario and “partial PS” refers to the partial price sensitivity scenario. . . . .	36
2.5	Statistics of hourly price shift in absolute value. This shows the hourly change of price spreads averaged on each year considering the price sensitivity. The percentiles here are set to 5% and 95%. . . . .	37
2.6	Profitability of virtual bids with different market shares. . . . .	39
2.7	Sharpe ratios of algorithmic virtual bids trading portfolios. . . . .	40
3.1	Geographical information of bidding nodes in 13 groups. . . . .	56
3.2	Cumulative profits of the proposed virtual bid trading strategy in a 2-year period for CAISO. The daily budget limits correspond to 5.48%, 2.83%, and 1.44% of virtual bid market shares. The daily risk limit is set to be half of the portfolio budget. . . . .	57
4.1	Electric power interconnection map . . . . .	68
4.2	Framework of the Prediction Model . . . . .	69
4.3	The structure of an STGCN model. (a) is the overall structure of STGCN. (b) is the internal structure of an ST-Conv block. . . . .	71
4.4	Hourly MEF on March 27, 2022 for PGE-TAC and SCE-TAC. . . . .	77
4.5	Average hourly MEF on weekdays/weekends in PGE-TAC. . . . .	78

# List of Tables

2.1	Architecture Hyperparameters of Neural Networks . . . . .	30
2.2	LMP Convergence considering Price Sensitivity . . . . .	38
3.1	Top 3 Most Profitable Virtual Bid Pairs by Season . . . . .	58
4.1	Summary of Input and Output Data . . . . .	66
4.2	Summary of the Regions and Balancing Authorities of the Load Data . . . . .	67
4.3	Prediction Performance of GHG Emissions. (Average Value / Optimal Value) . . . . .	76

# Chapter 1

## Introduction

### 1.1 Background

#### 1.1.1 Wholesale Electricity Market

The electricity markets in the U.S. comprise wholesale and retail components, where the wholesale markets involve the sales of electricity among utilities and traders before it is distributed to consumers, and the retail markets involve the sales of electricity to consumers.

In wholesale electricity markets, there are traditional electricity markets and electricity markets managed by market operators. Traditional wholesale electricity markets exist primarily in the Southeast, Southwest, and Northwest (shown in Figure 1.1). The utilities in these markets are vertically integrated, in the sense that they are solely responsible for generating, transmitting, and distributing electric power to consumers. Wholesale physical power trade typically occurs via bilateral transactions, historically via power-pool agreements. In addition to traditional electricity markets, there are independent system operators (ISOs) and regional transmission organizations (RTOs),

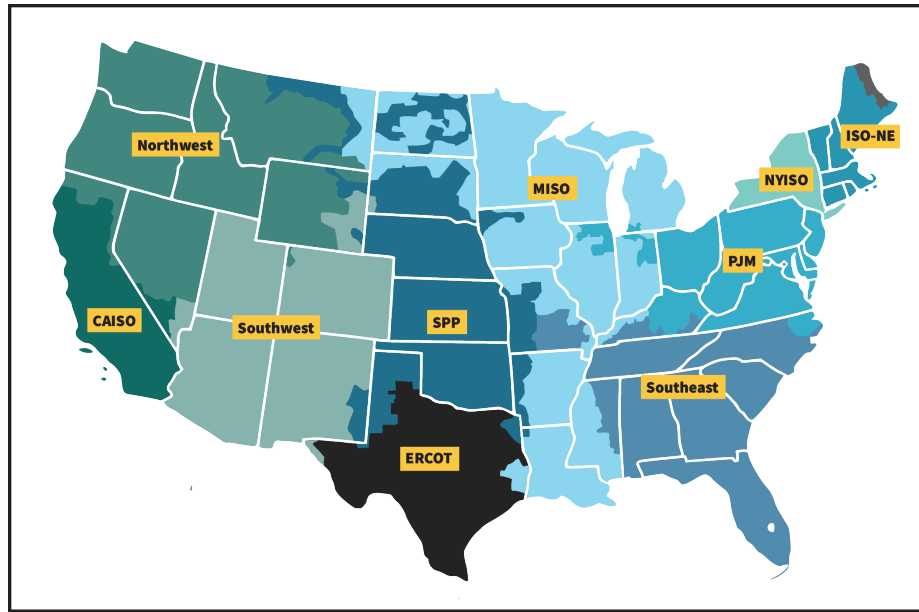


Figure 1.1: Electricity Markets in the U.S. (Source: Federal Energy Regulatory Commission)

which operate the transmission system independently of market participants and foster competition for electricity generation among market participants. They are regional federally organized organizations that coordinate, control, and monitor the power system. Each of the ISOs and RTOs have energy markets and ancillary service markets, which enable market participants to submit demand bids or supply offers for electricity. Such bid-based markets are used to determine the economic dispatch. Although major sections of the country operate under more traditional market structures, two-thirds of the nation's electricity load is served in regions managed by market operators. As shown in Figure 1.1, there are 7 market operators in the U.S., namely the California ISO (CAISO), the Midcontinent ISO (MISO), the ISO of New England (ISO-NE), the New York ISO (NYISO), the Southwest Power Pool (SPP), the Pennsylvania-New Jersey-Maryland Interconnection (PJM), and the Electric Reliability Council of Texas (ERCOT).



CAISO was founded in 1998 and became a fully functioning ISO in 2008. CAISO centrally dispatches generation and coordinates the movement of wholesale electricity in California and a portion of Nevada. CAISO's markets include energy (day-ahead and real-time), ancillary services, and congestion revenue rights. CAISO also operates an energy imbalance market (EIM), which currently includes CAISO and other balancing authority areas in the western United States.

MISO was organized as an ISO in December 2001. It began market operations in April 2005. In 2009, MISO started operating an ancillary services market and combined its 24 separate balancing areas into a single balancing area. In 2013, the ISO began operating in the MISO South region, including the utility footprints of Entergy, Cleco, and the South Mississippi Electric Power Association. Now, MISO centrally dispatches the market in portions of 15 states in the Midwest and the South, extending from Michigan and Indiana to Montana and from the Canadian border to the southern extremes of Louisiana and Mississippi.

ISO-NE was created in 1997 and began operating as an ISO in 2005. ISO-NE operates the transmission network and performs long-term planning and maintenance for the New England system, now serving 6 New England states: Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont.

NYISO was authorized as an ISO in December 1999. The chronic transmission in NY-ISO are in the southeastern part of the New York state, covering New York City and Long Island. NYISO is responsible for operating wholesale electricity markets that trade electricity, capacity, transmission congestion contracts, and related products.

PJM was founded in 1927 as a power pool of three utilities serving customers in Pennsylvania and New Jersey. In 1956, with the addition of two Maryland utilities, it became the

Pennsylvania-New Jersey-Maryland Interconnection, after which its name “PJM” came. PJM became a fully functioning ISO in 1996 and introduced markets with bid-based pricing and locational marginal pricing (LMP) in 1997. PJM was designated an RTO in 2001. Now, PJM centrally dispatches generation and coordinates the movement of wholesale electricity in all or part of 13 states (Delaware, Illinois, Indiana, Kentucky, Maryland, Michigan, New Jersey, North Carolina, Ohio, Pennsylvania, Tennessee, Virginia, and West Virginia) and the District of Columbia.

SPP was founded as an 11-member tight power pool in 1941 and became an RTO in 2004. In 2014, SPP implemented its integrated marketplace, including a day-ahead energy market, a real-time energy market, and an operating reserve market. In 2015, SPP expanded its footprint incorporating the Western Area Power Administration – Upper Great Plains (WAPA-UGP) region, the Basin Electric Power Cooperative, and the Heartlands Consumer Power District. Now, SPP manages transmission in portions of 14 states: Arkansas, Iowa, Kansas, Louisiana, Minnesota, Missouri, Montana, Nebraska, New Mexico, North Dakota, Oklahoma, South Dakota, Texas and Wyoming.

ERCOT was formed in 1970 and was deregulated for its wholesale energy market in 1995 and everything else in 1999 to fully encourage competitions. ERCOT manages the power flow to 24 million customers in the state of Texas, representing approximately 90% of the electrical load in Texas. ERCOT operates as an energy-only market with real-time, day-ahead, and ancillary services, performs financial settlement for the competitive wholesale energy market, and administers retail for 7 million premises in competitive-choice areas.

In the United States, the wholesale electricity market consists of two types, the day-ahead (DA) market and the real-time (RT) market, which is officially termed as “two-settlement”. The

DA market is a forward market that determines unit commitment plans, generation unit dispatch schedules, electric load, virtual bid awards, and the DA locational marginal prices (LMPs). The RT market is a spot market that balances between real-time demand and supply and determines the RT resource dispatch schedules and the RT LMPs. The LMPs are the marginal costs to supply the next incremental demand at different locations, accounting for transmission constraints, performance characteristics of generation resources, and load patterns.

### **1.1.2 Virtual Bidding**

Virtual bids are financial positions in wholesale electricity markets, which neither require physical assets nor provide/consume physical energy. Virtual bids consist of two types: increment (INC) offers and decrement (DEC) bids, also referred to as virtual supply offers and virtual demand bids. INCs are financial instruments that enable virtual traders to sell energy in the DA market and purchase the same amount of energy back in the RT market, while DECs enable virtual traders to purchase energy in the DA market and sell the same amount of energy back in the RT market.

The goal of introducing virtual bids to the electricity market is to drive price convergence between DA and RT LMPs, hedge financial risks, and increase market liquidity. However, the real impact of virtual bidding is controversial [1]. Theoretically, the introduction of virtual bidding increases market efficiency and reduces price spreads between DA LMP and RT LMP [2, 3, 4, 5]. On the other hand, market manipulation [6, 7], inappropriate market designs such as modeling discrepancies [2], and virtual bidding on interconnections [8] can lead to inefficient market solutions.

According to the most recent CAISO market report [9], the net profits paid to virtual bidders in 2022 is increased to \$106M, from \$38M in 2021 and \$45M in 2020. Specifically, during the summer heat wave of 2022, market participants received more than \$36M in net revenues from

virtual demand, accounting for 93% of the net revenues from virtual demand in all 2022. Virtual bidding is designed to align DA and RT LMPs by financial arbitrage. The higher the net profit, the larger the price spread exists between the DA and RT LMPs. Such statistics indicates an increasingly huge opportunity in algorithmic trading with virtual bids in the wholesale electricity market.

### **1.1.3 Greenhouse Gas Emission**

Greenhouse gas (GHG) traps the heat and makes the world warmer. Human activities are responsible for almost all the increase in GHGs in the atmosphere in the long run. According to the U.S. Environmental Protection Agency [10], electricity accounts for 25% of the total GHG emissions, which is the second largest share of GHG emissions. The reason is that approximately 60% of electricity comes from burning fossil fuels, mainly coal and natural gas.

The emission of electricity is mainly from the generation process. GHGs are released during the combustion of fossil fuels, including coal, oil, and natural gas, to produce electricity. Coal combustion is more carbon-intensive than natural gas or petroleum. In details [10], coal use accounted for 59% of CO<sub>2</sub> emissions in the United States in 2021, while it accounted for only 23% of electricity generation. Natural gas use accounted for 37% of electricity generation in 2021, and petroleum use accounted for less than 1%. The remaining generation in 2021 came from non-fossil fuel sources, including nuclear (20%) and renewable energy sources (20%), which involve hydroelectric, biomass, wind, and solar. Most of non-fossil sources, such as nuclear, hydroelectric, wind, and solar, cannot emit GHGs.

There are various ways to diminish GHG emissions associated with electricity. First, increase the efficiency of the existing fossil-based generations by replacing them with less carbon-intensive fuels. Second, encourage and promote the use of renewable energy, such as wind, solar,

hydro, and geothermal. Third, increase the efficiency of consuming electricity from end-users by providing more energy-efficient equipment to homes, businesses, and industries. Last, increase the use of nuclear energy in the substitution of fossil fuels in generations.

However, in order to reduce GHG emissions, an accurate prediction on GHG emissions is required. Normally, two major GHG emissions factors are considered: the average emissions factor (AEF) and the marginal emissions factor (MEF). The AEF is calculated as the ratio of total GHG emissions to total power consumption. The MEF is the ratio of the change in GHG emissions to the change in power consumption. MEF is more helpful compared to AEF, since distributed energy resource aggregators and end-use customers can determine how much electricity should be consumed at different time slots of a day using MEF. MEF signals can be sent along with electricity price signals to flexible loads and other smart technologies of the residential, commercial and industrial customers [11]. Based on the MEF signals, flexible loads, such as electric vehicles (EVs), smart thermostats, and batteries can consume or charge less power during high MEF hours and more power during low MEF hours. MEF can also be considered in the new design of electricity pricing to help further reduce GHG emissions.

## **1.2 Technical Challenges and Research Opportunities**

### **1.2.1 Machine Learning-Driven Virtual Bidding Strategy and Market Efficiency Analysis in Wholesale Electricity Market**

The goal of this research is to develop a machine learning based trading strategy to maximize the net profit for a virtual bid trader while analyzing the market efficiency of different U.S. wholesale electricity markets.

There are four major gaps in existing works. First, none of the existing work explicitly models the impacts of virtual bidders' trading activities on electricity market prices [3, 4, 12, 13]. We hereby develop a machine learning-based estimation algorithm for the LMP spread sensitivity with respect to virtual bid trading quantities. Second, recognizing that the inter-hour operational constraints such as resource ramping constraints have a great influence on LMPs, we accommodate the inter-hour dependencies by adopting a recurrent neural network framework to further improve the existing feedforward neural network based LMP spread forecasting model [13]. Third, most prior works perform market efficiency analysis for one electricity market at a time [3, 4, 13]. Our research performs a large-scale empirical market efficiency analysis across multiple wholesale electricity markets (PJM, ISO-NE, and CAISO). Last, the congestion difference in pairs of bidding nodes is rarely explored. We aim to predict and arbitrage the difference between DA and RT LMP spreads in pairs of nodes instead of separate nodes as in the prior works, so that the system energy components can be canceled.

## **1.2.2 Prediction of Marginal Greenhouse Gas Emission Factors of Electricity with Graph Neural Networks**

The goal of this research is to develop a hybrid machine learning model, which integrates feedforward neural networks (FNNs) with spatio-temporal graph convolutional networks (STGCNs) [14], to accurately predict GHG emissions and the MEF.

There exist three research gaps in predicting GHG emissions. First, previous works have difficulty capturing the complex interactions between multiple load zones, hard to provide a highly-accurate locational MEF prediction. Second, The short-term hourly MEF predictions are rarely explored. Third, an applicable framework based only on publicly available data is missing.

### 1.3 Contributions

This dissertation studies the two targeted topics mentioned in Section 1.2. The main contributions are as follows.

- Constrained gradient boosting tree based price sensitivity modeling: we are the first to develop the constrained gradient boosting tree based algorithm to model the monotonic function representing the LMP spread sensitivity with respect to net virtual bids.
- Neural network based price forecast and trading volume forecast: we develop a neural network based virtual bid trading quantity forecasting model to predict the hourly difference between market-wide cleared quantities of INC and DEC bids.
- Virtual bid portfolio optimization framework considering risk constraints and price sensitivities: we establish a virtual bid portfolio optimization framework considering both risk constraints and price sensitivities, which is shown to be much more profitable than the version without price sensitivity modeling.
- Virtual bid portfolio optimization framework with congestion arbitrage: we develop a virtual bid portfolio optimization framework to arbitrage congestion in electricity markets, which completely eliminates the significant uplift costs, taking into account the geographical relations among trading nodes.
- Graph neural networks based marginal greenhouse gas emission factors forecast: we develop a hybrid machine learning model incorporating FNNs and STGCNs to accurately predict locational MEFs with hourly granularity using only publicly available information.

## **1.4 Thesis Organization**

The remainder of this dissertation is organized as follows. Chapter 2 elucidates the entire research work on machine learning-driven virtual bidding with market efficiency analysis in electricity market. Chapter 3 introduces the work on learning to arbitrage congestion with virtual bids in electricity market. Chapter 4 explains the work on predicting locational MEF of GHG with STGCNs. Chapter 5 concludes the dissertation.



## **Chapter 2**

# **Machine Learning-Driven Virtual Bidding with Market Efficiency Analysis in Electricity Market**

### **2.1 Introduction**

The wholesale electricity markets in the United States operate under the two-settlement system, comprising the day-ahead (DA) market and the real-time (RT) market. The DA market clears bid-in supply against bid-in demand and determines DA physical schedules, virtual awards, and DA locational marginal prices (LMPs), which are defined as the marginal costs of serving the next increment of demand at pricing nodes consistent with the existing transmission constraints and performance characteristics of generation resources. The RT market procures “balancing” energy to meet the forecast RT grid energy demand and determines RT dispatch signals and RT LMPs.

Electricity price forecasting is one of the most fundamental inputs to decision making problems for electric utilities and energy trading companies. A general review of the complexity of different electricity price forecasting models with an emphasis on strengths and weaknesses is provided in [15]. Widely-used DA and RT LMP forecasting models include auto-regressive integrated moving average (ARIMA) model and its variants [16, 17], Markov regime-switching (MRS) model and its variants [18, 19], and the deep neural network-based models [20].

Electricity markets in the U.S. have two types of bids: physical bids and virtual bids. Physical bids must be backed by physical generation assets, loads, or imports/exports. Virtual bids are financial positions that are not backed by physical assets and do not deliver or consume physical energy. There are two types of virtual bids: increment (INC) offers and decrement (DEC) bids, also known as virtual supply offers and virtual demand bids respectively. INC (DEC) bids sell (buy) energy in the DA market and buy (sell) the same amount of energy back in the RT market.

Virtual bids are introduced in the U.S. electricity markets to drive the price convergence between DA and RT LMPs [4, 21, 22, 23], hedge financial risks [1], and increase the market liquidity. The impact of virtual bidding on electricity market has been a controversial topic [1]. In theory, the introduction of virtual bidding increases market efficiency and reduces price spreads between DA LMP and RT LMP [2, 3, 4, 5]. However, market manipulation [6, 7] and inappropriate market designs such as modeling discrepancies [2] and virtual bidding on the interties [8] can lead to inefficient market solutions.

Researchers tested the hypothesis that the electricity market is efficient by showing if one can find a virtual bid trading strategy that consistently achieves returns in excess of average market returns on a risk-adjusted basis. By identifying profitable virtual bid trading strategy based

solely on historical prices, it has been shown that California Independent System Operator (CAISO) [3, 4], Pennsylvania-New Jersey-Maryland Interconnection (PJM), and New York Independent System Operator (NYISO) [12] markets fail the weak form of the market efficiency hypothesis test. Most of the prior work [3, 4, 12] underestimate the potential profitability of virtual bid trading strategies and overestimate the market efficiencies by limiting the available information for virtual traders to historical LMPs. Furthermore, they either neglected virtual bid transaction fees [12] or left out the uplift costs of virtual bids [3] which resulted in overestimation of profitability of virtual bid trading strategies. Our recent work addressed these two issues by developing a machine learning-based algorithmic trading strategy for virtual bidding [13], which uses publicly available information such as load forecasts, meteorological variables, renewable generation forecasts, fuel prices, and historical LMPs to forecast the price spreads between DA and RT LMPs. After accounting for virtual bid transaction fees, uplift costs, and publicly available information, it showed that the algorithmic trading strategy consistently achieved significant profits in the Independent System Operator New England (ISO-NE) market. In other words, the ISO-NE market fails the semi-strong form of market efficiency hypothesis test.

This chapter extends the prior works in three ways. First, none of the existing work explicitly models the impacts of virtual bidders' trading activities on electricity market prices [3, 4, 12, 13]. We hereby develop a machine learning-based estimation algorithm for the LMP spread sensitivity with respect to virtual bid trading quantities. This enables us to develop an algorithmic virtual bidding portfolio optimization framework considering the price sensitivity. By explicitly modeling the impacts of virtual bidding activities on LMPs, the profitability of virtual bid portfolios and market efficiencies can finally be evaluated with different virtual bid market shares. Second,

recognizing that the inter-hour operational constraints such as resource ramping constraints have a great influence on LMPs, we accommodate the inter-hour dependencies by adopting a recurrent neural network framework to further improve the existing feedforward neural network-based LMP spread forecasting model [13]. Third, most of the prior works perform market efficiency analysis for one electricity market at a time [3, 4, 13]. This chapter performs a large-scale empirical market efficiency analysis across multiple wholesale electricity markets (PJM, ISO-NE, and CAISO).

The unique contributions of this chapter are as follows:

- We develop a constrained gradient boosting tree-based algorithm to model the monotonic function representing the LMP spread sensitivity with respect to net virtual bids.
- A virtual bid portfolio optimization framework considering both risk constraints and price sensitivities is established, which is shown to be much more profitable than the version without price sensitivity modeling.
- We develop a neural network-based virtual bid trading quantity forecasting model to predict the hourly difference between market-wide cleared quantities of INC and DEC bids.
- A large-scale empirical market efficiency analysis is conducted for multiple U.S. wholesale electricity markets with respect to different market shares of virtual bids.

The remainder of this chapter is organized as follows. Section 2.2 formulates the virtual bid portfolio optimization problem with price sensitivities. Section 2.3 presents the machine learning-based forecasts for LMP spreads, virtual trading quantities, and the price sensitivity. The empirical study on three U.S. wholesale electricity markets is conducted in Section 2.4. Section 2.5 concludes the chapter.

## 2.2 Virtual Bid Portfolio Optimization Problem with Price Sensitivity

In this section, we formulate the virtual bid portfolio optimization problem with price sensitivity and risk constraint. Note that the virtual trader under consideration is not treated as a price taker. The net profits of virtual bids are modeled in 2.2.1. The virtual bid portfolio optimization with budget and risk constraints are presented in 2.2.2. The sensitivity of LMP spread with respect to virtual bid trading quantities is modeled in 2.2.3. The portfolio optimization problem is reformulated and summarized in Subsections II.D and II.E.

A proprietary trading company engages in virtual bidding activities in wholesale electricity markets through the following process. On a daily basis, the proprietary trading company needs to ensure that it has posted a sufficient amount of collateral in a bank account monitored by the market operator to cover its virtual bid positions. One day before the operating day, the proprietary trading company submits INC offers and DEC bids through the DA market. Then the market operator clears the DA market and returns the virtual bid awards and LMP results back to the proprietary trading company. On the operating day, the virtual bids' positions are automatically liquidated by the market operator in the real-time market, which does not involve any further decision making from the proprietary trading company.

Three modeling assumptions are made here. First, it is assumed that the INC offers and DEC bids are guaranteed to be cleared in the DA market, which can be achieved by setting the offer (bid) price to be the price floor (price cap) for INCs and (DECs). Second, we assume that the impact of the virtual bids on congestion patterns in the market is negligible. We ensure that this assumption holds by setting the maximum bid quantity of virtual bids at each node to be 1 MWh. Third, we assume that the spatial-temporal distribution of the LMP spreads does not change much over time.

### 2.2.1 Modeling the Net Profits of Virtual Bids

Let  $\lambda_{i,h}^{DA}$  and  $\lambda_{i,h}^{RT}$  denote the DA and RT LMP for node  $i$  at hour  $h$ . The price spread  $\lambda_{i,h}^{dif}$  for node  $i$  at hour  $h$  is defined as the difference between DA and RT LMP,  $\lambda_{i,h}^{dif} = \lambda_{i,h}^{DA} - \lambda_{i,h}^{RT}$ .  $\lambda_{i,h}^{bid,I}$  and  $\lambda_{i,h}^{bid,D}$  are bid prices of INC and DEC for node  $i$  at hour  $h$ . Note that INCs are cleared when  $\lambda_{i,h}^{bid,I} \leq \lambda_{i,h}^{DA}$  and DEC are cleared when  $\lambda_{i,h}^{bid,D} \geq \lambda_{i,h}^{DA}$ . The bidding costs associated with INCs and DEC are denoted as  $\gamma^I$  and  $\gamma^D$ . The bidding costs include uplift cost and transaction fee. The expected net profit of the INC offer ( $r_{i,h}^I$ ) and the DEC bid ( $r_{i,h}^D$ ) for node  $i$  at hour  $h$ , can be calculated as:

$$E[r_{i,h}^I] = E[(\lambda_{i,h}^{dif} - \gamma^I) \mathbb{1}(\lambda_{i,h}^{bid,I} \leq \lambda_{i,h}^{DA})] \quad (2.1)$$

$$E[r_{i,h}^D] = E[(-\lambda_{i,h}^{dif} - \gamma^D) \mathbb{1}(\lambda_{i,h}^{bid,D} \geq \lambda_{i,h}^{DA})] \quad (2.2)$$

In this chapter, it is assumed that INCs and DEC are submitted into the DA market with bidding prices that will guarantee their clearance. This can be achieved by setting the bid price to be the price floor (price cap) for INCs (DECs).

### 2.2.2 Virtual Bid Portfolio Optimization with Budget and Risk Constraints Considering Price Sensitivity

The objective of an energy trading company is to develop a portfolio of virtual bids, which maximizes its profit with limited risks. The LMPs will be impacted by the submitted virtual bids. Thus, the price sensitivity with respect to the virtual bid trading quantities should be considered in the portfolio optimization process. In some related work like [24], the trader is assumed to be a price-taker in order to comply with the assumption that its bids are guaranteed to be cleared. In fact, to ensure the virtual bids are cleared, the trader can either make its bids self-schedules or setting

bid price to the price floor or the price cap for supply offers or demand bids respectively. The price sensitivity assumption is more realistic and does not necessarily affect the bid clearance.

The trading quantities of INC ( $z_{i,h}^I$ ) and DEC ( $z_{i,h}^D$ ) for node  $i$  at hour  $h$  across the operating day are the decision variables. To alleviate the impact of virtual bids on the congestion patterns in the electricity market, we assume that  $z_{i,h}^I$  or  $z_{i,h}^D$  are binary variables, where 0 represents no virtual bids and 1 represents a 1 MWh of virtual bid.

The portfolio optimization problem of virtual bidding is formulated as follows:

$$\max_{\mathbf{z}} \sum_{h=1}^{24} \sum_{i=1}^N (z_{i,h}^I E[r_{i,h}^I(\mathbf{z}_h)] + z_{i,h}^D E[r_{i,h}^D(\mathbf{z}_h)]) \quad (2.3)$$

$$s.t. \quad \sum_{i=1}^N \sum_{h=1}^{24} (z_{i,h}^I \text{prox}_{i,h}^I + z_{i,h}^D \text{prox}_{i,h}^D) \leq \mathcal{B} \quad (2.4)$$

$$\sum_{h=1}^{24} \text{CVaR}_{\beta}(f_h(\mathbf{z}_h, \boldsymbol{\lambda}_h^{dif})) \leq \mathcal{C} \quad (2.5)$$

$$f_h(\mathbf{z}_h, \boldsymbol{\lambda}_h^{dif}) = - \sum_{i=1}^N (z_{i,h}^I r_{i,h}^I + z_{i,h}^D r_{i,h}^D) \quad (2.6)$$

where  $\text{prox}_{i,h}^I$  and  $\text{prox}_{i,h}^D$  are the collaterals required by market operators for placing INC and DEC bids.  $\mathcal{B}$  is the portfolio budget limit.  $\mathcal{C}$  is the portfolio risk limit. The conditional value-at-risk (CVaR) is used to quantify the financial risk of the virtual bid portfolio as in (2.5), where  $\beta$ ,  $\mathbf{z}_h$ ,  $\boldsymbol{\lambda}_h^{dif}$ , and  $f_h(\mathbf{z}_h, \boldsymbol{\lambda}_h^{dif})$  are the confidence level associated with CVaR, the vector of decision variables, the vector of LMP spreads, and the portfolio loss at hour  $h$  respectively.

By explicitly specifying the impacts of virtual bids submitted by the energy trading company on LMP spreads, we can rewrite  $E[r_{i,h}^I(\mathbf{z}_h)]$  and  $E[r_{i,h}^D(\mathbf{z}_h)]$  as

$$E[r_{i,h}^I(\mathbf{z}_h)] = E[(\lambda_{i,h}^{dif}(\mathbf{z}_h, \mathbf{u}_h) - \gamma^I) \mathbb{1}(\lambda_{i,h}^{bid,I} \leq \lambda_{i,h}^{DA})] \quad (2.7)$$

$$E[r_{i,h}^D(\mathbf{z}_h)] = E[(-\lambda_{i,h}^{dif}(\mathbf{z}_h, \mathbf{u}_h) - \gamma^D) \mathbb{1}(\lambda_{i,h}^{bid,D} \geq \lambda_{i,h}^{DA})] \quad (2.8)$$

where  $\mathbf{u}_h$  denotes the vector of aggregated virtual bids from the rest of the electricity market at hour  $h$ . In other words, the LMP spread is influenced by the virtual bidding activities of the energy trading company under consideration and the rest of the market participants.

### 2.2.3 Sensitivity of LMP Spreads with Respect to Virtual Bid Trading Quantities

The impacts of the virtual bidding activities from the energy trading company and the rest of the market participants on LMP spreads can be approximated as:

$$\lambda_{i,h}^{dif}(\mathbf{z}_h, \mathbf{u}_h) = \lambda_{i,h}^{dif}(x_h + y_h) \quad (2.9)$$

The difference between the energy trading company's INC and DEC trading quantities at hour  $h$  ( $x_h$ ) and that of the other market participants' ( $y_h$ ) are defined as:

$$x_h = \sum_{i=1}^N z_{i,h}^I - \sum_{i=1}^N z_{i,h}^D \quad (2.10)$$

$$y_h = \sum_{i=1}^N u_{i,h}^I - \sum_{i=1}^N u_{i,h}^D \quad (2.11)$$

where  $u_{i,h}^I$  and  $u_{i,h}^D$  are the aggregated INC and DEC bids of the rest of the market for node  $i$  at hour  $h$  respectively.

It is extremely difficult to estimate the impact of virtual bidding on individual node's price spread due to the lack of nodal virtual bid trading quantity. Thus, the impacts of virtual bidding on an individual node is approximated by the impacts on the system reference node:

$$\lambda_{i,h}^{dif}(x_h + y_h) \approx \lambda_{i,h}^{dif}(y_h) + [\lambda_{ref,h}^{dif}(x_h + y_h) - \lambda_{ref,h}^{dif}(y_h)] \quad (2.12)$$

Note that  $y_h$  in equation (2.12) is determined outside the energy trading companies' portfolio optimization problem and can be estimated with a machine learning model.



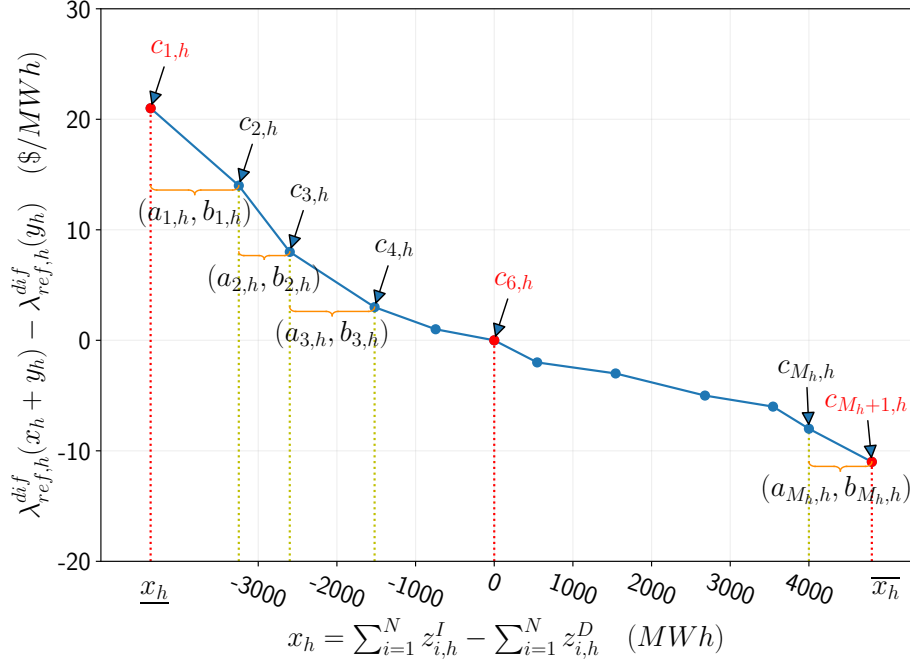


Figure 2.1: Illustration of the piece-wise linear LMP sensitivity function.

To make the portfolio optimization problem tractable, we model the impact of the energy trading company's virtual bids on the market reference LMP as a piece-wise linear function of  $x_h$  shown in Figure 1.

$$\lambda_{ref,h}^{dif}(x_h + y_h) - \lambda_{ref,h}^{dif}(y_h) = \sum_{j=1}^{M_h} (a_{j,h}x_h + b_{j,h})d_{j,h} \quad (2.13)$$

$$c_{j,h} - S(1 - d_{j,h}) \leq x_h \leq c_{j+1,h} + S(1 - d_{j,h}) \quad (2.14)$$

$$\sum_{j=1}^{M_h} d_{j,h} = 1 \quad (2.15)$$

$$\underline{x}_h \leq x_h \leq \bar{x}_h \quad (2.16)$$

Here  $\underline{x}_h$  and  $\bar{x}_h$  are the lower and upper bound of  $x_h$ , which are selected based on historical minimum and maximum hourly market-wide trading quantity of INCs minus DEC. Starting from the lower bound, we number the intervals in ascending order from  $j = 1$  to  $M_h$  which hits the

upper bound.  $c_{j,h}$  is the impact of virtual bids on market reference LMP at the starting point of  $j$ -th interval.  $S$  is a sufficiently large real number. Let  $a_{j,h}$  and  $b_{j,h}$  denote the slope and intercept of the linear function defined on  $j$ -th interval. The binary variable  $d_{j,h}$  indicates whether  $x_h$  belongs to  $j$ -th interval at hour  $h$ . Equation (2.14) enforces that  $x_h$  can only fall in a single interval. The parameters of the piece-wise linear function are derived from a gradient boosting tree (GBT) model, which will be discussed in detail in 2.3.3.

Note that  $a_{j,h}$  should be less than 0 for all intervals. This is because as the INC (DEC) trading quantity of the energy trading company increases, the hourly LMP spread at the reference node decreases (increases). We will describe how to enforce the piece-wise linear function to be monotonically decreasing by using a GBT model in 2.3.3. Finally, it should be noted that parameters of the piecewise linear function depend on  $y_h$ , the difference between INC and DEC trading quantities of the other market participants at hour  $h$  and other factors that affect market clearing outcomes.

## 2.2.4 Reformulation of Portfolio Optimization Problem

By substituting equations (2.7), (2.8), (2.9), (2.12), and (2.13) into (2.3), the objective function of the portfolio optimization problem can be rewritten as:

$$\sum_{h=1}^{24} \left\{ \sum_{i=1}^N z_{i,h}^I E[(\lambda_{i,h}^{dif} - \gamma^I)] + \sum_{i=1}^N z_{i,h}^D E[(-\lambda_{i,h}^{dif} - \gamma^D)] \right. \\ \left. + E\left[\sum_{j=1}^{M_h} (a_{j,h}x_h^2 + b_{j,h}x_h)d_{j,h}\right] \right\} \quad (2.17)$$

(See Appendix A.1 for proof)

To make the objective function concave, we introduce slack variables  $v_{j,h}$  and rewrite the equation (17) as:

$$\begin{aligned} & \sum_{h=1}^{24} \left\{ \sum_{i=1}^N z_{i,h}^I E[(\lambda_{i,h}^{dif} - \gamma^I)] \right. \\ & \left. + \sum_{i=1}^N z_{i,h}^D E[(-\lambda_{i,h}^{dif} - \gamma^D)] + E\left[\sum_{j=1}^{M_h} v_{j,h} d_{j,h}\right] \right\} \end{aligned} \quad (2.18)$$

$$s.t. \quad v_{j,h} \leq a_{j,h} x_h^2 + b_{j,h} x_h \quad \forall j, h \quad (2.19)$$

Note that the term  $E[\sum_{j=1}^{M_h} v_{j,h} d_{j,h}]$  still makes the objective function non-concave. The objective function is further relaxed by introducing additional slack variables  $w_{j,h}$  as:

$$\begin{aligned} & \sum_{h=1}^{24} \left\{ \sum_{i=1}^N z_{i,h}^I E[(\lambda_{i,h}^{dif} - \gamma^I)] \right. \\ & \left. + \sum_{i=1}^N z_{i,h}^D E[(-\lambda_{i,h}^{dif} - \gamma^D)] + E\left[\sum_{j=1}^{M_h} w_{j,h}\right] \right\} \end{aligned} \quad (2.20)$$

$$s.t. \quad -S d_{j,h} \leq w_{j,h} \leq S d_{j,h} \quad \forall j, h \quad (2.21)$$

$$-S(1 - d_{j,h}) \leq w_{j,h} - v_{j,h} \leq S(1 - d_{j,h}) \quad \forall j, h \quad (2.22)$$

## 2.2.5 Summary of the Portfolio Optimization Formulation

In summary, the virtual bid portfolio optimization problem can be formulated as follows:

$$max_z \quad (2.20)$$

$$s.t. \quad (2.4) - (2.6), (2.10), (2.11), (2.14) - (2.16), (2.19), (2.21), (2.22)$$

This is a mixed-integer quadratically-constrained programming problem and can be solved by optimization engines such as CPLEX[25].

## 2.2.6 Formulation of Risk Constrained Portfolio Optimization

In this work, we adopt CVaR as the risk measure for the virtual bid portfolio. It is chosen as the preferred risk measure because it is not only a coherent measurement of risk, but also accurately captures the tail distribution of portfolio loss function. In order to introduce CVaR, we need to first define value-at-risk (VaR)[26]. Let us first define the probability of the loss  $f_h(\mathbf{z}_h, \boldsymbol{\lambda}_h^{dif})$  at hour  $h$  not exceeding  $\alpha_h$  as:

$$\Psi(\mathbf{z}_h, \alpha_h) = \int_{f_h(\mathbf{z}_h, \boldsymbol{\lambda}_h^{dif}) \leq \alpha_h} p(\boldsymbol{\lambda}_h^{dif}) d\boldsymbol{\lambda}_h^{dif} \quad (2.23)$$

Here  $p(\boldsymbol{\lambda}_h^{dif})$  is the density function of LMP spread vector.  $\Psi(\mathbf{z}_h, \alpha_h)$  is the cumulative distribution function of the portfolio loss associated with decision vector  $\mathbf{z}_h$ . The  $VaR_\beta$  of the portfolio is the minimum portfolio loss such that the probability of having a smaller loss is  $\beta$ :

$$VaR_\beta(\mathbf{z}_h) = \min\{\alpha_h : \Psi(\mathbf{z}_h, \alpha_h) \geq \beta\} \quad (2.24)$$

The CVaR of the portfolio  $CVaR_\beta$  is defined as the expected loss given that the loss is no less than  $VaR_\beta$ :

$$\begin{aligned} CVaR_\beta(f_h(\mathbf{z}_h, \boldsymbol{\lambda}_h^{dif})) \\ = \frac{1}{1-\beta} \int_{f_h(\mathbf{z}_h, \boldsymbol{\lambda}_h^{dif}) \geq VaR_\beta(\mathbf{z}_h)} f_h(\mathbf{z}_h, \boldsymbol{\lambda}_h^{dif}) p(\boldsymbol{\lambda}_h^{dif}) d\boldsymbol{\lambda}_h^{dif} \end{aligned} \quad (2.25)$$

It has been proved that  $CVaR_\beta$  is upper bounded by the function  $F_\beta(\mathbf{z}_h, \alpha_h)$  [27]:

$$\begin{aligned} F_\beta(\mathbf{z}_h, \alpha_h) = \alpha_h \\ + \frac{1}{1-\beta} \int_{\boldsymbol{\lambda}_h^{dif}} [f_h(\mathbf{z}_h, \boldsymbol{\lambda}_h^{dif}) - \alpha_h]^+ p(\boldsymbol{\lambda}_h^{dif}) d\boldsymbol{\lambda}_h^{dif} \end{aligned} \quad (2.26)$$

In other words, CVaR can be represented as:

$$CVaR_\beta(f_h(\mathbf{z}_h, \boldsymbol{\lambda}_h^{dif})) = \min_{\alpha_h} F_\beta(\mathbf{z}_h, \alpha_h) \quad (2.27)$$

Under the assumption that the spatial-temporal distribution of the LMP spread does not change much over time,  $F_\beta(\mathbf{z}_h, \alpha_h)$  can be approximated by Monte Carlo sampling with historical LMP spread samples as:

$$F_\beta(\mathbf{z}_h, \alpha_h) = \alpha_h + \frac{1}{(1-\beta)N_s} \sum_{k=1}^{N_s} [f_h(\mathbf{z}_h, \boldsymbol{\lambda}_{h,k}^{dif}) - \alpha_h]^+ \quad (2.28)$$

where  $N_s$  is the number of historical LMP spread samples.

To remove the  $\max(0, x)$  function on the last term, (3.12) can be further relaxed as:

$$F_\beta(\mathbf{q}_h, \alpha_h) = \alpha_h + \frac{1}{(1-\beta)N_s} \sum_{k=1}^{N_s} q_h^k \quad (2.29)$$

$$s.t. \quad q_h^k \geq f_h(\mathbf{z}_h, \boldsymbol{\lambda}_{h,k}^{dif}) - \alpha_h \quad (2.30)$$

$$q_h^k \geq 0 \quad (2.31)$$

Then, the portfolio optimization problem in 2.2 can be reformulated as:

$$\max_{\mathbf{z}, \mathbf{q}, \alpha} \quad (2.20)$$

$$s.t. \quad \sum_{h=1}^{24} F_\beta(\mathbf{q}_h, \alpha_h) \leq \mathcal{C} \quad (2.32)$$

$$(3.5), (3.7), (2.10), (2.11), (2.14) - (2.16), (3.14), (3.15), (2.22), (3.13) - (2.31)$$

## 2.3 Data-Driven Forecasting for Virtual Bidding

### 2.3.1 Neural Network-based LMP Spread Forecast

The LMP spread between DA and RT market  $\lambda_{i,h}^{dif}$  is a highly nonlinear function of explanatory variables such as meteorological variables at key weather stations, fuel price forecasts, zonal load forecasts, and renewable generation forecasts. It has been shown that feedforward neural

networks such as multilayer perceptrons (MLP) and mixture density networks are quite effective in learning the nonlinear function [13]. In feedforward neural networks, the training samples at different hours are considered to be independent. However, the 24 hours of LMPs of the DA market are determined jointly via the security constrained unit commitment (SCUC) and security constrained economic dispatch (SCED) processes in practice. In particular, many inter-hour operational constraints such as resource ramping constraints are enforced in SCUC and SCED. For example, it is more likely to observe spikes in LMPs when the net-load of the hour and the increase in net-load from the previous hour are both very high.

To accommodate the inter-hour dependencies, we decide to adopt the long short-term memory [28] network, which is capable of learning long-term dependencies in the data. We use the cell state of LSTM to carry electricity market operation status information. LSTM employs three types of gates to control the information flow. The forget gate and input gate control which information should be discarded and added to the cell state. The output gate influences how the information in the cell state is used to predict the LMP spread.

In addition to the typical input feature normalization, we also need to perform special pre-processing for the target variables, i.e., the LMP spreads. Note that the LMP spreads are extremely volatile and have many spikes. If the output LMP spread is not scaled to flatten its distribution, then the LMP spikes will dominate the loss function of the neural network. This essentially makes all other training samples ineffective. To mitigate this problem, we leverage the parameterized sigmoid function  $f(x) = \frac{1}{1+e^{-x/\theta}}$  to scale the LMP spread to  $(0, 1)$ , with a hyper-parameter  $\theta$ . The activation functions are chosen to be hyperbolic tangent function “tanh” for the hidden layers and the sigmoid function for the output layer.

### 2.3.2 Neural Network-based Virtual Bid Trading Quantity Forecast

To estimate the impact of the energy trading company’s virtual bids on LMP spreads, we need to first forecast the aggregated virtual trading quantities  $y_h$  from the rest of the market participants. An MLP is adopted to address this problem. Similar to the LMP spread forecast model, the inputs to the MLP also consist of hourly market-wide features such as zonal load forecast, wind and solar generation forecast, meteorological variables, and one-hot encoding for trading hour. The output of the neural network is the hourly difference between market-wide cleared quantities of INC and DEC bids. These features can be found in the data archives maintained by the market operators. For this regression task, the typical input feature normalization and target scaling are required as well. As discussed in 2.3.1, we leverage a parameterized sigmoid function to scale the LMP spreads. Similarly, here we apply the sigmoid scaling function  $f(x) = \frac{1}{1+e^{-x/\theta_v}}$  to scale the target, which is the market-wide cleared virtual trading quantity of INCs minus DECs. The hyperparameter  $\theta_v$  is much greater than  $\theta$ , which is used in the scaling function of the LMP spread. This is because the range of the virtual bid trading quantity is much wider than that of the LMP spread.

### 2.3.3 Constrained Gradient Boosting Tree-based Price Sensitivity Modeling

We expand the XGBoost [29] method to model the monotonic piece-wise linear function representing the LMP spread sensitivity with respect to the net virtual bids. XGBoost is a variation of the gradient boosting tree method [30]. It is selected as the base model due to its scalability and capability to handle sparse datasets. The modification we make to the algorithm ensures that the learned LMP spread sensitivity function is monotonic.

As a supervised learning model, XGBoost searches in the space of regression trees to minimize a regularized objective. XGBoost addresses the optimization problem over the function space by additive training. This incremental approach, also called “tree boosting”, helps to learn the tree structure with the optimal score. Specifically, at the  $t$ -th iteration, a new tree  $f_t$  is added to optimize the objective:

$$\mathcal{L}^{(t)} = \sum_{i=1}^N l(\mathbf{y}_i, \hat{\mathbf{y}}_i^{(t-1)} + f_t(\mathbf{x}_i)) + \Omega(f_t)$$

where  $N$  is the number of training samples,  $\mathbf{x}_i$  is the  $i$ -th input,  $\mathbf{y}_i$  is the actual output,  $\hat{\mathbf{y}}_i$  is the predicted output, and  $\Omega(f_t)$  is the regularization term. On top of the inputs for the neural network-based LMP spread forecast model, we add the market-wide net INC offer quantity as an additional input for the XGBoost-based model. The outputs are the LMP spreads at the reference node.

By taking Taylor expansion at  $(\mathbf{y}_i, \hat{\mathbf{y}}_i^{(t-1)})$ , the second-order approximation of the objective can be derived. For a fixed tree structure  $q(\mathbf{x})$ , the optimal leaf weight  $w_j^*$  and objective can be derived as:

$$\mathcal{L}^{*(t)} = -\frac{1}{2} \sum_{j=1}^T \frac{G_j^2}{H_j + \lambda} + \gamma T$$

$$w_j^* = -\frac{G_j}{H_j + \lambda}$$

where  $G_j$  and  $H_j$  denote the sum of first-order derivatives  $g_i$  and the sum of second-order derivatives  $h_i$  of leaf  $j$  respectively.  $T$  is the total number of leaves in the tree. This score can be regarded as a quality measure of the tree structure at  $t$ -th iteration. When a new split is made, we can calculate the change of the score, also called the gain, as follows:

$$\mathcal{L}_{split} = \frac{1}{2} \left[ \frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right] - \gamma$$



Note that after a new split is made, the left leaf weight is not guaranteed to be higher than that of the right leaf. To ensure the learned LMP spread function is monotonically decreasing as a function of the net INC quantity, we modify the algorithm for finding split points in the XGBoost framework. The proposed greedy algorithm for finding splits that ensures monotonicity of the learned function is summarized in Algorithm 1.

The inputs to the algorithm include the dimension of the input feature space ( $D$ ), the index of the input feature involved with the monotonicity constraint ( $p$ ), and the combination set ( $I$ ) of indices of input samples assigned to each leaf. The indices assigned to leaf  $j$  is defined as  $I_j = \{i \mid q(\mathbf{x}_i) = j\}$ , where  $q(\mathbf{x}_i)$  represents the mapping function that assigns the  $i$ -th sample to the  $j$ -th leaf.

Starting from line 3, the algorithm searches through all  $D$  features for split candidates. For each feature  $k$ , the algorithm searches through all possible split points in an ascending order. If the feature is the one that is involved in the monotonicity constraint, then the weight of the left leaf must be higher than that of the right leaf, before the split's gain in objective is saved and recommended as a candidate split. Finally, the candidate split that satisfies the monotonicity constraint with highest gain is selected.

With a finite number of splits, the output of the modified XGBoost model is a step function. To convert this function into the piece-wise linear function in equation (13), we can simply connect the adjacent splitting points. The number of intervals  $M_h$  of the piece-wise linear function in equation (13) is determined by the final output of the modified XGBoost algorithm.

---

**Algorithm 1** Greedy Algorithm for Finding Splits for Learning Monotonically Decreasing Functions.

---

**Input:**  $I, D, p$

**Output:** Split with best gain and monotonicity constraint

```

1:  $gain \leftarrow 0$ 
2:  $G \leftarrow \sum_{i \in I} g_i, H \leftarrow \sum_{i \in I} h_i$ 
3: for  $k = 1$  to  $D$  do
4:    $G_L \leftarrow 0, H_L \leftarrow 0$ 
5:   for  $j$  in Ascending sort( $I$ , by  $x_{jk}$ ) do
6:      $G_L \leftarrow G_L + g_j, H_L \leftarrow H_L + h_j$ 
7:      $G_R \leftarrow G - G_L, H_R \leftarrow H - H_L$ 
8:     if ( $k == p$ ) then
9:       if ( $-\frac{G_L}{H_L + \lambda} \geq -\frac{G_R}{H_R + \lambda}$ ) then
10:         $gain \leftarrow$ 
11:           $max(gain, \frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda})$ 
12:        end if
13:      else
14:         $gain \leftarrow max(gain, \frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda})$ 
15:      end if
16:   end for
17: end for
18: return Split with the highest feasible gain

```

---

## 2.4 Numerical Study

### 2.4.1 Setup for Numerical Study

We validate the proposed algorithmic trading strategy and perform market efficiency analysis on three wholesale electricity markets in the US: Pennsylvania-New Jersey-Maryland Interconnection (PJM), California Independent System Operator (CAISO) and ISO New England (ISO-NE). Three years of historical data are collected for each market. The first year of data is used for initial model training and the last two years of data is used for rolling forecast. PJM and ISO-NE’s historical data ranges from January 2015 to December 2017. CAISO’s historical data ranges from July 2018 to June 2020. We perform rolling forecasts and update the forecast model on a monthly basis. The amount of training data is always kept at one year. The common inputs to deep neural networks and gradient boosting trees for all three electricity markets include load forecast, meteorological variables (temperature, humidity, wind speed, and precipitation), fuel price, and one-hot encoding for operating hour. The ISO-NE’s models include an additional input: wind generation forecast. The CAISO’s models include three extra inputs: estimated import, wind, and solar generation forecast. Furthermore, we conduct the forecasting task using only the common inputs for all three markets to show the impacts on forecasting performance from the additional input features for ISO-NE and CAISO.

The architectural hyperparameters of the MLP and LSTM models are summarized in Table 2.1. For the LSTM model, the first two numbers shown in the table represent the dimensions of two stacked LSTM layers. The other numbers in the list represent the dimensions of the remaining fully-connected layers. The first LSTM layer returns the sequence of all hidden states while the second LSTM layer simply returns a single output at the last time step. Both neural networks employ

Table 2.1: Architecture Hyperparameters of Neural Networks

Model Type	Market	Hidden Units
MLP	PJM	[128, 64, 32]
	ISO-NE	[64, 32]
	CAISO	[128, 64, 32]
LSTM	PJM	[64, 128, 128, 64, 32]
	ISO-NE	[32, 64, 64, 32]
	CAISO	[64, 128, 128, 64, 32]

the hyperbolic tangent function ( $\tanh$ ) as the activation function for the hidden layers. Dropout is introduced to regularize both neural networks. The dropout rate is selected to be 20%. The learning rate is set at 0.001. The Adam optimizer is used for both neural network models. The training batch sizes are 2048 for both LSTM and MLP models. Finally, the LMP spread scaling parameter  $\theta$  in the sigmoid function is set to be 20 for CAISO and 10 - 40 for PJM and ISO-NE depending on the price spread volatility of the node.

In this work, all the scripts are written in *Python 3.7*. The neural network models are based on *Tensorflow 1.15* software platform and *Nvidia RTX 2080 Ti* hardware platform. The portfolio optimization is based on *CVXPY 1.1* and the optimizer used is *Gurobi 9.1*.

## 2.4.2 Performance Comparison of LMP Spread Forecasting Algorithms

The LMP spreads between DA and RT markets quantify the potential net revenue of virtual bids without considering trading costs. Therefore, we forecast the LMP spreads directly instead of forecasting DA and RT LMPs separately. To better compare the performance of different LMP spread forecasting models, we introduce two evaluation metrics, which are tailored for the virtual bidding setup.

The first evaluation metric quantifies if the LMP spreads forecast led the virtual trader to place a virtual bid in the right direction. If the LMP spread forecast correctly forecasted that the RT LMP is higher than DA LMP, then the virtual trader will place a DEC bid that is profitable. The LMP spread forecast accuracy metric is formally defined as the proportion of the time that the LMP spread predictions have the correct sign. To better quantify the capability of the LMP spread forecasting algorithms to capture spikes that lead to massive profit, the forecasting accuracy is evaluated on spikes that are in the top 1th percentile of absolute LMP spreads. The second evaluation metric is the root mean square error (RMSE) of the LMP spread forecasts for the top 1th percentile of absolute LMP spreads.

We compared the LMP spread forecasting performance of the proposed MLP and LSTM models with a benchmark support vector regression machine (SVRM) [31]. The performance metrics are calculated by averaging the results on three electricity markets and across 2 years. The LMP spread spike forecast accuracies for SVRM, MLP and LSTM are 40.21%, 51.78%, and 56.18%. The RMSE for the LMP spread spike forecasts for SVRM, MLP and LSTM are \$217.35/MWh, \$216.15/MWh, and \$214.98/MWh. The LSTM-based LMP spread forecasting algorithm outperforms the MLP and SVRM-based approaches.

As mentioned in 2.4.1, we also conduct the forecasting task again using the common input features instead of different ones for all three markets using the LSTM models. The removal of additional input features for ISO-NE and CAISO lowers the LMP spread spike forecast accuracy by 3.96% and increases the RMSE for LMP spread forecast by 0.20%. This result shows that these additional input features are helpful in improving the LMP spread forecast performance.

### **2.4.3 Profitability of Algorithmic Trading Strategies Without Considering Price Sensitivity**

We first evaluate the profitability of algorithmic trading strategy without considering the impacts of the trading company’s virtual bids on LMP spreads. The evaluation is conducted over a 2-year period. The daily virtual bid portfolio budget limits in the 1st-year are set to \$600K, \$25K, and \$85K for PJM, ISO-NE and CAISO, respectively. In the 2nd-year, the daily budget limit for PJM is altered to \$330K, while the other two remain the same. The selected budget limits represent approximately 5% of the market share for virtual bids in the corresponding markets for that year. The virtual bid market share is defined as the portion of the market-wide cleared virtual trading quantity controlled by the proprietary trading company. The daily portfolio risk limit is set to be the same as the budget limit. As explained in 2.2.3, the upper and lower bound of  $x_h$  are selected based on historical minimum and maximum hourly market-wide bidding quantity of INCs minus DECs in the corresponding wholesale electricity markets. Specifically, for PJM, ISO-NE, and CAISO, the upper and lower bounds  $(\underline{x}_h, \overline{x}_h)$  in MWh are (-7812, 6821), (-331, 378), and (-3525, 3450) for the first testing year and (-8815, 5622), (-255, 435), and (-5647, 2798) for the second testing year.

The cumulative net profits of algorithmic trading strategy using MLP and LSTM models for the three electricity markets are depicted in Figure 2.2. As shown in the figure, the proposed al-

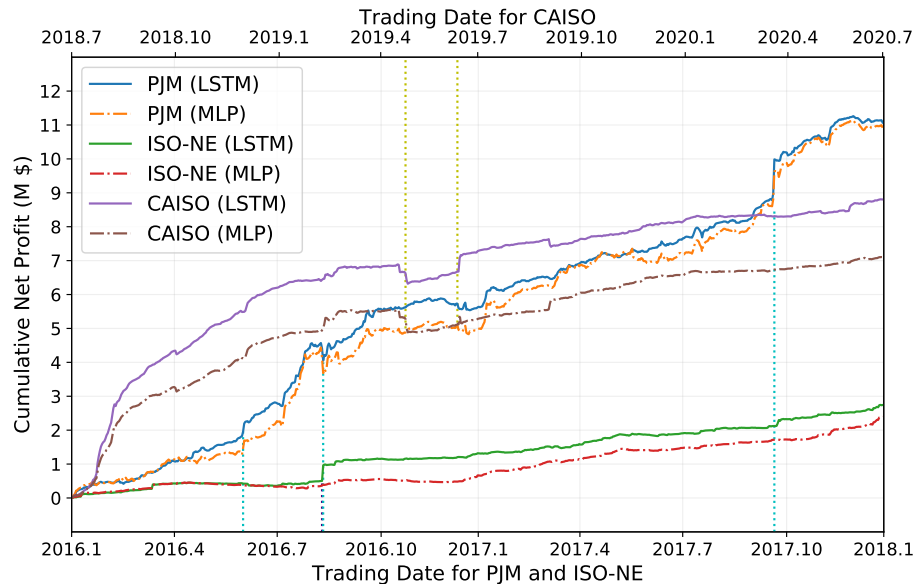


Figure 2.2: Cumulative net profits of algorithmic trading strategy without considering price sensitivity.

algorithmic trading strategy is very profitable in all three electricity markets when the price sensitivity is not considered. When the LSTM model is used to predict LMP spreads, the algorithmic trading strategy yields approximately \$11M, \$9M, and \$3M of cumulative net profits for PJM, CAISO, and ISO-NE in a 2-year period.

Figure 2.2 also shows that algorithmic trading strategy based on the LSTM model is much more profitable than that of the MLP model in CAISO market. In ISO-NE market, the LSTM algorithm slightly outperforms MLP. In PJM market, the cumulative net profit achieved by the LSTM-based and the MLP-based algorithmic trading strategies are roughly the same. In terms of net profit, the virtual bid portfolio derived from the LSTM-based LMP spread forecast outperforms MLP-based portfolio by 1.4%, 14.1% and 23.9% for PJM, ISO-NE, and CAISO respectively. The share of non-hydro renewable generation in CAISO, 27%, is much higher than that of PJM, 3%. During early spring and summer days, CAISO is much more likely to experience significantly faster

ramping in net-load than PJM. By considering the explanatory variables in the past few hours, the LSTM model is more capable of capturing the potential shortage in supply and LMP spikes than the MLP model in CAISO market.

Note that rare market events could lead to dramatic virtual bid portfolio gains and losses. In the PJM market, a significant gain of \$0.95M and a notable loss occurred on Sep. 20, 2017 and Aug. 12, 2016 respectively due to high temperature and peak load conditions. In ISO-NE, an unusual gain of \$0.65M took place on Aug. 11, 2016 due to severe generation capacity deficiency caused by thunder storms. In CAISO, notable portfolio gain and loss happened on Apr. 18 and Jun. 10, 2019 respectively. These two events are caused by the unforeseen and sharp drop in renewable generation.

#### **2.4.4 Impact of Portfolio Risk Limit on the Profitability of the Algorithmic Trading Strategy**

This subsection evaluates the impact of portfolio risk limit on the profitability of the algorithmic virtual bid trading strategy. We conduct the virtual bid portfolio optimization under three risk limit scenarios ranging from risk-averse to risk-neutral. In scenario 1, the portfolio risk level equals to one half of the portfolio budget limit. In scenario 2, the portfolio risk level equals to the portfolio budget limit. In scenario 3, we completely remove the risk limit, which makes the proprietary trading company risk neutral. The cumulative net profits of the virtual bid trading strategy under three risk limit scenarios are calculated based on LSTM model and depicted in Fig. 2.3.

As shown in Fig. 2.3, by focusing solely on potential gains regardless of the risk, the risk-neutral portfolio achieves notably higher cumulative net profits than both risk-averse portfolios for



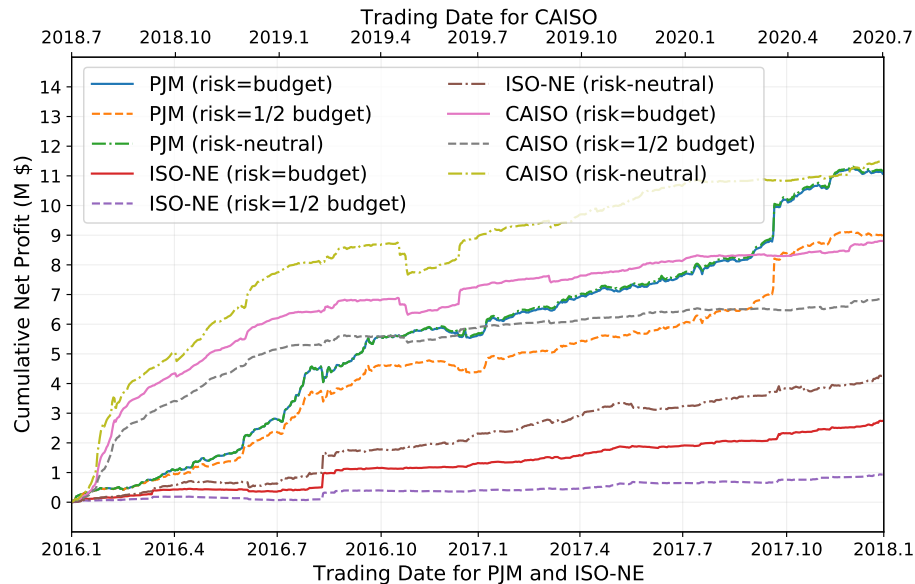


Figure 2.3: Cumulative net profits of virtual bid trading strategy under different risk limits without considering price sensitivity.

ISO-NE and CAISO. For PJM, when we increase the risk limit from one half of the portfolio budget to the portfolio budget, the cumulative net profit increases significantly. When we further relax the risk constraint by removing it, the improvement in net profit becomes negligible. This is because the budget constraint rather than the risk limit constraint is binding most of time for the portfolio optimization under scenario 2 in PJM. When we reduce the risk limit from the budget limit to one half of the budget limit, the cumulative net profits decrease by 18%, 60%, and 22% respectively for PJM, ISO-NE, and CAISO.

### 2.4.5 Profitability of Algorithmic Trading Strategies Considering Price Sensitivity

In this subsection, we quantify the profitability of algorithmic trading strategy considering price sensitivity. Here we analyze two scenarios. In both scenarios, when reporting net profit, the impact of the trading company's virtual bids on the LMP is taken into consideration. The

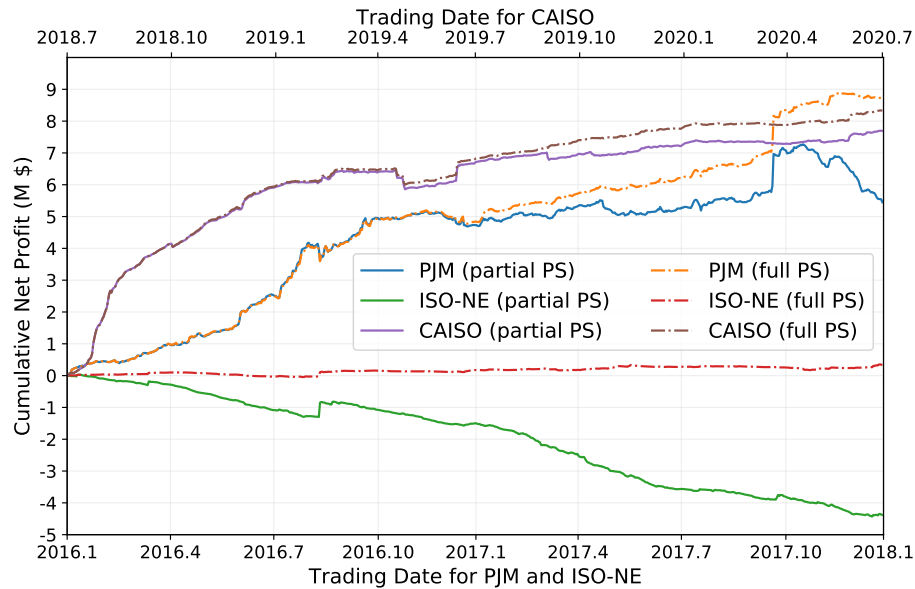


Figure 2.4: Cumulative net profits of algorithmic trading strategy considering the impact of virtual bids on LMP when reporting profit and loss. “Full PS” refers to the full price sensitivity scenario and “partial PS” refers to the partial price sensitivity scenario.

first scenario is called the full price sensitivity scenario, where the trading company considers the impacts of its virtual bids on LMP spread while solving the portfolio optimization problem. The second scenario is called the partial price sensitivity scenario, where the trading company does not consider price sensitivity while performing virtual bid portfolio optimization.

Figure 2.4 depicts the cumulative net profits of algorithmic trading strategy with LSTM model under the full and partial price sensitivity scenarios. By comparing the scenario without price sensitivity in Fig. 2 and the full price sensitivity scenario in Fig. 3, it can be seen that the trading company’s virtual bidding activity reduce a sizable portion of its algorithmic trading strategies’ cumulative net profits. The reductions in cumulative net profits in PJM, CAISO and ISO-NE are approximately \$2 Million, \$0.7 Million, and \$2.5 Million. By comparing the full and partial price sensitivity scenarios in Fig. 3, we conclude that the proposed virtual bids portfolio optimization

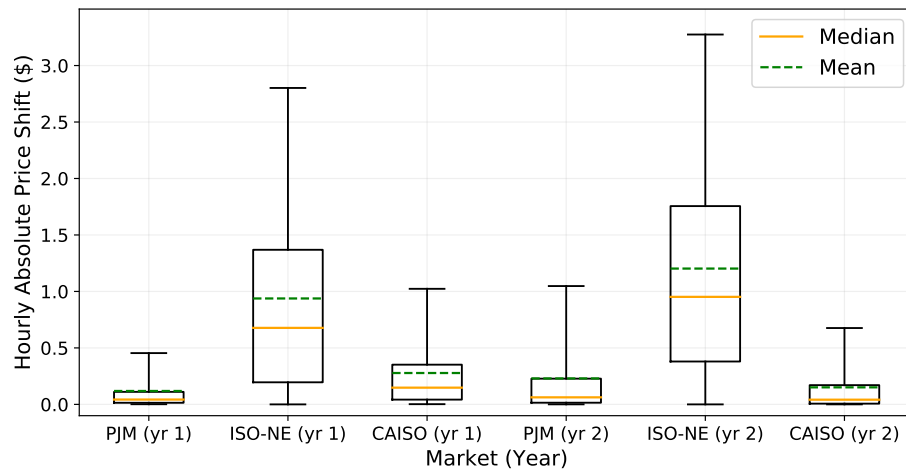


Figure 2.5: Statistics of hourly price shift in absolute value. This shows the hourly change of price spreads averaged on each year considering the price sensitivity. The percentiles here are set to 5% and 95%.

strategy that considers price sensitivity explicitly performs much better than the one that ignores price sensitivity. The differences in cumulative net profit between the full and the partial price sensitivity scenarios is the largest for ISO-NE (\$4 Million), followed by PJM (\$2.5 Million), and CAISO (\$0.9 Million). This is because the impacts of virtual bids on LMP spread is the largest in ISO-NE due to its small market size and large price sensitivity. Figure 2.5 shows the box plot of the changes in LMP spread due to the trading company’s virtual bids that correspond to 5% of virtual bid market share verifies the statement above. As shown in Fig. 2.5, the median LMP spread change in ISO-NE is much higher than that of PJM and CAISO.

Table 2.2 shows the convergence of LMP considering price sensitivity with 5% virtual bid market share. As shown in the table, the presence of virtual bid does lead to reduction in LMP spread between DA and RT markets.

Table 2.2: LMP Convergence considering Price Sensitivity

Market	Year	Average Absolute LMP Spread (\$/MWh)	
		without virtual bidding	with virtual bidding
PJM	yr 1	6.37	6.25
	yr 2	6.07	5.84
ISO-NE	yr 1	9.39	8.45
	yr 2	10.38	9.18
CAISO	yr 1	13.25	12.97
	yr 2	7.11	6.96

#### 2.4.6 Efficiency Analysis of Two-settlement Power Markets

In this subsection, we evaluate the efficiency of three wholesale power markets' two-settlement system by measuring the performance of virtual bid portfolio with different market shares. The first performance metric quantifies the cumulative net profit of virtual bids portfolio per dollar of collateral and risk limit. The second metric is the Sharpe ratio which is often used in finance to measure the performance of an investment portfolio. Specifically, Sharpe ratio measures the performance of an investment portfolio compared to a risk-free asset after adjusting for its risk. The Sharpe ratio  $S_p$  of an investment portfolio can be calculated as:

$$S_p = \frac{E[R_p - R_f]}{\sqrt{\text{var}[R_p - R_f]}}$$

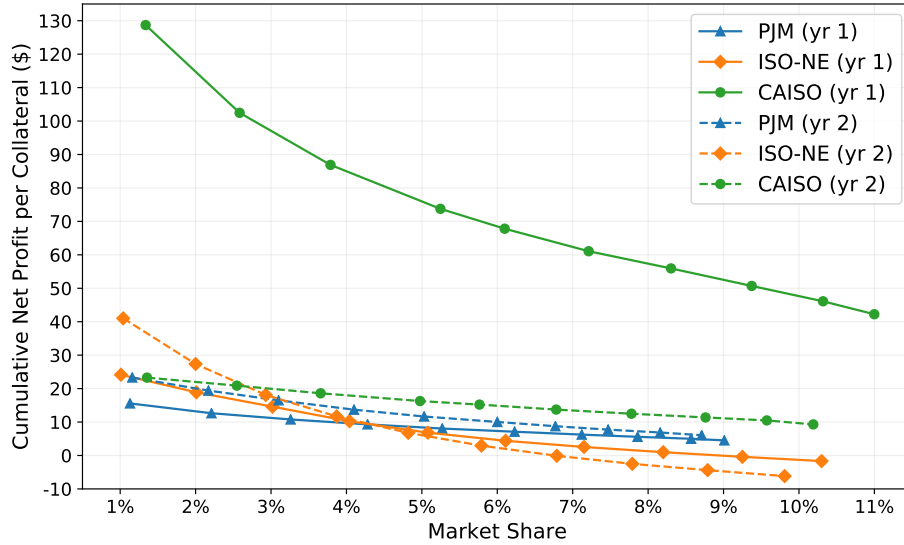


Figure 2.6: Profitability of virtual bids with different market shares.

where  $R_a$  is the portfolio's rate of return,  $R_f$  is the rate of return of the risk-free asset. If the algorithmic virtual bidding strategy yields a portfolio with higher performance, the two-settlement system of the corresponding wholesale market is expected to have lower efficiency.

To vary the market share of the virtual bidding portfolio, we choose different daily portfolio budgets. Note that the risk limit is set equal to the daily portfolio budget limit. From the perspective of the trading company, we assume that the virtual bid portfolio can take market shares from 1% to 10% with approximately 1% stepsize.

Figure 2.6 shows the annual cumulative portfolio net profit per dollar of budget and collateral from the three markets in two separate years with ten different market shares. It can be observed that the net profit per dollar decreases as the market share increases. This result can be explained in two aspects. First, after taken the most profitable bidding positions, only less profitable virtual bids can be identified with increased portfolio budget. Secondly, margins between DA and RT LMPs decrease as trading quantities increase, further slowing the increase in portfolio net profit.

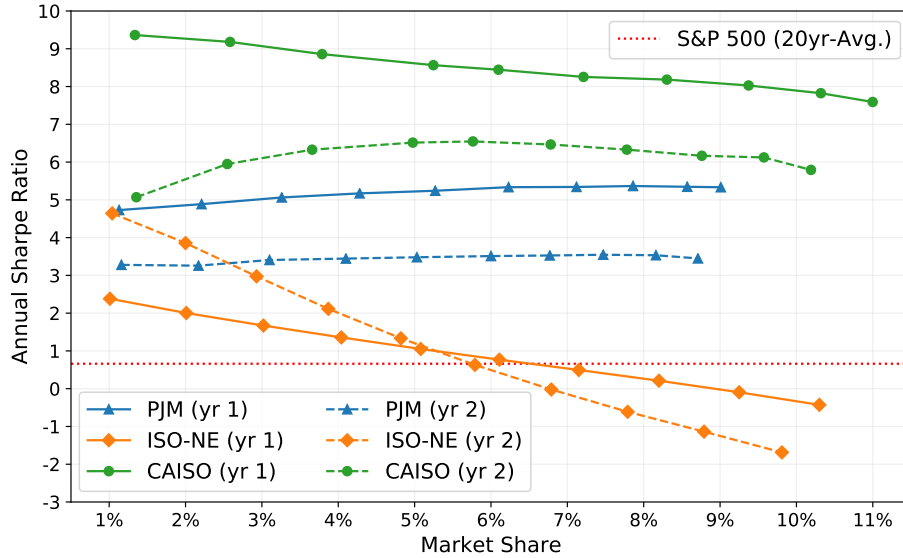


Figure 2.7: Sharpe ratios of algorithmic virtual bids trading portfolios.

CAISO’s two settlement system is shown to be the least efficient among the three wholesale markets, as the algorithmic trading strategy achieves the highest profitability in CAISO market during the second year of study. The LSTM model manages to forecast the LMP spreads in CAISO with 70% accuracy, compared with 58% and 60% for PJM and ISO-NE (accuracy here refers to the ratio of correctly-predicted directions of LMP spreads, either positive or negative). With 1% market share, the algorithmic trading strategy secured a profitability of up to \$41 and \$23 per dollar of collateral in ISO-NE and PJM. The reason profitability is the lowest in PJM is that it is the most competitive market with the largest number of virtual traders and trading volume.

To better understand the market efficiency and portfolio performance, we calculate the annual virtual bid portfolio’s Sharpe ratio with different market shares. In contrast with the first portfolio performance metric, the Sharpe ratio measures the performance of the virtual bids portfolio compared to a risk-free asset after adjusting for its risk.

Figure 2.7 illustrates the Sharpe ratios of virtual bid portfolio in three power markets and S&P 500 index for the corresponding years. With a wealth of historical data, we report the 20-year average Sharpe ratio for the S&P 500 index, which is a stock market index measuring the stock performance of 500 large companies listed on stock exchanges in the United States. The Sharpe ratio of virtual bid portfolios for CAISO and PJM are much higher than that of the S&P 500 index for all market shares. When the ISO-NE's virtual bid portfolio's market share is 5% or below, its Sharpe ratio is also higher than that of S&P 500 index. This indicates that the electricity markets' two settlement systems are in general much less efficient than the stock market. According to the Sharpe ratios, CAISO has the least efficient two settlement system among the three wholesale power markets. Note that the Sharpe ratio of PJM virtual bid portfolio is higher than that of ISO-NE. This result is different from the portfolio profitability curves shown in Figure 2.6. This is because our proposed algorithmic trading strategy captures the extremely high price spread between DA and RT LMPs on August 11, 2016, which results in a 1100% daily return. It significantly increases the total net profit, but reduces the Sharpe ratio which penalizes volatility in portfolio returns.

## **2.5 Conclusion**

This chapter develops an algorithmic virtual bid trading strategy that considers the impacts of virtual bids on LMPs. A constrained gradient boosting tree is developed to model the monotonic function representing the sensitivity of LMP spread. The risk-constrained virtual bid portfolio optimization problem is reformulated as a mixed-integer quadratically-constrained problem via convex relaxation. The results of comprehensive empirical studies on the three U.S. electricity markets show that the proposed virtual bid portfolio optimization framework considering price sensitivity

outperforms the one that ignores it. Among the three U.S. electricity markets, the proposed algorithmic virtual bid trading strategy achieves the highest profit in CAISO. The Sharpe ratios of virtual bid portfolios for PJM, ISO-NE, and CAISO are all significantly higher than that of S&P 500 index when the virtual bidder's market share is lower than 5%. Given the high uplift cost of net virtual supply in U.S. wholesale electricity markets such as CAISO, it would be interesting to develop algorithmic trading strategies to exploit the differences in congestion patterns between day-ahead and real-time markets.

## 2.6 List of Symbols in Chapter 2

$\lambda_{i,h}^{DA}, \lambda_{i,h}^{RT}$	DA LMP and RT LMP for node $i$ at hour $h$ (\$/MWh)
$\lambda_{i,h}^{dif}$	Price spread for node $i$ at hour $h$ (\$/MWh)
$\lambda_h^{dif}$	Vector of price spreads at hour $h$ (\$/MWh)
$\lambda_{i,h}^{bid,I}, \lambda_{i,h}^{bid,D}$	Bid prices of INC and DEC for node $i$ at hour $h$ (\$/MWh)
$r_{i,h}^I, r_{i,h}^D$	Net profits of INC and DEC for node $i$ at hour $h$ (\$)
$z_{i,h}^I, z_{i,h}^D$	Decision variables of INC and DEC for node $i$ at hour $h$ (MWh)
$z_h$	Vector of decision variables at hour $h$ (MWh)
$u_{i,h}^I, u_{i,h}^D$	Aggregated virtual bids of INC and DEC from the rest of the market for node $i$ at hour $h$ (MWh)
$u_h$	Vector of aggregated virtual bids from the rest of the market at hour $h$ (MWh)
$x_h$	Difference between the energy trading company's INC and DEC bidding quantities at hour $h$ (MWh)



$y_h$	Difference between the other market participants' INC and DEC bidding quantities at hour $h$ ( $MWh$ )
$\lambda_{ref,h}^{dif}$	Price spread on the system reference node at hour $h$ ( $$/MWh$ )
$d_{j,h}$	Binary variables indicating whether $x_h$ belongs to the $j$ -th interval at hour $h$
$v_{j,h}, w_{j,h}, \alpha_h, q_h^k$	Slack variables introduced in convex relaxation (\$)
$\gamma^I, \gamma^D$	Bidding costs of INC and DEC ( $$/MWh$ )
$prox_{i,h}^I, prox_{i,h}^D$	Collaterals required by market operators for placing INC and DEC bids for node $i$ at hour $h$ ( $$/MWh$ )
$\mathcal{B}$	Portfolio budget limit (\$)
$\mathcal{C}$	Portfolio risk limit (\$)
$a_{j,h}$	Slope of the linear function defined on the $j$ -th interval at hour $h$ ( $$/MWh2)$
$b_{j,h}$	Intercept of the linear function defined on the $j$ -th interval at hour $h$ ( $$/MWh$ )
$c_{j,h}$	Starting point of the $j$ -th interval at hour $h$ ( $MWh$ )

## **Chapter 3**

# **Learning to Arbitrage Congestion with Virtual Bids in Electricity Market**

### **3.1 Introduction**

Virtual bids are financial positions in wholesale electricity markets, which neither require physical assets nor deliver/consume physical energy. Virtual bids consist of two types: increment (INC) offers and decrement (DEC) bids, also referred to as virtual supply offers and virtual demand bids. INCs are financial instruments that enable virtual traders to sell energy in the DA market and purchase the same amount of energy back in the RT market, while DEC enables virtual traders to purchase energy in the DA market and sell the same amount of energy back in the RT market.

The objective of this study is to develop an algorithmic trading strategy for a virtual bid trading company. The goal of a virtual bid trading company is to maximize its profit while satisfying the portfolio risk constraints. Several studies aim to develop virtual bid trading strategies, including

an online learning algorithm using dynamic programming [12] and a neural network-based algorithmic trading framework using the mixture density network (MDN) [32]. The virtual bid trading costs (transaction and uplift costs) and key factors that influence LMPs such as load forecast, outage, and meteorological data are ignored in [12]. The estimated uplift costs are considered as a constant outside the virtual bid portfolio optimization in [32]. Our recent work addressed this issue by explicitly considering the estimated average uplift and transaction costs in the virtual bid portfolio optimization formulation [33]. However, the uplift costs associated with virtual supply and virtual demand are different and change with time. The actual uplift costs depend on the net virtual supply of the proprietary trading company and the entire market as well as the operating status of the power plants. Thus, it is extremely difficult to accurately forecast the uplift costs associated with virtual bids.

In some wholesale electricity markets, e.g., the California Independent System Operator market, the uplift costs of virtual bids become negligible if pairs of virtual supply and virtual demand bids are placed to arbitrage congestion spreads between DA and RT markets. To circumvent the difficulties of accounting for high and uncertain uplift costs, this work proposes an algorithmic trading strategy by exploiting the difference between congestion patterns in DA and RT markets.

Electricity price forecasting and congestion forecasting are both essential tasks for market participants. Congestion occurs when the least-cost energy cannot be entirely transmitted to load centers, due to the transmission constraints. Power system congestion is often accompanied by high shadow price for transmission constraints and large marginal congestion components in LMPs of nodes severely affected by the congestion.

Although the topic of electricity price forecasting has been well studied, very few researchers explored congestion forecasting. The widely used algorithms for LMP forecasting include the Markov regime-switching (MRS) model and its variants [18, 19], the auto-regressive integrated moving average (ARIMA) model and its variants [16, 17], and the neural network-based model [20]. The convex hull algorithm [34] and the Monte Carlo Simulation (MCS)-based approach with probabilistic models [35] are proposed to forecast the short-term congestion. Multi-parametric programming-based approach [36] is developed to forecast the network congestion together with RT LMPs. A factor model with least-square estimation [37] is designed for DA shadow price forecasting, which is useful for congestion management.

In this work, instead of forecasting congestion patterns or DA shadow prices, we directly estimate the difference between congestion patterns of DA and RT markets. In other words, we aim to predict the difference between DA and RT LMP spreads in pairs of nodes instead of separate nodes as in previous works. Due to the cancellation in system energy components of both DA and RT LMPs, our proposed forecasting model could focus on modeling the congestion components of LMPs. The main contributions of this chapter are as follows:

- We develop a virtual bid portfolio optimization framework to arbitrage congestion in electricity markets, which completely eliminates the significant uplift costs.
- We develop a machine learning framework to exploit the geographical relations among bidding nodes and estimate the congestion patterns of node pairs.

The remainder of the chapter is arranged as follows. Section 3.2 formulates the virtual bid portfolio optimization problem as a linear programming (LP) problem. Section 3.3 reveals the key technical methods, involving the clustering of bidding nodes, the forecasting of LMP congestion

spreads, and the allocation of virtual bid quantities. Section 3.4 shows the empirical results of clustering and the profit of virtual bid portfolio in CAISO. Section 3.5 summarizes the chapter.

## 3.2 Portfolio Optimization with Pairs of Virtual Supply Offers and Virtual Demand Bids

### 3.2.1 Modeling the Profits of Virtual Bid Pairs

Let  $G$  denote the network topology of the power. Formally,  $G = (\mathcal{V}, \mathcal{E})$ , where  $\mathcal{V}$  is the vertex set containing all network nodes, and  $\mathcal{E}$  is the edge set containing all transmission lines. A subset  $\tilde{\mathcal{V}} \subseteq \mathcal{V}$  is defined as a vertex set including all nodes eligible for virtual bidding, where  $|\tilde{\mathcal{V}}| = N$ . Let  $\mathcal{T}$  represent the set of 2-permutations of  $N$  from  $\tilde{\mathcal{V}}$ , therefore  $|\mathcal{T}| = P(N, 2)$ . The 2-tuple  $\tau_i \in \mathcal{T}$  is represented by  $\tau_i = (i_1, i_2)$ , where  $i = 1, \dots, |\mathcal{T}|$  is the index of the tuple. Namely,  $\tau_i$  is an ordered pair of two virtual bidding nodes, called “bidding pair” in this chapter.

The DA LMPs for nodes  $i_1$  and  $i_2$  in bidding pair  $\tau_i$  at hour  $h$  are denoted by  $\lambda_{i_1, h}^{DA}$  and  $\lambda_{i_2, h}^{DA}$ , respectively. Similarly, the RT LMPs for nodes  $i_1$  and  $i_2$  in bidding pair  $\tau_i$  at hour  $h$  are denoted by  $\lambda_{i_1, h}^{RT}$  and  $\lambda_{i_2, h}^{RT}$ , respectively. We ensure that INCs and DEC bids are cleared in the DA market by setting the bid price at the price floor/cap for INCs/DECs.

Assuming an INC offer is placed at node  $i_1$  and a DEC bid is placed at node  $i_2$ , the expected return  $r_{i, h}$  of the bidding pair  $\tau_i$  with 1 MWh bidding quantity on both nodes can be written as:

$$\begin{aligned}
E[r_{i,h}] &= E[r_{i_1,h}^I] + E[r_{i_2,h}^D] - 2\gamma \\
&= E[(\lambda_{i_1,h}^{DA} - \lambda_{i_2,h}^{DA}) - (\lambda_{i_1,h}^{RT} - \lambda_{i_2,h}^{RT})] - 2\gamma \\
&= E[\lambda_{i,h}^{dif,c}] - 2\gamma,
\end{aligned} \tag{3.1}$$

where  $r_{i_1,h}^I$  and  $r_{i_2,h}^D$  denote return of a 1 MWh INC offer and DEC bid at node  $i_1$  and  $i_2$  respectively.  $\gamma$  denotes the fixed transaction cost associated with an 1 MWh INC offer and DEC bid, which includes the bid segment fee and the market service fee. The difference between congestion spreads in DA and RT market,  $\lambda_{i,h}^{dif,c} = (\lambda_{i_1,h}^{DA} - \lambda_{i_2,h}^{DA}) - (\lambda_{i_1,h}^{RT} - \lambda_{i_2,h}^{RT})$ , is also called the LMP congestion spread. Since the energy component of the LMPs in DA and RT markets cancel out and the marginal loss component of LMPs are often small,  $\lambda_{i,h}^{dif,c}$  is mostly affected by the congestion patterns in DA and RT markets.

### 3.2.2 Virtual Bid Portfolio Optimization with Budget and Risk Constraints for Congestion Arbitrage

The objective of a virtual bid trading company is to design a virtual bid portfolio that maximizes its profit with budget and risk constraint. The trading quantity  $z_{i,h}$  for the two nodes in pair  $\tau_i$  at hour  $h$  is treated as a non-negative continuous decision variable. To mitigate the impact of congestion arbitrage on congestion pattern, we enforce two quantity constraints  $\mathcal{C}_h$  and  $\mathcal{D}_i$ .  $\mathcal{C}_h$  limits the total hourly virtual bid quantity at hour  $h$ , and  $\mathcal{D}_i$  limits the daily pairwise trading quantity of pair  $\tau_i$ .

The portfolio optimization problem for congestion arbitrage is formulated as follows:

$$\max_{\mathbf{z}} \sum_{h=1}^{24} \sum_{i=1}^{|\mathcal{T}|} z_{i,h} E[r_{i,h}] \quad (3.2)$$

$$s.t. \quad \sum_{i=1}^{|\mathcal{T}|} 2z_{i,h} \leq \mathcal{C}_h, \quad h = 1, \dots, 24 \quad (3.3)$$

$$\sum_{h=1}^{24} 2z_{i,h} \leq \mathcal{D}_i, \quad i = 1, \dots, |\mathcal{T}| \quad (3.4)$$

$$\sum_{i=1}^{|\mathcal{T}|} \sum_{h=1}^{24} z_{i,h} \beta_{i,h} \leq \mathcal{B} \quad (3.5)$$

$$\sum_{h=1}^{24} CVaR_{\alpha}(f_h(\mathbf{z}_h, \boldsymbol{\lambda}_h^{dif,c})) \leq \mathcal{M} \quad (3.6)$$

$$f_h(\mathbf{z}_h, \boldsymbol{\lambda}_h^{dif,c}) = - \sum_{i=1}^{|\mathcal{T}|} z_{i,h} r_{i,h}, \quad \forall h \quad (3.7)$$

$$z_{i,h} \geq 0, \quad \forall i, h, \quad (3.8)$$

where  $\beta_{i,h}$  denotes the collateral required by the market operator for placing 1 MWh of virtual bids on the two nodes in bidding pair  $\tau_i$ .  $\mathcal{B}$  is the portfolio budget limit.  $\mathcal{M}$  is the portfolio risk limit. Both portfolio budget and risk are modeled as constraints in the portfolio optimization formulation[38]. The conditional value-at-risk (CVaR) is leveraged to measure the financial risk of virtual bid portfolio in (3.6), where  $\alpha$  is the confidence level of CVaR,  $\mathbf{z}_h$  and  $\boldsymbol{\lambda}_h^{dif,c}$  are the vectors of decision variables and LMP congestion spreads, and  $f_h(\mathbf{z}_h, \boldsymbol{\lambda}_h^{dif,c})$  is the portfolio loss at hour  $h$ .

### 3.2.3 Reformulation of Risk Constrained Portfolio Optimization

In this work, CVaR is selected as the risk measurement for the portfolio. CVaR, also called mean shortfall, improves upon the Value-at-risk (VaR)[39] risk measure. CVaR controls not only

VaR but also the potential risk beyond VaR. CVaR has been shown to be a coherent measurement of risk, capturing the tail distribution of the portfolio loss [40].

CVaR of the portfolio with confidence level  $\alpha$ ,  $CVaR_\alpha$ , is defined as the expected portfolio loss provided that the loss is no less than its  $VaR_\alpha$ :

$$CVaR_\alpha(f_h(\mathbf{z}_h, \boldsymbol{\lambda}_h^{dif,c})) = \frac{1}{1-\alpha} \int_{f_h(\mathbf{z}_h, \boldsymbol{\lambda}_h^{dif,c}) \geq VaR_\alpha(\mathbf{z}_h)} f_h(\mathbf{z}_h, \boldsymbol{\lambda}_h^{dif,c}) p(\boldsymbol{\lambda}_h^{dif,c}) d\boldsymbol{\lambda}_h^{dif,c}, \quad (3.9)$$

where  $\alpha$  is selected to be 95%.

It was shown that  $CVaR_\alpha$  is upper-bounded by a convex and continuously-differentiable function  $F_\alpha(\mathbf{z}_h, \eta_h)$  [41]:

$$F_\alpha(\mathbf{z}_h, \eta_h) = \eta_h + \frac{1}{1-\alpha} \int_{\boldsymbol{\lambda}_h^{dif,c}} [f_h(\mathbf{z}_h, \boldsymbol{\lambda}_h^{dif,c}) - \eta_h]^+ p(\boldsymbol{\lambda}_h^{dif,c}) d\boldsymbol{\lambda}_h^{dif,c} \quad (3.10)$$

In fact, CVaR can be rewritten as follows [42]:

$$CVaR_\alpha(f_h(\mathbf{z}_h, \boldsymbol{\lambda}_h^{dif,c})) = \min_{\eta_h \in \mathbb{R}} F_\alpha(\mathbf{z}_h, \eta_h) \quad (3.11)$$

Assuming that the spatial-temporal distribution of the LMP congestion spread stays the same over time,  $F_\alpha(\mathbf{z}_h, \eta_h)$  can be sufficiently estimated by Monte Carlo sampling using historical LMP congestion spreads, which can be written as:

$$F_\alpha(\mathbf{z}_h, \eta_h) = \eta_h + \frac{1}{(1-\alpha)N_s} \sum_{j=1}^{N_s} [f_h(\mathbf{z}_h, \boldsymbol{\lambda}_{h,j}^{dif,c}) - \eta_h]^+, \quad (3.12)$$

where  $N_s$  is the number of historical samples of LMP congestion spreads.



To solve the optimization problem, the function  $\max(0, x)$  in (3.12) can be further relaxed with the slack variable  $\mathbf{u}_h$ :

$$F_\alpha(\mathbf{u}_h, \eta_h) = \eta_h + \frac{1}{(1-\alpha)N_s} \sum_{j=1}^{N_s} u_h^j \quad (3.13)$$

$$s.t. \quad u_h^j \geq f_h(z_h, \lambda_{h,j}^{dif,c}) - \eta_h \quad (3.14)$$

$$u_h^j \geq 0 \quad (3.15)$$

The virtual bid portfolio optimization problem in 3.2.2 can be reformulated as:

$$\begin{aligned} & \max_{z, \mathbf{u}, \eta} \quad (3.2) \\ s.t. \quad & \sum_{h=1}^{24} F_\alpha(\mathbf{u}_h, \eta_h) \leq \mathcal{M} \end{aligned} \quad (3.16)$$

$$(3.3) - (3.5), (3.7), (3.8), (3.13) - (3.15)$$

### 3.3 Data-Driven Methodology for Congestion Arbitrage

We propose a three-stage data-driven framework to arbitrage congestion using virtual bids. In stage one, we group nodes eligible for virtual bids with similar LMP patterns into clusters. In stage two, we develop deep neural network (DNN)-based LMP congestion spread forecast for typical nodes in different clusters. In stage three, we allocate the virtual bid quantity within each cluster.

#### 3.3.1 Cluster Analysis for Nodes Eligible for Virtual Bids

As discussed in 3.2.1, the cardinality of  $\mathcal{T}$  is  $P(N, 2)$ . CAISO has around 400 aggregated nodes eligible for virtual bids. Thus, the number of bidding pairs is approximately 160,000. It

would take a tremendous amount of computing resource to consider all possible pairs in the virtual bid portfolio optimization framework. A power system can be naturally divided into different zones connected by major transmission interfaces. The inter-zonal congestion is often more pronounced than the intra-zonal congestion. Thus, we propose selecting a representative node for each zone/cluster and using virtual bids to arbitrage DA and RT inter-zonal congestion patterns.

We first collect one year of historical DA and RT LMPs for each node into a  $2 \times 8760$  matrix. Then we perform dimension reduction using principal component analysis (PCA) to represent each node with a vector of length 8760. Afterwards, we apply k-means clustering to group nodes eligible for virtual bids into different zones. The dissimilarity between nodes is measured by the Euclidean distance between their dimension reduced LMP vectors. The number of clusters is selected using the “elbow method”[43]. Within each cluster, only one representative node with the shortest Euclidean distance to the cluster centroid is considered in the virtual bid portfolio optimization framework. The set of ordered pairs  $\mathcal{T}$  introduced in 3.2.1 is constructed based on these representative nodes with a dramatically-reduced cardinality.

### 3.3.2 DNN-based LMP Congestion Spread Forecast

Instead of separately predicting the DA and RT LMPs, we directly forecast the LMP congestion spread  $\lambda_{i,h}^{dif,c}$ . The cancellation of system energy components of the LMPs in  $\lambda_{i,h}^{dif,c}$  helps the forecast model to focus on capturing the congestion patterns. The LMP congestion spread is a nonlinear function of interpretable variables, consisting of regional load forecasts, fuel price forecasts, renewable generation forecasts, outage forecasts, and meteorological variables. Feedforward neural networks have been shown to be capable of learning such nonlinear mappings [32]. In this work, we develop a deep neural network to forecast the LMP congestion spreads.

On top of normalization for the input variables, we also need to perform data preprocessing for the target variables. This is because the LMP congestion spreads exhibit extremely high volatility. To mitigate the problem of oversized squared error on LMP spikes, appropriate scaling is applied on the LMP congestion spreads. Specifically, the parametric sigmoid function  $f(x) = \frac{1}{1+e^{-x/\theta}}$  is utilized to scale the LMP congestion spread to the range of (0, 1), where the hyper-parameter  $\theta$  needs to be carefully tuned. The activation functions of the DNNs are selected to be the hyperbolic tangent function “tanh” for the hidden layers and the standard sigmoid function for the output layer. Note that a DNN will be fit for each pair of representative nodes in  $\mathcal{T}$ .

### 3.3.3 Allocation of Virtual Bid Quantity within Each Group

As mentioned in 3.3.1, the virtual bid portfolio optimization are conducted using representative nodes from different clusters. Within each cluster, the virtual bid quantity derived from the portfolio optimization is imposed on the representative node. This raises a practical issue that large trading quantity allocated on a single node can cause inner-zonal congestion, which may reduce the profitability of the virtual bid strategy. To alleviate this problem, we need to reallocate the virtual bid quantity on the representative node to nodes in the same cluster/zone. The minimum virtual bid quantity for a node is 1 MWh. To ensure the reproducibility of our experiments, we first sort the nodes eligible for virtual bids alphabetically within each group. The node eligibility for virtual bids is determined by the market operator. Then we allocate 1 MWh of virtual bid quantity to each node in the cluster recursively according to the ordering until the total trading quantity for the representative node in the cluster is completely distributed.

## 3.4 Numerical Study

### 3.4.1 Setup of Numerical Study

We validate the proposed algorithmic virtual bid trading strategy on California Independent System Operator (CAISO) managed electricity market. 3 years of historical data are collected for CAISO from July 2017 to June 2020, where the 1st year of data is used for model training and the last 2 years of data is used for rolling forecast and testing. The rolling forecasts and forecast model updates are performed on a monthly basis. The inputs to the DNN model are composed of load forecast, fuel price forecast, renewable generation forecast (wind and solar), outage forecast by generation type (natural gas, hydro, geothermal, nuclear, and others), meteorological variables (temperature, humidity, wind speed, and precipitation), and one-hot encoding for the operating hour.

The DNN consists of four fully-connected layers with 200, 100, 50, and 25 hidden units respectively. The activation functions of the hidden layers are “tanh”. The output layer has 1 hidden unit with sigmoid function as the activation. The dropout rate is selected as 20% for each hidden layer. The learning rate is set to 0.001. The optimizer used to train the DNN is Adam. The training batch size is 2048. The training epoch is limited to 1000 with early-stopping and the patience is set to 20. The scaling parameter  $\theta$  in the parametric sigmoid function of LMP congestion spreads is set to 10.

K-means cluster analysis is performed to group the nodes eligible for virtual bids into different clusters. By using the elbow method, the number of clusters  $M$  is selected to be 30. Increasing the number of cluster beyond  $M = 30$  does not significantly reduce the within-cluster sum-of-squares.

### **3.4.2 Results of K-Means Clustering for Bidding Nodes**

In this subsection, we depict the geographical information for some of the clustered nodes eligible for virtual bids from CAISO. The longitude and latitude data of non-aggregated nodes are collected from CAISO's LMP price contour maps. Since the nodes eligible for virtual bids considered in this chapter are aggregated bidding nodes, we manually mapped a subset of the aggregated bidding nodes to the non-aggregated ones. To avoid overcrowding, we only show a subset of nodes from 13 clusters of the k-means clustering results.

The geographical information for 13 groups of bidding nodes is depicted in Figure 3.1. The proposed k-means clustering algorithm manages to discover groups of nodes in geographically separated zones based on the DA and RT LMP features. The nodes eligible for virtual bids within the same group are usually close to each other geographically.

### **3.4.3 Profitability of Algorithmic Virtual Bid Trading Strategy for Congestion Arbitrage**

To evaluate the profitability of algorithmic virtual bid trading strategy for congestion arbitrage, we conduct the virtual bid portfolio optimization using data from July 2018 to June 2020 with LMP congestion spread forecasts from the DNN model. The daily portfolio budget limits are set to \$100K, \$50K, and \$25K for the first year and \$120K, \$60K, and \$30K for the second year, corresponding to 5.48%, 2.83%, and 1.44% virtual bid market share respectively. The daily portfolio risk limits are set to be half of the budget limits for all cases. The market service fee is \$0.0836/MWh and the bid segment fee is \$0.005/bid segment.

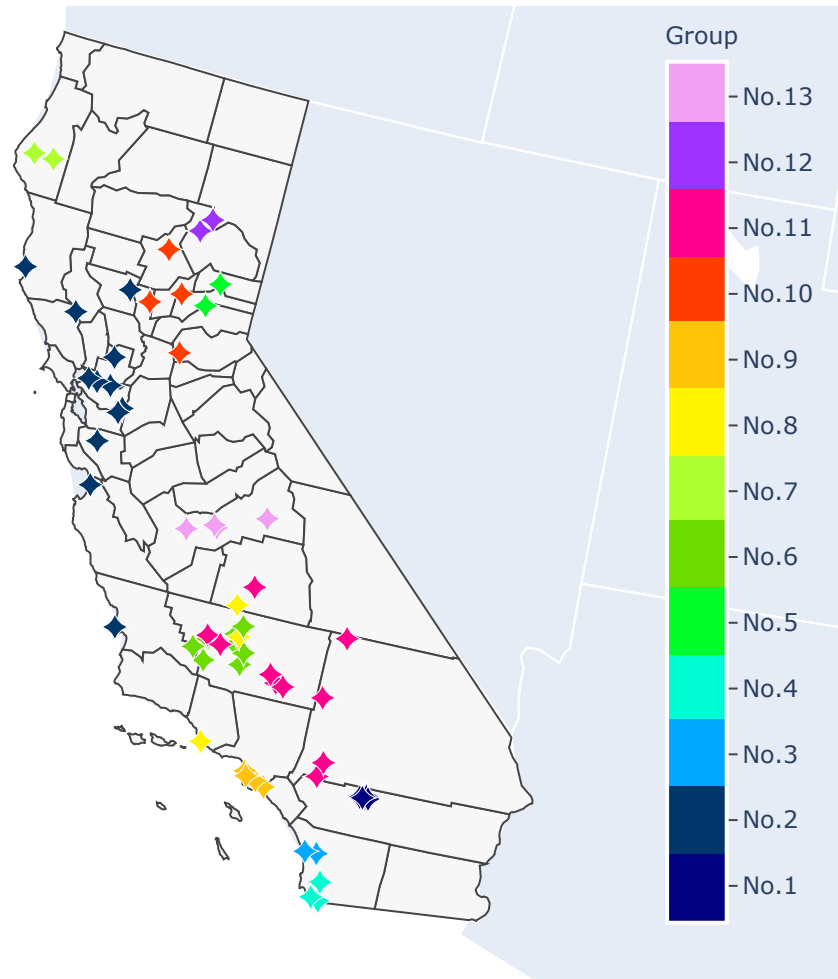


Figure 3.1: Geographical information of bidding nodes in 13 groups.

As discussed in 3.3.3, the portfolio optimization is first performed on the representative nodes from the clusters. The virtual bid quantities are then assigned to all nodes eligible for virtual bids in the clusters. Figure 3.2 shows the cumulative profits of the proposed virtual bid trading strategy with different budget and risk limits. After accounting for all trading costs including the market service fee and the bid segment fee, the cumulative profits of the proposed strategy are \$4.59M, \$2.83M, and \$1.58M for 5.48%, 2.83%, and 1.44% virtual bid market share respectively. The cumulative profit per dollar of trading budget is higher as the portfolio budget is lower.

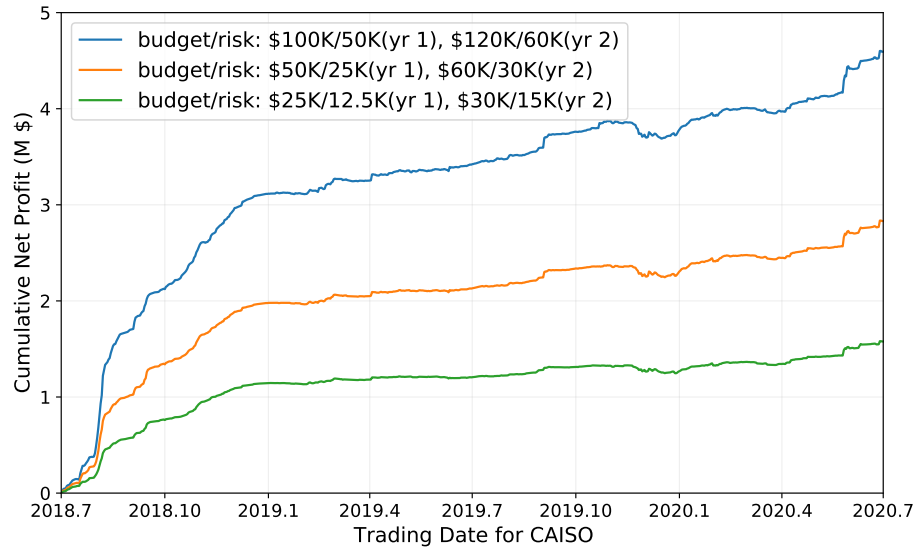


Figure 3.2: Cumulative profits of the proposed virtual bid trading strategy in a 2-year period for CAISO. The daily budget limits correspond to 5.48%, 2.83%, and 1.44% of virtual bid market shares. The daily risk limit is set to be half of the portfolio budget.

Table 3.1 shows the top 3 most profitable virtual bid pairs by season. Winter is from December to February, Spring is from March to May, Summer is from June to August, and Fall is from September to November. In each pair, the first number denotes the group ID of INC offer and the second number denotes the group ID of DEC bid. It is interesting to observe that the most profitable virtual bid pairs in summer and fall arbitrage the difference in congestion pattern between Bay area and California/Oregon border.

### 3.5 Conclusion

This chapter develops an algorithmic trading strategy to arbitrage congestion with virtual bids in the wholesale electricity market. The k-means clustering algorithm is implemented to group the bidding nodes based on their LMP patterns. The algorithmic trading strategy aims to exploit the

Table 3.1: Top 3 Most Profitable Virtual Bid Pairs by Season

Season	1st Pair	2nd Pair	3rd Pair
Winter	(1, 13)	(6, 5)	(6, 3)
Spring	(11, 2)	(11, 10)	(11, 13)
Summer	(10, 7)	(1, 7)	(11, 2)
Fall	(2, 7)	(13, 7)	(6, 7)

inter-zonal congestion patterns. The risk-constrained portfolio optimization problem with virtual bids is formulated as a linear programming problem via convex relaxation. The empirical results with CAISO managed electricity market demonstrate that the proposed framework is capable of capturing the congestion differences in node pairs and making notable profits by arbitraging the congestion. In the future, we plan to further refine the proposed model by studying the impact of different distance measures in k-means cluster analysis and neural network architectures and hyperparameters on the portfolio profitability. We also plan to develop a model to jointly capture the LMP congestion spreads among all possible pairs of nodes eligible for virtual bids.

### 3.6 List of Symbols in Chapter 3

- $\tau_i$  Ordered pair of two virtual bidding nodes
- $\lambda_{i,h}^{dif,c}$  Difference between congestion spreads for pair  $\tau_i$  at hour  $h$  (\$/MWh)
- $\lambda_h^{dif,c}$  Vector of difference between congestion spreads at hour  $h$  (\$/MWh)



$r_{i,h}$	Net profit for pair $\tau_i$ at hour $h$ (\$)
$z_{i,h}$	Decision variables for pair $\tau_i$ at hour $h$ (MWh)
$\mathbf{z}_h$	Vector of decision variables at hour $h$ (MWh)
$\beta_{i,h}$	Collateral of virtual bids for pair $\tau_i$ at hour $h$ (\$/MWh)
$\gamma$	Transaction cost (\$/MWh)
$\mathcal{C}_h$	Total hourly virtual bid quantity at hour $h$ (MWh)
$\mathcal{D}_i$	Daily pairwise trading quantity of pair $\tau_i$ (MWh)
$\mathcal{B}$	Portfolio budget limit (\$)
$\mathcal{M}$	Portfolio risk limit (\$)

## **Chapter 4**

# **Predict Locational Marginal**

# **Greenhouse Gas Emission Factors of**

# **Electricity with Spatial-Temporal Graph**

# **Convolutional Networks**

## **4.1 Introduction**

Excessive emission of greenhouse gas (GHG) can cause global climate change and notable environment impact, such as global warming and rising sea-levels. Reducing GHG emission is the key to slowing such detrimental processes. The U.S. government has announced a target of 50-52% GHG emission reduction below 2005 levels by 2030 [10]. The electric power system is one of the main contributors of GHG emissions, producing about 25% of the total GHG emissions [44].

In power system operations, a mixture of generation resources are coordinated not only to meet the varying electricity demand with least cost while satisfying a number of operational constraints. Different generation resources have different levels of GHG emissions. Fossil-fueled power plants are major GHG emission sources while solar and wind resources do not emit GHG at all in daily operations. The GHG emission from power systems is influenced by many factors [45] such as generation mix, time of the day, season, electric load level, and the topology of the power system.

To reduce GHG emissions, accurate GHG emission predictions are in critical need. There are two major GHG emission factors: average emissions factor (AEF) and marginal emissions factor (MEF). The AEF is calculated as the ratio of total GHG emissions to the total power consumption. The MEF is the ratio of the change of GHG emissions to the change of power consumption. Compared with AEF, MEF is a more useful tool for distributed energy resources aggregators and end-use customers to make intelligent decisions about how much electricity should be consumed at different time slots of a day. MEF signals can be sent along with electricity prices signals to flexible loads and other smart technologies of the residential, commercial and industrial customers [11]. Based on the MEF signals, flexible loads, such as electric vehicles (EVs), smart thermostats, and batteries can consume or charge less power during high MEF hours and more power during low MEF hours. MEF can also be considered in new electricity pricing design to help reduce GHG emissions.

GHG emission and marginal GHG emission prediction methods are in their early stage of research and development. The existing methods can be classified into three groups. In the first group, GHG emission or MEF is estimated through production cost simulations of power systems and electricity markets. Reference [45] used load duration curve and economic dispatch to emulate power plants dispatch. In [46], the order of dispatching was empirically derived to calculate AEF

and MEF. In [47], simulations in detailed transmission system models were used to estimate GHG emissions. These approaches have two drawbacks. First, to accurately predict AEF or MEF, high-fidelity production cost simulations of electricity markets are needed. This can be computationally expensive if MEF needs to be calculated at high granularity in space and time. Second, only the market operators have access to accurate models of transmission networks and propriety bids and offers submitted by power producers and load serving entities, making it difficult for others to apply these methods.

The second group of methods are based on clustering and linear regression. In [48, 49], linear regression was used to predict GHG emissions from load. In [50], cluster analysis was first conducted on daily load curves and linear regression models were developed for each cluster to predict MEF. The drawback of these approaches is that linear regression models can not accurately capture the complex interactions between different influential factors in determining GHG emissions.

In the third group, machine learning models such as feed-forward neural networks (FNNs) [51, 52], support vector machines (SVMs) [53], and long short-term memory (LSTM) networks [52] were proposed to predict GHG emissions. In [54], an ensemble model combining multiple basic models such as FNN, LSTM, and random forest (RF) was trained to predict GHG emissions. Although machine learning models have shown good prediction accuracy, most of them were designed to do long-term predictions with very low time granularity, such as yearly GHG emissions, which are not sufficient for short-term control of flexible loads and smart buildings. Furthermore, these models were designed separately to forecast GHG emissions in each region, ignoring the interactions between different load serving zones, which greatly limited the prediction accuracy.

The marginal GHG emission factor varies by load zones, or electric buses, due to the limited power transfer capability between zones and nodes in the power system. To accurately predict the locational MEF, the information from the entire transmission network should be effectively leveraged. Graph neural network (GNN) [55, 56] is an ideal candidate for processing and learning from information collected from a complex network such as the power grid. In fact, GNN has received increasing attention in recent years from researchers to tackle a number of prediction, estimation and optimization problems in power systems such as optimal power flow [57], solar energy prediction [58], parameter estimation [59], state estimation [60], and system health index prediction [61].

In this chapter, we propose a hybrid machine learning model, which integrates FNNs with spatio-temporal graph convolutional networks (STGCNs) [14] to predict GHG emissions and the MEF. We adopted the STGCN due to its capability to efficiently capture temporal and spatial structure of the network data. Compared with existing GHG emission prediction methods, our proposed model has three advantages. First, it captures the complex interactions between multiple load zones and thus provides highly accurate locational MEF prediction. Second, it provides short-term MEF predictions with hourly granularity to aggregators and end-users to control flexible loads. Third, the model uses only publicly available electricity market and power system information, making it widely applicable. A comprehensive case study in California electricity market shows that the proposed method has more accurate predictions of GHG emissions and MEF than baseline methods.

The rest of the chapter is organized as follows. Section 4.2 describes the problem setup and the dataset. Section 4.3 presents the technical details of the proposed hybrid machine learning

model. Section 4.4 evaluates the GHG emission prediction performance and the analyzes the MEF in different scenarios. Section 4.5 states the conclusion.

## 4.2 Problem Setup and Dataset Description

### 4.2.1 Problem Setup

The objective of this work is to obtain an accurate prediction of GHG emissions as well as locational MEF of electricity. Let  $E_t(x_{1t}, x_{2t}, \dots, x_{Rt})$  be the GHG emission of a system at time  $t$  with influential factors  $x_{1t}, x_{2t}, \dots, x_{Rt}$ .  $x_{rt}$  is the load of a subsystem  $r$  at time  $t$ . Then the locational MEF due to a local load change  $\Delta x$  at subsystem  $r$  at time  $t$  is defined in (4.1). If the GHG emission is measured in metric ton of carbon dioxide equivalent per hour ( $mTCO_2/h$ ) and the load is measured in megawatt ( $MW$ ), then the MEF is measured in  $mTCO_2/MWh$ .

$$\frac{\Delta E}{\Delta G} = \frac{E_t(x_{1t}, \dots, x_{rt} + \Delta x, \dots, x_{Rt}) - E_t(x_{1t}, \dots, x_{rt}, \dots, x_{Rt})}{\Delta x} \quad (4.1)$$

The prediction of GHG emissions and locational MEF are conducted in two steps. First, machine learning models are trained to predict GHG emissions based on electric power load and other input features. Second, machine learning models are used to make new GHG emission predictions with perturbed load levels. The locational MEF is then calculated following the definition in (4.1).

### 4.2.2 Description of the Dataset

In this chapter, we used real-world GHG emissions of California in the test case. We utilized data from six sources, which are summarized in Table 4.1. All of the data, except time,

were collected from the Open Access Same-time Information System (OASIS) of the California Independent System Operator (CAISO), which is publicly-available [62]. In the dataset, CAISO measured GHG emissions by tracking the power generations of power plants, the heat rate provided by generation asset owners, and the GHG emission factors by resource types [63]. The data were collected between July 2018 and September 2022. After data cleansing and preprocessing, about 37,000 hours of data remained.

The electric load data were collected from not only California, but also other parts of western United States. These regions and the corresponding balancing authorities are summarized in Table 4.2 and their locations are illustrated in Fig. 4.1. All the load data used were one day-ahead predictions made by system operators with mean absolute percentage error (MAPE) below 3%.

The solar and wind power generation data in California were from three regions (coded as NP15, ZP26, and SP15) specified by CAISO, which represent the north, middle, and south of California. They were two day-ahead predictions made by CAISO with MAPE below 7%. Natural gas is the main fuel source for thermal power plants in California, and thus its price in different regions of California was collected. The power supply by resource type reported by CAISO are used as input features. The output of the prediction model is the total GHG emissions of all the electric power resources in California.

Table 4.1: Summary of Input and Output Data

	Data Type
Input	Day-ahead hourly electric load predictions of different balancing authorities (MWh).
	Two day-ahead predictions of solar and wind power generation of different regions (MWh).
	Hourly natural gas price of different regions (dollars per thousand cubic feet).
	Hourly power supply from different generation resources: renewables, natural gas, large hydro, imports, batteries, nuclear, and coal (MWh).
	Time: month, hour, weekday/weekend.
Output	Total hourly California power grid emission from all resources: imports, natural gas, biogas, biomass, geothermal, and coal (mTCO <sub>2</sub> )



Table 4.2: Summary of the Regions and Balancing Authorities of the Load Data

Code Name	Full Name	State
MWD-TAC	Metropolitan Water District Transmission System	California
PGE-TAC	Pacific Gas And Electric	California
SCE-TAC	Southern California Edison	California
SDGE-TAC	San Diego Gas And Electric	California
VEA-TAC	Valley Electric Association	California
IPCO	Idaho Power Company	Idaho
PACE	PacifiCorp East	Utah
PACW	PacifiCorp West	Utah
PGE	Portland General Electric	Oregon
PSEI	Puget Sound Energy	Washington
NEVP	Nevada Energy	Nevada

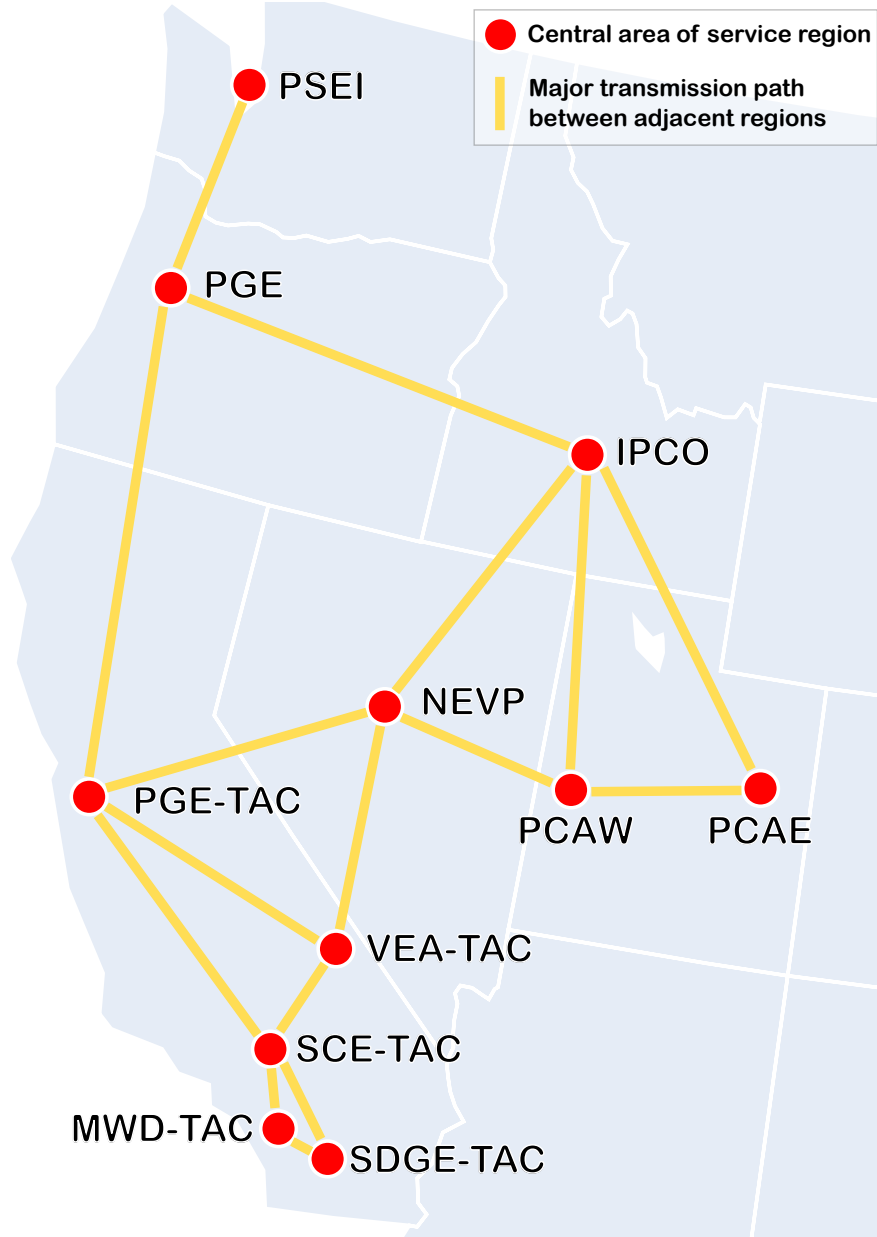


Figure 4.1: Electric power interconnection map

### 4.3 Technical Methodology

#### 4.3.1 Overall Framework of the GHG Emission Prediction Model

The overall framework of the proposed GHG emission prediction model is illustrated in Fig. 4.2. As shown in the figure, we design a hybrid model that combines both STGCNs and

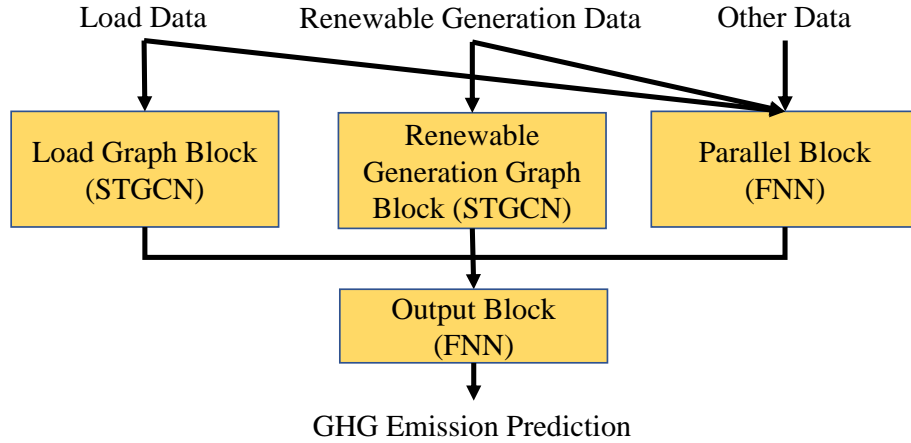


Figure 4.2: Framework of the Prediction Model

FNNs. The load graph block and the renewable generation block are two STGCN models, while the parallel block is an FNN model. The load data and the renewable generation data are fed into the load graph block and renewable generation graph block respectively; these two types of data and other input data are also fed into the parallel block. The outputs of these three blocks are concatenated into one tensor for each time instance, and fed into the output block, which is an FNN network and its output is the predicted GHG emission. The details of the proposed method is described in the following subsections, including the data preprocessing, brief introduction of STGCN, the design of each block, and the data split and hyperparameter tuning.

## 4.3.2 Data Preprocessing

### 4.3.2.1 Preprocessing of Time Data

We used a binary variable to represent weekday (value 0) and weekend (value 1). To represent month and hour, we use cyclical encoding. In cyclical encoding, The  $k$ -th hour ( $k = 1, 2, \dots, 24$ ) is encoded by  $[\cos \frac{2\pi k}{24}, \sin \frac{2\pi k}{24}]$ . Similarly, the  $k$ -th month ( $k = 1, 2, \dots, 12$ ) is encoded

by  $[\cos \frac{2\pi k}{12}, \sin \frac{2\pi k}{12}]$ .

#### 4.3.2.2 Transformation of Data

To improve the convergence in training and prediction accuracy, we applied two types of transformation to the data: z-score normalization and quantile transformation. In the z-score normalization, the data are centered and normalized by their standard deviation. In the quantile transformation, the data are transformed to follow a normal distribution. The z-score normalization is a linear transformation, which preserves the correlations and distances within the data; on the other hand, the nonlinear quantile transformation smooths out unusual distributions and is less influenced by outliers than z-score.

### 4.3.3 Brief Introduction of STGCN

Here we briefly introduce the design of STGCN and more details can be found in [14]. STGCN is designed to process and learn data set collected from a graph. Let  $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$  be a graph, in which  $\mathcal{V}$  is the set of vertices (nodes),  $\mathcal{E}$  is the set of edges, and  $W \in \mathbb{R}^{n \times n}$  is the weighted adjacency matrix ( $n = |\mathcal{V}|$ ). The structure of the STGCN model is illustrated in Fig. 4.3. An STGCN contains three parts: two spatial-temporal convolutional (ST-Conv) blocks and an output layer. Each ST-Conv block contains a “sandwich” structure of two temporal gated convolution (Gated-Conv) layers and one spatial graph convolution (Graph-Conv) layer. The output layer is a fully connected layer.

The Graph-Conv layer uses Chebyshev polynomials to approximate a graph convolution

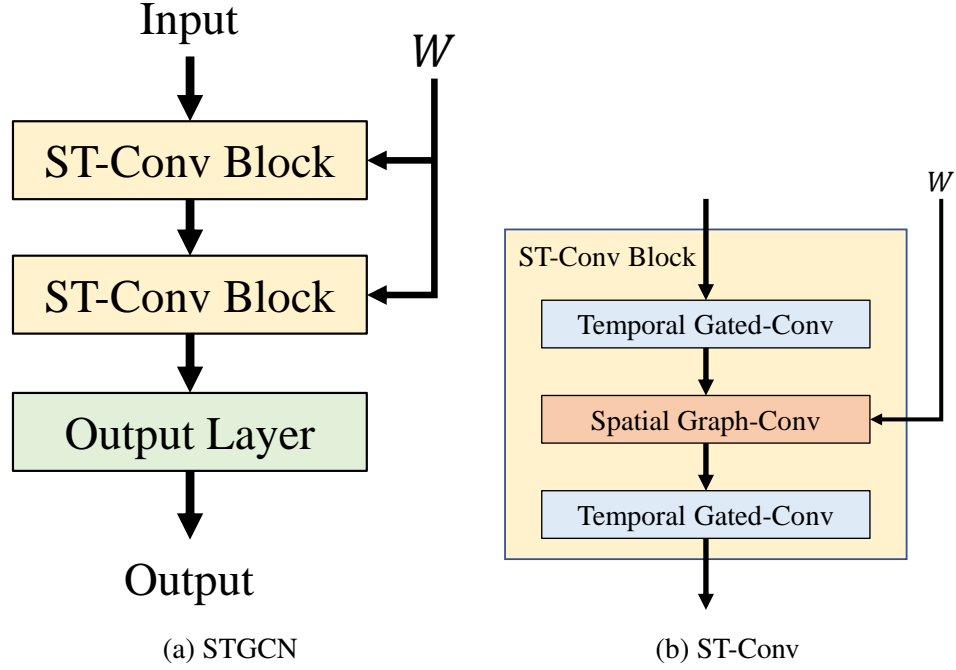


Figure 4.3: The structure of an STGCN model. (a) is the overall structure of STGCN. (b) is the internal structure of an ST-Conv block.

operation “ $\ast_{\mathcal{G}}$ ” with a kernel  $\Theta$  as defined in (4.2):

$$\Theta_{\ast_{\mathcal{G}}}\mathbf{x} = \Theta(L) \approx \sum_{k=0}^{K-1} \theta_k T_k(\tilde{L})\mathbf{x} \quad (4.2)$$

Here,  $T_k(\tilde{L}) \in \mathbb{R}^{n \times n}$  is the Chebyshev polynomial of order  $k$ ,  $\tilde{L} = 2L/\lambda_{max} - I_n$  is the scaled Laplacian, and  $\mathbf{x}$  is a  $n$ -dimensional vector representing the inputs at the graph nodes.  $L = I_n - D^{-\frac{1}{2}}WD^{-\frac{1}{2}}$ .  $I_n$  is an identity matrix.  $D \in \mathbb{R}^{n \times n}$  is the diagonal degree matrix derived from  $W$ , and  $\lambda_{max}$  is the largest eigenvalue of  $L$ .  $K$  is the kernel size, which determines the maximum radius of the convolution. When each node has a  $C_i$ -channel input and  $C_o$ -dimensional output, the graph is generalized to (4.3):

$$\mathbf{y}_j = \sum_{i=1}^{C_i} \Theta_{i,j}(L)\mathbf{x}_i \in \mathbb{R}^n, 1 \leq j \leq C_o \quad (4.3)$$

Here  $\mathbf{y}_j \in \mathbb{R}^n$  is the nodal output of channel  $j$ .  $\mathbf{x}_i \in \mathbb{R}^n$  is the nodal input of channel  $i$ . When

multiple time steps are considered, the input and output will have an additional dimension for time steps.

The temporal Gated-Conv layer is applied to each node in the graph. Let  $M$  be the input time steps; let  $C_i$  and  $C_o$  be numbers of the input and output channels. The Gated-Conv layer first uses a convolution kernel  $\Gamma \in \mathbb{R}^{K_t \times C_i \times 2C_o}$  to perform 1-D causal convolutions of width  $K_t$  and obtains two elements  $P$  and  $Q$ , such that  $[P, Q] \in \mathbb{R}^{(M-K_t+1) \times (2C_o)}$ . Then this layer uses element-wise Hadamard product ( $\odot$ ) to obtain its out as  $P \odot \sigma(Q) \in \mathbb{R}^{(M-K_t+1) \times C_o}$ , where  $\sigma()$  is the sigmoid gate.

#### 4.3.4 Load Graph Block

To process the load data with a STGCN model, we need to determine the graph, model structure, the inputs and outputs of the model. We use the graph shown in Fig. 4.1, in which each node represents a balancing authority and each edge represents a major transmission path between adjacent regions. We construct the adjacency matrix  $W$  for the model by using the same weight for each edge in the graph. We use a single-step in both the inputs and outputs of the STGCN. Thus, in the temporal Gated-Conv layers, we set  $K_t = 1$ . We assume that each region have direct interactions with only its nearest neighbors, thus we use a maximum radius of 1 in the Spatial Graph-Conv layer, i.e.  $K = 2$ . To predict the GHG emission at hour  $h$ , the inputs of each node in this graph are the load at hour  $h$  and  $h - 24$  of the corresponding region with both the  $z$ -score and quantile transformation. Note that load used here are all one day-ahead predictions.

### 4.3.5 Renewable Generation Graph Block

Similar to the load data, the renewable generation data of solar power and wind power also have a graph structure. The renewable generation data was recorded separately in the three regions: NP15, ZP26, and SP15, representing the north, central, and the south parts of California. Thus, the graph of this block was designed as three nodes connected to each other with three edges. We use a single time step in both the input and output, thus we set  $K_t = 1$ . Since it is a small graph, any non-zero convolution radius will average the nodal features. Hence, we set  $K = 1$  to avoid this. To predict the GHG emission at hour  $h$ , the inputs of this block are the renewable generation power at hour  $h$  and  $h - 24$  with both the  $z$ -score and quantile transformation for each of the three nodes. Note that the renewable generation data used here are all two day-ahead predictions.

### 4.3.6 Parallel Block

Since not all input data are collected from a graph, we also design a parallel block to extract information that is not captured by the STGCN blocks. The inputs to the parallel block include the load data, renewable generation data and other input data. The other input data are the time feature data (described in Section 4.3.2) of hour  $h$  and  $h - 24$ , the natural gas price at hour  $h$  and  $h - 24$ , the supply resource mix at hour  $h - 24$ , and the historical GHG emission at hour  $h - 24$ . Every input feature except time and historical GHG emission uses both  $z$ -score and quantile transformation. The historical GHG emission uses only the quantile transformation.

### **4.3.7 Output Block**

The output block is designed as an FNN with batch normalization before each layer. Its output is the quantile transformation of the GHG emission. The proposed hybrid model is trained to minimize the mean squared error of the quantile-transformed GHG emission. To obtain the final prediction, inverse of quantile transform is performed.

### **4.3.8 Data Split and Hyperparameter Tuning**

To train the proposed model and tune its hyperparameters, the dataset is split into three parts. The first 80% of samples are used as the training and validation dataset while the last 20% of the samples were used as the testing dataset to evaluate the model's GHG emission prediction performance. In the first 80% of samples, for every five day's data (120 samples), we put the first four days into the training dataset and the last day's data into the validation dataset. Thus, 64% of the whole data set is used as the training dataset and 16% is used as the validation dataset. The proposed machine learning model contains many hyperparameters: the number of layers in each block, the dimension of each layer, the size of spatial kernel and temporal kernel, learning rate, etc. To systematically tune hyperparameters, for each hyperparameter setup, we trained the model 10 times using the training dataset, and calculate the average prediction error using the validation dataset. The hyperparameter setup with the lowest average error in the validation dataset is chosen as the best hyperparameter setup.



## 4.4 GHG Emission Prediction Performance and Analysis of Marginal

### GHG Emission

In this section, we evaluate the prediction performance for California’s GHG emissions of the proposed hybrid machine learning model and compared it with two baseline algorithms. We also calculate and analyze the locational MEF under different scenarios.

#### 4.4.1 GHG Emission Prediction Performance

We compare the GHG emission prediction performance of our proposed model (hybrid STGCN) and two other baseline models: FNN and gradient boosted trees (GBT). Note that the FNN and GBT models use the same input features as our proposed model.

The hyperparameters all three machine learning models were tuned following the approach in Subsection 4.3.8. We trained the hybrid STGCN model and FNN model using the Adam algorithm, with batch size = 10 and early stopping patience = 10 epochs. We trained the GBT with early stopping patience = 10 rounds. By tuning hyperparameters, the numbers of channels of the three “sandwich” layers in ST-Conv and the output layer of STGCN are 4–2–4–2 respectively in the load graph block, and 8–4–8–2 in the renewable generation graph block; the FNN in the parallel block and the output block are two three-layer FNNs of dimension 20–20–1. The number of neurons of the FNN model is 20–20–20–1. All FNN models use batch normalization before each layer.

To evaluate the prediction accuracy of the machine learning models, we train each model 10 times using the training dataset, and then test the models with the testing dataset. Three error metrics are used to evaluate the prediction accuracy: mean squared error (MSE), mean absolute

error (MAE), and mean absolute percentage error (MAPE). The prediction performance of the three machine learning models are compared in Table 4.3. For each type of measurements, two values were recorded: the average performance over 10 tests, and the optimal value, i.e. the performance of the model with the lowest validation loss. We can see that our proposed hybrid STGCN model has the lowest prediction error in MSE, MAE, and MAPE in both average value and the optimal value. In addition, by choosing the optimal value from multiple trained models, our proposed prediction model can further improve the prediction accuracy. These results show that by capturing the complex spatio-temporal relationship of the data, our proposed model can significantly improve the accuracy of GHG emission prediction.

Table 4.3: Prediction Performance of GHG Emissions. (Average Value / Optimal Value)

Model	MSE $(mT\text{CO}_2/h)^2$	MAE $(mT\text{CO}_2/h)$	MAPE (%)
Hybrid STGCN	2.90/2.77E+05	4.02/3.88E+02	9.46/9.15
FNN	3.01/3.11E+05	4.09/4.18E+02	10.80/10.36
GBT	3.25/3.23E+05	4.27/4.25E+02	11.43/11.43

#### 4.4.2 Analysis of Locational Marginal GHG Emission

We use the trained hybrid model to calculate the locational MEF and  $\Delta E$ , following the definition in (4.1). The locational MEF and  $\Delta E$  are calculated for each hour and the load shift for each region in the data, with  $\Delta G = 100MW$ . We analyze the locational marginal MEFs in two aspects: the hourly pattern of locational MEF with high renewable energy and the weekday effect.

#### 4.4.2.1 Locational MEF on a Day with High Renewable Energy Output

California has very high renewable energy penetration rate. On March 27, 2022, California hit a record that 94.5% of the electricity on the grid came from renewable energy [64]. We calculate the 24-hour locational MEF on this day in the balancing authority of PGE-TAC and SCE-TAC respectively. The result is illustrated in Fig. 4.4. From this figure, we can see that the marginal GHG emission is significantly lower during the day. This is because solar photovoltaic (PV) generation is very high during these hours and does not emit any GHG. This result shows that the proposed hybrid STGCN model successfully recognizes the contribution of renewable energy in reducing GHG emission. Furthermore, the MEF for SCE-TAC is slightly lower than that of PGE-TAC between 12:00 pm and 18:00 pm. This is because Southern California has much higher solar PV generation and not all renewable energy can be moved from Southern California to Northern California due to limited power transfer capability.

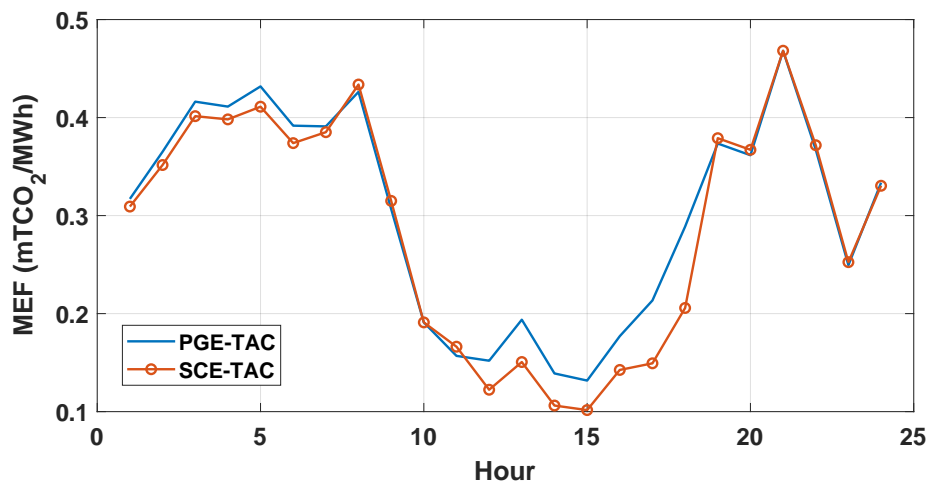


Figure 4.4: Hourly MEF on March 27, 2022 for PGE-TAC and SCE-TAC.

#### 4.4.2.2 Weekday/Weekend Effect on Marginal GHG Emission

We calculate the average locational MEF for 24 hours on weekdays and weekends in the PGE-TAC area. The result is illustrated in Fig. 4.5. From this figure, we can see that the MEF is lower during the day, which has been explained in Subsection 4.4.2.1. We can also observe that the weekends have lower MEF than the weekdays. This is due to the lower power demand on weekends. When there are lower power demand, system operators can turn off the less fuel-efficient power stations and keep running the power plants with higher fuel-efficiency. These results show that the proposed hybrid STGCN model can reflect the GHG emission differences between weekday and weekends and between different operation conditions.

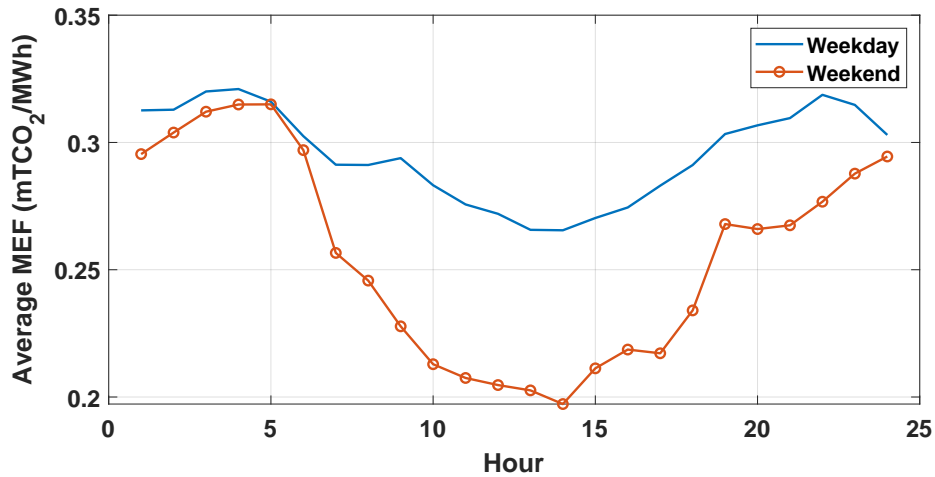


Figure 4.5: Average hourly MEF on weekdays/weekends in PGE-TAC.

## 4.5 Conclusion

In this chapter, we developed a hybrid machine learning model by integrating FNN with STGCNs to predict GHG emissions and locational MEF. The STGCN components of the model

allows us to capture the complex spatio-temporal correlations in the network data and improves the prediction accuracy. The proposed model can provide short-term locational MEF predictions at hourly granularity to aggregators and end-users to manage flexible loads and it does not require accurate power system model. The numerical study on California’s electricity market shows that the proposed method has more accurate GHG emission predictions than the baseline machine learning models. Detailed analysis also showed how the locational MEF is influenced by load level, hour, and renewable generation levels.

## 4.6 List of Symbols in Chapter 4

$\mathcal{G}$	Graph
$\mathcal{V}$	Set of Vertices
$\mathcal{E}$	Set of edges
$\mathbf{x}_i$	$\mathbf{x}_i \in \mathbb{R}^n$ Nodal input of channel $i$
$\mathbf{y}_j$	$\mathbf{y}_j \in \mathbb{R}^n$ Nodal output of channel $j$
$W$	Weighted adjacency matrix
$D$	Diagonal degree matrix derived from $W$
$\tilde{L}$	Scaled Laplacian
$M$	Input time steps
$K$	Spatial kernel size
$K_t$	Temporal kernel size
$C_i, C_o$	Numbers of input and output channels

## Chapter 5

# Conclusions

In this dissertation, we demonstrate all the research works that have been done regarding the topic: data-driven modeling and algorithmic trading in electricity market.

In Chapter 2, we discussed the algorithmic virtual bid trading framework considering the price sensitivity. A constrained gradient boosting tree is developed to model the monotonic function of price sensitivity. The risk-constrained virtual bid portfolio optimization problem is reformulated as a mixed-integer quadratically-constrained problem via convex relaxation. The results of comprehensive empirical studies on the three U.S. electricity markets show that the proposed virtual bid portfolio optimization framework considering price sensitivity outperforms the one without it.

In Chapter 3, we explored the algorithmic virtual bid trading strategy to arbitrage the congestion in wholesale electricity market. The k-means clustering algorithm is implemented to group the bidding nodes based on their price patterns. The risk-constrained portfolio optimization problem with virtual bids is formulated as a linear programming problem via convex relaxation. The empirical results on CAISO demonstrate that the proposed framework is capable of capturing

the congestion differences in node pairs and making notable profits by arbitraging the congestion, while eliminating the significant uplift costs.

In Chapter 4, we interrogated the hybrid machine learning framework of FNNs and STGCNs to predict GHG emissions and locational MEF. The STGCN components enable us to capture the complex spatio-temporal correlations in the network data and improves the prediction accuracy. The proposed data driven model can provide short-term locational MEF predictions at hourly granularity without power system modeling. The numerical study on CAISO shows that the proposed method has more accurate GHG emission and locational MEF predictions than the baseline machine learning models.

# Bibliography

- [1] W. W. Hogan, “Virtual bidding and electricity market design,” *The Electricity Journal*, vol. 29, no. 5, pp. 33–47, 2016.
- [2] PJM Interconnection, “Virtual transactions in the PJM energy markets,” 2015. [Online]. Available: <https://www.pjm.com/-/media/committees-groups/committees/mc/20170622/20170622-item-01-emustf-presentation.ashx>
- [3] R. Li, A. Svoboda, and S. Oren, “Efficiency impact of convergence bidding in the California electricity market,” *Journal of Regulatory Economics*, vol. 48, no. 3, pp. 245–284, 2015.
- [4] A. Jha and F. Wolak, “Testing for market efficiency with transactions costs: An application to convergence bidding in wholesale electricity markets,” *Industrial Organization Seminar, Yale University*, May 2013.
- [5] J. Kazempour and B. F. Hobbs, “Value of flexible resources, virtual bidding, and self-scheduling in two-settlement electricity markets with wind generation—Part II: ISO models and application,” *IEEE Transactions on Power Systems*, vol. 33, no. 1, pp. 760–770, 2017.
- [6] J. Birge, A. Hortaçsu, I. Mercadal, and M. Pavlin, “Limits to arbitrage in electricity markets: A case study of MISO,” *MIT Center for Energy and Environmental Policy Research*, January 2017. [Online]. Available: <http://cepr.mit.edu/files/papers/2017-003.pdf>
- [7] C. L. Prete, N. Guo, and U. V. Shanbhag, “Virtual bidding and financial transmission rights: An equilibrium model for cross-product manipulation in electricity markets,” *IEEE Transactions on Power Systems*, vol. 34, no. 2, pp. 953–967, 2018.
- [8] J. E. Parsons, C. Colbert, J. Larrieu, T. Martin, and E. Mastrangelo, “Financial arbitrage and efficient dispatch in wholesale electricity markets,” *MIT Center for Energy and Environmental Policy Research*, February 2015. [Online]. Available: <http://cepr.mit.edu/files/papers/2015-002.pdf>
- [9] California Independent System Operator (CAISO), “2022 annual report on market issues and performance.” [Online]. Available: <https://www.caiso.com/Documents/2022-Annual-Report-on-Market-Issues-and-Performance-Jul-11-2023.pdf>



- [10] The White House, “Fact sheet: President Biden sets 2030 greenhouse gas pollution reduction target aimed at creating good-paying union jobs and securing U.S. leadership on clean energy technologies,” April 2022.
- [11] WattTime, “Marginal emissions modeling: WattTime’s approach to modeling and validation,” Oct 2022. [Online]. Available: <https://www.watttime.org/app/uploads/2022/10/WattTime-MOER-modeling-20221004.pdf>
- [12] S. Baltaoglu, L. Tong, and Q. Zhao, “Algorithmic bidding for virtual trading in electricity markets,” *IEEE Transactions on Power Systems*, vol. 34, no. 1, pp. 535–543, 2019.
- [13] W. Wang and N. Yu, “A machine learning framework for algorithmic trading with virtual bids in electricity markets,” in *2019 IEEE Power & Energy Society General Meeting (PESGM)*. IEEE, 2019, pp. 1–5.
- [14] B. Yu, H. Yin, and Z. Zhu, “Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting,” *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, no. 7, pp. 3634—3640, 2018.
- [15] R. Weron, “Electricity price forecasting: A review of the state-of-the-art with a look into the future,” *International Journal of Forecasting*, vol. 30, no. 4, pp. 1030–1081, 2014.
- [16] J. Contreras, R. Espinola, F. J. Nogales, and A. J. Conejo, “ARIMA models to predict next-day electricity prices,” *IEEE Transactions on Power Systems*, vol. 18, no. 3, pp. 1014–1020, 2003.
- [17] R. Angamuthu Chinnathambi, A. Mukherjee, M. Champion, H. Salehfar, T. M. Hansen, J. Lin, and P. Ranganathan, “A multi-stage price forecasting model for day-ahead electricity markets,” *Forecasting*, vol. 1, no. 1, pp. 26–46, 2019.
- [18] R. Weron, M. Bierbrauer, and S. Trück, “Modeling electricity prices: Jump diffusion and regime switching,” *Physica A: Statistical Mechanics and Its Applications*, vol. 336, no. 1, pp. 39–48, 2004.
- [19] N. P. Yu, H. Y. Sheng, and R. Johnson, “Economic valuation of wind curtailment rights,” in *2013 IEEE Power & Energy Society General Meeting (PESGM)*, July 2013, pp. 1–5.
- [20] H. Ebrahimian, S. Barmayoon, M. Mohammadi, and N. Ghadimi, “The price prediction for the energy market based on a new method,” *Economic Research-Ekonomska Istraživanja*, vol. 31, no. 1, pp. 313–337, 2018.
- [21] J. Mather, E. Bitar, and K. Poolla, “Virtual bidding: Equilibrium, learning, and the wisdom of crowds,” *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 225–232, 2017.
- [22] W. Tang, R. Rajagopal, K. Poolla, and P. Varaiya, “Model and data analysis of two-settlement electricity market with virtual bidding,” in *2016 IEEE 55th Conference on Decision and Control (CDC)*. IEEE, 2016, pp. 6645–6650.
- [23] C. K. Woo, J. Zarnikau, E. Cutter, S. Ho, and H. Leung, “Virtual bidding, wind generation and California’s day-ahead electricity forward premium,” *The Electricity Journal*, vol. 28, no. 1, pp. 29–48, 2015.

- [24] A. Attarha, N. Amjady, and S. Dehghan, “Affinely adjustable robust bidding strategy for a solar plant paired with a battery storage,” *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 2629–2640, 2019.
- [25] C. Bliet, P. Bonami, and A. Lodi, “Solving mixed-integer quadratic programming problems with IBM-CPLEX: A progress report,” 2014.
- [26] P. Artzner, F. Delbaen, J. Eber, and D. Heath, “Coherent measures of risk,” *Mathematical Finance*, vol. 9, no. 3, pp. 203–228, 1999.
- [27] R. T. Rockafellar and S. Uryasev, “Optimization of conditional value-at-risk,” *Journal of Risk*, vol. 2, pp. 21–41, 2000.
- [28] F. A. Gers, J. Schmidhuber, and F. Cummins, “Learning to forget: Continual prediction with LSTM,” *Neural Computation*, vol. 12, no. 10, pp. 2451–2471, 2000.
- [29] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD ’16, 2016, pp. 785–794.
- [30] J. H. Friedman, “Greedy function approximation: A gradient boosting machine,” *The Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001.
- [31] H. Drucker, C. J. C. Burges, L. Kaufman, A. Smola, and V. Vapnik, “Support vector regression machines,” in *Advances in Neural Information Processing Systems 9 (NIPS 1996)*. MIT Press, 1997, pp. 155–161.
- [32] W. Wang and N. Yu, “A machine learning framework for algorithmic trading with virtual bids in electricity markets,” in *2019 IEEE Power & Energy Society General Meeting (PESGM)*, 2019, pp. 1–5.
- [33] Y. Li, N. Yu, and W. Wang, “Machine learning-driven virtual bidding with electricity market efficiency analysis,” *IEEE Transactions on Power Systems*, 2021.
- [34] Q. Zhou, L. Tesfatsion, and C. Liu, “Short-term congestion forecasting in wholesale power markets,” *IEEE Transactions on Power Systems*, vol. 26, no. 4, pp. 2185–2196, 2011.
- [35] L. Min, S. T. Lee, P. Zhang, V. Rose, and J. Cole, “Short-term probabilistic transmission congestion forecasting,” in *2008 Third International Conference on Electric Utility Deregulation and Restructuring and Power Technologies*, 2008, pp. 764–770.
- [36] Y. Ji, R. J. Thomas, and L. Tong, “Probabilistic forecasting of real-time LMP and network congestion,” *IEEE Transactions on Power Systems*, vol. 32, no. 2, pp. 831–841, 2017.
- [37] G. Li, C. Liu, and H. Salazar, “Forecasting transmission congestion using day-ahead shadow prices,” in *2006 IEEE PES Power Systems Conference and Exposition*, 2006, pp. 1705–1709.
- [38] S. Uryasev, “Conditional value-at-risk: Optimization algorithms and applications,” in *Proceedings of the IEEE/IAFE/INFORMS 2000 Conference on Computational Intelligence for Financial Engineering (CIFEr)*, 2000, pp. 49–57.

- [39] R. T. Rockafellar and S. Uryasev, "Conditional value-at-risk for general loss distributions," *Journal of Banking and Finance*, vol. 26, no. 7, pp. 1443–1471, 2002.
- [40] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath, "Coherent measures of risk," *Mathematical Finance*, vol. 9, no. 3, pp. 203–228, 1999.
- [41] R. T. Rockafellar and S. Uryasev, "Optimization of conditional value-at-risk," *The Journal of Risk*, vol. 2, pp. 21–41, 2000.
- [42] N. Yu, A. Somani, and L. Tesfatsion, "Financial risk management in restructured wholesale power markets: Concepts and tools," in *2010 IEEE Power & Energy Society General Meeting (PESGM)*, 2010, pp. 1–8.
- [43] R. L. Thorndike, "Who belongs in the family?" *Psychometrika*, vol. 18, no. 4, pp. 267–276, 1953.
- [44] U.S. Environmental Protection Agency, "Sources of greenhouse gas emissions," August 2022. [Online]. Available: <https://www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions>
- [45] C. Marnay, D. C. Fisher, S. Murtishaw, A. A. Phadke, L. K. Price, and J. A. Sathaye, "Estimating carbon dioxide emissions factors for the California electric power sector," *Proceedings of the 2002 ACEEE's Summer Study on Energy Efficiency*, vol. 33, no. 99, 2002.
- [46] R. Bettle, C. H. Pout, and E. R. Hitchin, "Interactions between electricity-saving measures and carbon emissions from power generation in england and wales," *Energy Policy*, vol. 34, no. 18, pp. 3434–3446, 2006.
- [47] D. L. Shawhan, J. T. Taber, D. Shi, R. D. Zimmerman, J. Yan, C. M. Marquet, Y. Qi, B. Mao, R. E. Schuler, W. D. Schulze, and D. Tylavsky, "Does a detailed model of the electricity grid matter? Estimating the impacts of the regional greenhouse gas initiative," *Resource and Energy Economics*, vol. 36, no. 1, pp. 191–207, 2014.
- [48] A. Hawkes, "Estimating marginal  $CO_2$  emissions rates for national electricity systems," *Energy Policy*, vol. 38, no. 10, pp. 5977–5987, 2010.
- [49] K. Siler-Evans, I. L. Azevedo, and M. G. Morgan, "Marginal emissions factors for the U.S. electricity system," *Environmental Science & Technology*, vol. 46, no. 9, pp. 4742–4748, 2012.
- [50] D. S. Callaway, M. Fowlie, and G. McCormick, "Location, location, location: The variable value of renewable energy and demand-side efficiency resources," *Journal of the Association of Environmental and Resource Economists*, vol. 5, no. 1, pp. 39–75, 2018.
- [51] L.-N. Guo, C. She, D.-B. Kong, S.-L. Yan, Y.-P. Xu, M. Khayatnezhad, and F. Gholinia, "Prediction of the effects of climate change on hydroelectric generation, electricity demand, and emissions of greenhouse gases under climatic scenarios and optimized ANN model," *Energy Reports*, vol. 7, pp. 5431–5445, 2021.
- [52] L. Alfaseeh, R. Tu, B. Farooq, and M. Hatzopoulou, "Greenhouse gas emission prediction on road network using deep sequence learning," *Transportation Research Part D: Transport and Environment*, vol. 88, p. 102593, 2020.

- [53] M. S. Bakay and Ümit Ağbulut, “Electricity production based forecasting of greenhouse gas emissions in Turkey with deep learning, support vector machine and artificial neural network algorithms,” *Journal of Cleaner Production*, vol. 285, p. 125324, 2021.
- [54] M. Emami Javanmard and S. Ghaderi, “A hybrid model with applying machine learning algorithms and optimization model to forecast greenhouse gas emissions with energy market data,” *Sustainable Cities and Society*, vol. 82, p. 103886, 2022.
- [55] J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun, “Graph neural networks: A review of methods and applications,” *AI open*, vol. 1, pp. 57–81, 2020.
- [56] T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” in *2017 International Conference on Learning Representations (ICLR)*, April 2017, pp. 1–11.
- [57] D. Owerko, F. Gama, and A. Ribeiro, “Optimal power flow using graph neural networks,” in *ICASSP 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 5930–5934.
- [58] A. M. Karimi, Y. Wu, M. Koyuturk, and R. H. French, “Spatiotemporal graph neural network for performance prediction of photovoltaic power systems,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 17, 2021, pp. 15 323–15 330.
- [59] W. Wang and N. Yu, “Estimate three-phase distribution line parameters with physics-informed graphical learning method,” *IEEE Transactions on Power Systems*, vol. 37, no. 5, pp. 3577–3591, 2022.
- [60] Q. Yang and A. Sadeghi, “Graph convolutional networks for power system state estimation,” in *2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, 2020.
- [61] K. Yamashita, J. Qin, N. Yu, E. Farantatos, and L. Zhu, “Predicting power system voltage health index with PMUs and graph convolutional networks,” in *2023 IEEE Power & Energy Society General Meeting (PESGM)*, 2023.
- [62] California Independent System Operator (CAISO), “Open access same-time information system (OASIS).” [Online]. Available: <http://oasis.caiso.com/mrioasis/logon.do>
- [63] A. Hundiwale, “Greenhouse gas emission tracking methodology,” California Independent System Operator, Tech. Rep., 2016. [Online]. Available: <http://www.caiso.com/Documents/GreenhouseGasEmissionsTracking-Methodology.pdf>
- [64] California Independent System Operator (CAISO), “California ISO hits all-time peak of more than 97% renewables,” April 2022. [Online]. Available: <https://www.caiso.com/Documents/California-ISO-Hits-All-Time-Peak-of-More-Than-97-Percent-Renewables.pdf>

# Appendix A

## Appendix for Chapter 2

### A.1 Proof of (2.17)

The sum among hours ( $\sum_{h=1}^{24}$ ) is omitted as it does not affect the proof mathematically.

$$\begin{aligned}
& \sum_{i=1}^N (z_{i,h}^I E[r_{i,h}^I(\mathbf{z}_h)] + z_{i,h}^D E[r_{i,h}^D(\mathbf{z}_h)]) \\
&= \sum_{i=1}^N z_{i,h}^I E[(\lambda_{i,h}^{dif} - \gamma^I)] + \sum_{i=1}^N z_{i,h}^I E[\sum_{j=1}^{M_h} (a_{j,h} x_h + b_{j,h}) d_{j,h}] + \sum_{i=1}^N z_{i,h}^D E[(-\lambda_{i,h}^{dif} - \gamma^D)] \\
&\quad - \sum_{i=1}^N z_{i,h}^D E[\sum_{j=1}^{M_h} (a_{j,h} x_h + b_{j,h}) d_{j,h}] \\
&= \sum_{i=1}^N z_{i,h}^I E[(\lambda_{i,h}^{dif} - \gamma^I)] + \sum_{i=1}^N z_{i,h}^D E[(-\lambda_{i,h}^{dif} - \gamma^D)] \\
&\quad + (\sum_{i=1}^N z_{i,h}^I - \sum_{i=1}^N z_{i,h}^D) E[\sum_{j=1}^{M_h} (a_{j,h} x_h + b_{j,h}) d_{j,h}] \\
&= \sum_{i=1}^N z_{i,h}^I E[(\lambda_{i,h}^{dif} - \gamma^I)] + \sum_{i=1}^N z_{i,h}^D E[(-\lambda_{i,h}^{dif} - \gamma^D)] + x_h E[\sum_{j=1}^{M_h} (a_{j,h} x_h + b_{j,h}) d_{j,h}] \\
&= \sum_{i=1}^N z_{i,h}^I E[(\lambda_{i,h}^{dif} - \gamma^I)] + \sum_{i=1}^N z_{i,h}^D E[(-\lambda_{i,h}^{dif} - \gamma^D)] + E[\sum_{j=1}^{M_h} (a_{j,h} x_h^2 + b_{j,h} x_h) d_{j,h}]
\end{aligned}$$