A HYBRID MODEL FOR IMPROVING SPATIOTEMPORAL RESOLUTION IN MARGINAL EMISSION FACTOR ESTIMATES

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

Arjun Malhotra

August 2023

ABSTRACT

Electricity generation is a significant source of global carbon emissions and a major driver of climate change. In the interest of mitigating climate change impacts, there is substantial interest in minimizing the carbon footprint associated with energy generation through demand-side shifts and supply-side transitions to cleaner power sources. Effective implementation of these measures requires a thorough spatiotemporal understanding of marginal emission factors (MEFs) that are associated with the generators that respond to changes in electricity demand. In this study, we present a hybrid model that estimates MEFs in real-time and forecasts hour-ahead values at a zonal level. The model proposed combines the capabilities of a dispatch model to capture the physical characteristics of the electricity grid with an advanced machine learning algorithm to project these values in the future. This analysis focuses on the NYS electricity grid, based on a publicly available digital twin of the NYISO. The proposed model provides insights into seasonal and diurnal patterns of zonal MEFs and the expected future behavior to support informed decision-making toward reducing carbon emissions.

BIOGRAPHICAL SKETCH

Arjun Malhotra was born in Uttar Pradesh, India in 1996. He received his bachelors in Chemical Engineering from National Institute of Technology, Hamirpur in 2019. In August of 2021, he moved to Ithaca, New York to pursue Master of Science in Chemical Engineering at Cornell University. At Cornell, he resonated well with the research carried out by Professor C. Lindsay Anderson and Prof. Jefferson W. Tester and joined their group. Since then, he has spent the last two years focusing on problems involving power system engineering, computational and data-driven modeling, and efficient grid management strategies.

To my family without whose constant support this work could not have been possible

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisor, Professor C. Lindsay Anderson, for her invaluable support and guidance throughout the duration of this challenging and rewarding endeavor. I am truly privileged to have had her belief in my abilities, especially considering my lack of formal background in power systems. I still recall the day she sat down with me and a piece of paper and pen to explain the fundamental concepts of energy and demand, which I admire to this day. In addition, she meticulously evaluated my written work and provided exceptional insights and expert feedback, allowing me to present my research in the most effective manner possible. The dedication and mentoring of Professor Anderson have been instrumental in shaping my academic journey at Cornell University.

I am also grateful to Prof. Jefferson W. Tester for embracing my research idea and supporting my pursuit outside the chemical engineering department while ensuring my inclusivity with the CBE department. Despite his busy schedule, he went above and beyond to address my requests with the department promptly. His visionary perspective has consistently inspired me to consider the broader implications of my work.

Furthermore, I extend my heartfelt appreciation to my fellow lab members, Vivienne, Shikhar, and Gerald. Vivienne, in particular, has been an invaluable guide, consistently steering me in the right direction when faced with research challenges. Her expertise, attention to detail, and unwavering commitment to this project have played an immeasurable role that words cannot adequately express. I have always admired her meticulous work style and high standards, which have motivated me to continually strive for improvement. Additionally, Vivienne has been a trusted friend with whom I could share not only project-related concerns but also personal struggles. Shikhar and Gerald have also been instrumental in supporting me with their expertise and motivation, consistently pushing me in the right direction. I am especially grateful to Shikhar for devoting his weekends to helping me with my research.

I would like to acknowledge and extend my gratitude to all my friends who have made the past few years both enjoyable and memorable. Nandini, thank you for always having my back and serving as a constant source of inspiration during challenging times. Your belief in me has propelled me to go the extra mile and turn difficult situations into opportunities for growth. To my roommate, Partha, thank you for not only creating a home away from home with your exceptional cooking skills in Indian cuisine but also for fostering a deep friendship that has been a source of comfort throughout my journey. I want to express my sincerest appreciation to my family for their constant support. To my mom, thank you for your unconditional love. To my dad, thank you for always instilling in me positivity and helping me keep my cool throughout this journey. And to my sister, thank you for having the belief in me.

To all the individuals mentioned here and many others who have extended their help, I will always be grateful. In the future, I genuinely hope to have the opportunity to reciprocate your generosity and aim to live up to the example you have set through your selflessness and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Climate change due to increased anthropogenic greenhouse gas (GHG) emissions has become a significant and urgent challenge to present-day society. There have been collaborative global initiatives to reduce GHG emissions on a worldwide scale in order to address and mitigate the consequences of climate change. COP27, held in Sharm el-Sheikh, Egypt, represents the latest international effort by nations to collectively tackle the objective of attaining net-zero global emissions by 2050 and limiting global warming. However, these COP meetings have had little to no impact on the reduction of GHG emissions as governments around the world fail to take measures such as phasing out fossil fuels, halting deforestation, and expanding renewable energy [4]. Simultaneously, New York State's Community Leadership and Climate Protection Act (CLCPA) [5] is among the most ambitious climate legislations globally. In summary, the CLCPA requires the state to cut GHG emissions by 40% by 2030 and no less than 85% by 2050, relative to 1990 levels. Electricity generation is the second-largest source of GHG emissions after transportation, accounting for 25% of the emissions in the US [6]. To date, a significant proportion of electricity comes from fossil-fueled power plants; for instance, in New York State, 41% of the electricity generated is from thermal power plants [7], highlighting the need to manage emissions from electricity generation.

To address the problem of GHG emissions from electricity generation, it is critical to understand the factors that contribute to these emissions and to develop strategies to mitigate them. The implementation of both supply-side and demand-side measures will lead to the displacement of emissions generated by conventional sources. There is an increasing inclination on the supply side to incorporate intermittent renewable energy sources, such as wind and solar, which exhibit variable output ranging from short-term fluctuations to seasonal variations. The concept of demand response on the demand side requires an increased awareness of the impact of demand shifting on carbon emissions. As policymakers evaluate various options, such as renewable energy integration, smart grid initiatives, and demand response systems, it becomes evident that there is a need to calculate the emissions that can be avoided through the implementation of these measures. Marginal emissions factors (MEFs) provide a standardized method to quantify the avoided emissions that are linked with these improvements [8]. However, MEFs vary by time and geographical location, making it essential for policymakers and practitioners to assess their spatiotemporal heterogeneity to understand the associated environmental impacts.

Generation and consumption of electricity are coordinated in the United States through regional energy markets administered by regional transmission organizations (RTOs) or independent system operators (ISOs). RTOs and ISOs hold ownership rights to their system information, and most of that information is kept private due to confidentiality and security concerns. Information on marginal generators is one such piece of information that is not available to the public, making it difficult to make informed decisions to reduce the carbon footprint of electricity generation [9].

Various studies have been conducted in the field of power systems to estimate MEFs using different methods. The methodologies can be classified into two fundamental categories: physics-based power system models and statistical models [10]. The majority of the power systems-based methods employ dispatch models that assume that generators are dispatched based on their marginal cost of operation, and the last generator that responds to a load change determines the marginal emission factor for the region under study [11]. These studies, however, require difficult-to-obtain operational data for the power grid, and they typically do not estimate MEFs taking regional variations into account, limiting the ability to make informed decisions at the local level. Statistical methods, on the other hand, rely on historical data to make estimations based on the patterns present in the data The inability of the data-driven methodology to account for the physical and operational constraints of the power system leads to inaccurate MEF estimates.

Considering these limitations, this research proposes a hybrid model that combines a dispatch model with a machine learning algorithm to estimate and forecast hourly MEFs in the NYS electricity grid at a zonal level. By combining the detailed operational information provided by dispatch model with the forecasting capabilities of the machine learning model, the hybrid model will provide a more accurate and complete understanding of the MEFs at an improved spatiotemporal resolution.

1.1 Organization of the Thesis

The rest of this thesis is organized as follows:

Chapter 2: This chapter begins with an introduction of MEF and its significance in electricity grid management. Difference between MEF and Average Emission Factor (AEF) is discussed, emphasizing the significance of MEF in accurate quantification of avoided emissions. In addition, the existing methodologies for estimating MEFs are presented along with their merits and demerits. In the end, both the need for a hybrid model to estimate MEFs in the electricity grid and the overview of the New York State (NYS) electricity grid in 2019 are presented.

Chapter 3: In this chapter, the data and methodology of the proposed hybrid model are discussed. The chapter begins with an overview of the open-source digital NYS state electricity grid model that is used to inform the dispatch model. Next, the modeling details of the dispatch model used to identify marginal generators and the method for calculating MEFs are presented. Finally, the training of the machine learning algorithm used to predict MEF values is described.

Chapter 4: In this chapter, simulated MEFs are assessed for diurnal, seasonal, and inter-zonal patterns and compared to a NYSERDA study to ensure they accurately represent historical values. Subsequently, marginal generators are analyzed for interzonal variations. Finally, performance of the forecasting model is evaluated.

Chapter 5: This chapter concludes the thesis with a detailed case study of one of the potential applications, a discussion of potential future work, and a summary of the research findings and contributions.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter introduces the concept of MEF and emphasizes its importance in the effective management of electricity grids. Fundamental differences between MEF and AEF are discussed, highlighting the significance of MEF in accurately quantifying avoided emissions. In addition, methodologies currently used to estimate MEFs are elaborated, analyzing their advantages and disadvantages. Few recent approaches used for estimating emission factors in the NYS electricity grid are also discussed. This analysis enables us to identify the gap in the existing literature, facilitating the development of an improved model. Recognizing the complexities and nuances of accurately quantifying MEFs, we advocate for a combined approach that capitalizes on the strengths of dispatch models and machine learning algorithms. This hybrid model enables grid administrators, consumers, and policymakers to obtain more reliable and accurate MEF estimates at an improved spatiotemporal resolution, allowing them to make informed decisions and develop effective emission reduction strategies. Finally, we provide an overview of the unique characteristics and challenges of the NYS electricity grid, establishing a context for the subsequent application of our proposed model to estimate and forecast MEFs for this grid. Understanding the unique characteristics of the NYS electricity grid enables us to tailor our insights and recommendations to this region, resulting in more targeted and effective actions toward sustainable energy management.

2.2 Power System and Marginal Emission Factor

The power system is the backbone of modern society, providing a reliable and continuous electricity supply to households, commercial, and industrial buildings. Power systems are complex networks of generation, transmission, and distribution processes that work together to provide a continuous electricity supply to consumers. The processes involved in power systems are highly interconnected and dynamic.

The MEFs quantify the carbon intensity of electricity generation. The MEF is defined as the change in carbon emissions per unit of electricity produced when the demand for electricity increases by one unit [10]. In other words, it represents the carbon emissions associated with the marginal unit of electricity produced.

The MEF is influenced by a range of factors, including electricity demand, the time of day, grid location, energy mix, regulations, and renewable energy integration. For example, the MEF is typically higher during periods of high demand when peaking power plants with higher carbon intensity are dispatched to meet the increased demand [12]. Similarly, the MEF is higher in regions where the mix of generation sources is dominated by fossil fuels, such as coal and natural gas. These factors can result in significant differences in the MEF across different regions of the power system and timeframes, highlighting the need for more localized MEF estimation models at a fine temporal resolution.

2.3 MEF's Significance in Electricity Grid Management

MEF is a key metric that can be used in effective electricity grid management. As MEF considers the specific generators that are being used to meet changes in electricity demand at a particular point in time and location, it provides an accurate measurement of the carbon emissions associated with electricity generation in real-time. Therefore, MEFs can help make informed decisions both from supply and demand side.

2.3.1 Storage Optimization

Estimating MEFs hold importance in energy planning as it helps utilities to identify the ideal combination of energy technologies to fulfill energy requirements while minimizing the release of greenhouse gases. For example, utility companies can strategically charge their storage units during periods of low emission intensity. They can subsequently utilize this stored clean energy during periods when emission factors are high, effectively managing and reducing overall emissions [10].

2.3.2 Demand Shifting

Having knowledge of MEFs can assist consumers in shifting their energy consumption when a less carbon-intensive generator is available. During periods of elevated MEFs, consumers can refrain from utilizing energy-intensive appliances and equipment. This can help reduce the grid's overall emissions and obtain other incentives [13].

2.3.3 Effective Investment Decisions and Policy Design

MEFs are crucial in designing energy policies and investment decisions that aim to reduce greenhouse gas emissions. By understanding the impact of different energy technologies on the MEFs and the economic benefits of investing in various renewable technologies, policymakers can determine the most effective policies for reducing emissions while minimizing the cost of the energy system [14].

However, one of the challenges in managing electricity grids using MEF is the variability of the MEFs over time and location. The MEFs can change rapidly based on changes in electricity demand, operating cost, and the grid's availability of different generation sources. As a result, real-time monitoring, and accurate forecasting of the MEFs are critical for effective grid management.

2.4 MEFs versus AEFs

MEFs and AEFs are the two metrics commonly used to calculate carbon emissions from the electricity sector, especially in electricity generation [10]. They play a critical role to inform decisions related to energy supply, demand, and policymaking.

MEFs represent the incremental change in CO_2 emissions corresponding to an incremental change in electricity demand at a particular location and time. MEFs are particularly important in quantifying the emission changes resulting from interventions in the electricity system such as renewable energy integration and demand shifts.

AEFs, on the other hand, assume that any changes in demand are met with proportional responses from all generating units. AEFs are calculated by dividing the total CO_2 emissions from electricity generation by the total electricity generated for a specified location and time. Therefore, AEFs reflect the average emissions of all regional power plants. They are commonly used to estimate emissions for policy analysis aimed at emission reduction and carbon accounting [10].

To properly assess the electricity generation patterns, it is crucial to analyze the hourly demand as it tends to vary throughout the day. The level of demand determines the dispatching of generation plants. Renewable generators such as nuclear, hydro, and wind are typically dispatched first because they have low operating costs. However, fossil fuel generators are dispatched when the demand exceeds the renewable plants' capacity. It is important to note that dispatching practices vary based on the plant's availability and fuel prices in a particular location.

The marginal power plant is the last one dispatched to meet demand at a given time. If consumption were to change during this period, the resulting emissions change would come from this marginal plant. During the summer months in New York City, for instance, peak-load generators produce more emissions than those used during other regular hours. Generally, natural gas combined cycle plants are used to satisfy the base load demand in New York City, whereas steam turbines are utilized to accommodate the normal fluctuations. However, combustion turbines are dispatched during the summer to satisfy the increased electricity demand. Since these older peak plants are infrequently utilized and generate so little energy annually, they are exempt from specific emission regulations. To account for emissions from marginal plants, such as these older peak plants in New York City, the marginal emissions factor should be used rather than the average emissions factor. [15].

If a power grid intervention in New York City used average emission factors to calculate the total emissions avoided, the results could be misleading. AEFs evaluate the avoided emissions from all New York City power plants, with natural gas and a small amount of oil used to meet the city's peak electricity demand. In the AEF, avoided emissions from natural gas plants bear more weight than avoided emissions from oil plants. Nonetheless, if the oil plant is the marginal plant, it would reduce its electricity output and avoid emissions from the intervention [15]. The MEF would only consider the offset emissions from this marginal oil plant. As the AEF approach emphasizes avoiding emissions from natural gas, whereas the MEF approach considers avoiding emissions from oil, the outcomes of the two approaches will differ substantially.

Thus, it is noticed from the above discussion that MEFs are a more accurate representation of avoided emissions than AEFs; therefore, we focus on estimating MEFs for our present analysis, which is essential for effective grid management and emission reduction in the power sector. In the following section, an example calculation illustrating the difference between MEF and AEF CO_2 emission values is presented.

2.4.1 Calculation of CO_2 Emissions using AEF Method

To illustrate how significantly the outcomes of the two approaches can differ, let us consider a very simple example. Imagine an electricity grid having a generation mix of 50% wind and 50% natural gas. The wind plant produces no CO_2 emissions, while the natural gas plant produces 2,000 pounds of CO_2 per MWh. If an energy efficiency improvement resulted in avoiding 1 MWh of energy, the AEF approach would compute the avoided emissions to be 1000 pounds by considering a proportional response from both generators (Equation 2.1).

$$\begin{aligned} CO_2 \text{ emission avoided} &= (1 \text{ MWh} \times 0.5 \times 0 \text{ lb/MWh}) \\ &\quad + (1 \text{ MWh} \times 0.5 \times 2,000 \text{ lb/MWh}) \\ &= 1000 \text{ lb} \end{aligned} \tag{2.1}$$

2.4.2 Calculation of CO_2 Emissions using MEF Method

However, since the wind plant has a lower operating cost, it would be dispatched first to meet demand. Once the demand exceeds the wind plant's capacity, the natural gas plant would be dispatched and considered the marginal generator. If the natural gas plant is the marginal generator, the MEF approach will estimate the CO_2 emissions avoided to be 2000 pounds, as all the demand reductions would occur at the natural gas plant (Equation 2.2).

$$
\begin{aligned}
CO_2 \text{ emission avoided} &= (1\text{ MWh} \times 0 \times 0\,\text{lb}/\text{MWh}) \\
&\qquad + (1\text{ MWh} \times 1 \times 2,000\,\text{lb}/\text{MWh}) \\
&= 2,000\,\text{lb}
\end{aligned} \tag{2.2}
$$

Additionally, it could be the case that a hydropower plant is the marginal plant, and in that case, the marginal emission factor would be zero. Thus, MEFs are a more precise method of determining avoided emissions. This example shows that the outcomes can vary significantly between these two methods [16].

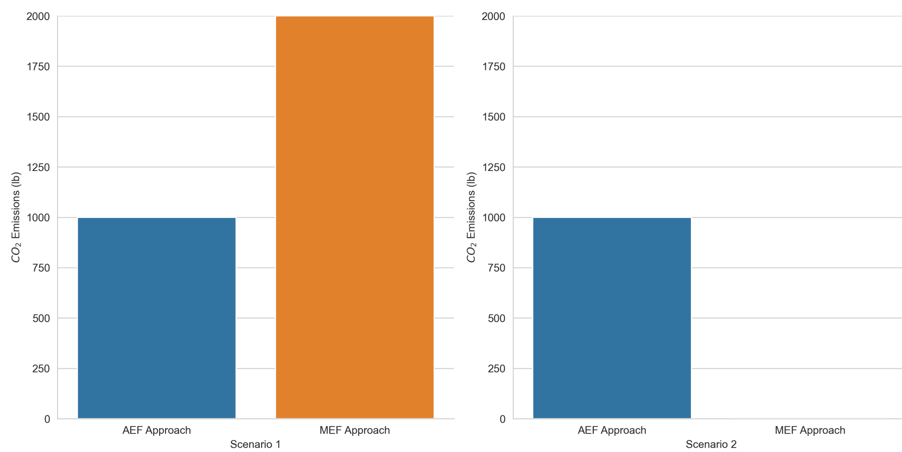

Figure 2.1: MEF vs AEF CO_2 emission comparison under different scenarios

2.5 Modeling Approaches for MEF with their Merits and Demerits

Various methods have been proposed for MEF estimation and forecasting, broadly including power system-based dispatch models and statistical methods. Each of these methods has its benefits and limitations. The choice of method depends on the type of data available, the objective of the study, and the computational resources available. This section provides an in-depth analysis of all these methods and their relative merits and demerits.

2.5.1 Power System-Based Modeling Approach

Power system-based modelling methods for MEF estimation simulate the operation of the power system and provide information on the emissions associated with each unit of energy added to the grid. Power system modeling-based methods primarily include merit order dispatch models. These models decide the dispatch of power plants to minimize the cost of supplying electricity to the system subject to transmission constraints and operating conditions of the grid. These models can provide information on the MEF at different times of the day or under different scenarios.

Raichu et al. [17] use an electricity dispatch model to estimate marginal emissions in the ERCOT and NYISO regions. The model considers hourly load, information about generators, and operating constraints such as plant maintenance, fuel changing, outages, additional unused capacity information, seasonal capacities, and seasonal hydro availability. The model is able to generate an approximation of the hourly electricity generation by each power plant in the network when these inputs are provided. This enables the estimation of marginal resource use as well as emissions. However, this study has some limitations, as it does not consider wind and solar integration in the generation mix, does not model transmission constraints and ramping limits, and requires a large amount of operational data that may not be easily available.

Several other studies based on power system-based methods have utilized energy system tools that are either not transparent, i.e., have algorithms that are not publicly available, or require a significant learning curve to use effectively. One such tool is the EnergyPLAN™ model, as described in [18], which is a proprietary software based on an analytical framework that identifies the marginal impacts of demand-side interventions in an energy system. The modeled data obtained through EnergyPLAN can then be used to calculate marginal emissions. These limitations of proprietary tools and models underline the importance of developing open-source representations of electricity grids that facilitate the identification of marginal units for more transparent and accessible marginal emissions estimation.

Azevedo et al. [19] present a reduced-order dispatch model for modeling MEFs in the US electricity sector. The model considers the hourly variability of electricity demand and power plant availability to minimize the cost of electricity production while meeting demand. The model is validated with historical generation and emissions data in the US power sector for 2014–2017. The study focuses on simulating MEFs at the North American Electric Reliability Corporation (NERC) regional level; however, the paper does not consider the variability of MEFs within a region, which could be significant due to differences in the types of power plants and fuel sources used in different areas. A finer spatial resolution allows for a more accurate representation of the power grid and its emissions profile, which could be useful for identifying opportunities to reduce emissions at a more localized level.

Power system modeling-based methods have the following advantages:

- They can provide detailed information on the operation of the electricity grid.
- They can be used to simulate different future-generation scenarios.

However, power system modeling-based methods have the following disadvantages:

- They require detailed data on the power system, such as the operating characteristics of power plants, transmission line data, etc., which can be challenging to obtain.
- They also require significant computational resources and can be time-consuming.

2.5.2 Statistical-Based Modeling Approach

Statistical-based methods estimate the MEF based on historical data. These methods use statistical models to establish the relationship between carbon emission and other explanatory variables such as demand, price, weather, and fuel mix. The detailed critical analysis of the above methods and their strengths and weaknesses is discussed below:

Siler-Evans et al. [11] use a linear regression-based approach to estimate hourly MEFs for different NERC regions. The slope of the linear regression between the change in emissions and generation for each time step is taken as an approximation of MEFs. Li et al. [20] and Thind et al. [21] extend this approach by incorporating the impacts of renewable energy sources, such as wind and solar generation, with a specific focus on the subregions of the Midcontinent Independent System Operator (MISO) system. The strengths of these approaches are their ability to provide detailed analyses of MEFs for specific subregions and incorporate the impacts of renewable energy sources. However, these regression methods assume a linear relationship between generation and emissions, which is not true in complex electricity grids with physical constraints. These methods highlight the need for more comprehensive and dynamic modeling approaches.

Additionally, there are other advanced statistical methods that have been used recently to capture the complex relationships and non-linearities in the data, making them particularly useful for MEF forecasting. Several machine learning algorithms have been applied to MEF estimation, including Support Vector Machines (SVM), ARIMA with exogenous inputs (ARIMAX), and Artificial Neural Networks (ANN) methods. Wang et al. [2] used an SVM-based method to estimate the hourly marginal emissions for the PJM market area in real time. The authors used data from various sources, including weather, electricity demand, and generation, to develop a machine-learning model that predicts MEFs. The paper's merit lies in its ability to capture the complex relationships between MEF and other explanatory variables to estimate emissions in real time. However, the accuracy of the proposed model depends on the quality of the underlying data used to train the model. For example, the model relies on weather forecasts to predict MEFs, which may not always be accurate, especially for extreme weather events. Additionally, the SVM method assumes that the data can be separated by a linear boundary. If the data is not linearly separable, the SVM may not be able to classify or predict the output accurately.

Leerbeck et al. [22] employed ARIMAX to forecast the CO_2 emission intensity in power grids on a short-term basis. The algorithm was trained on a dataset of 473 explanatory variables, including power production, demand, imports, and weather conditions. The paper's strength lies in its ability to provide forecasts at different time horizons (1–24 hours) for making informed decisions and reducing emissions. Nonetheless, the study's major limitation is that the MEF data used for training the machine learning model is acquired from a third-party (electricitymaps.org) that does not consider the physical limits of the electricity grid, raising questions on the quality of the MEF data used for training the forecasting model. Maji et al. [23] presented a neural network-based machine learning approach for forecasting the average emission factor for the following day using historical electricity generation data and future weather forecasts. The paper's strengths include the use of multiple data sources and a comprehensive validation process for evaluating the model's accuracy compared to other state-of-the-art techniques. However, the study forecasts average emissions instead of marginal emissions and does not consider electricity flow from neighboring regions. While these techniques have advantages in capturing complex relationships and non-linearities, they also have limitations such as sensitivity to input data, requirements for large amounts of training data, and ignorance of the physical characteristics of the electricity network.

Statistical-based methods have some advantages:

- By relying on historical data, these studies avoid the complexity of using power flow models to determine generator dispatch based on transmission constraints and operating conditions.
- Statistical models require relatively lesser amounts of operational data that

can be difficult to obtain compared to a dispatch model.

- They are less computationally intensive than the power system modeling-based methods.
- Statistical models can be effectively used to make short-term forecasting of MEFs.

The disadvantages of statistical models include the following:

- Statistical models are only as good as the quality of the data used to train the models.
- Statistical models do not capture the operating constraints of the power system.

Overall, the statistical models require lesser operational data, but they have some limitations, such as their inability to capture energy systems' complex and dynamic nature. In the following section we will investigate a few studies conducted specifically on NYS electricity grid.

2.5.3 Modeling Approach for the NYS electricity grid

The few studies that have focused on NYS have had a limited geographical scope or used proprietary software, which restricts the adoption and flexibility of the models employed. For instance, B. Howard et al. [24] estimated average emission factors for NYS but only evaluated the marginal emissions for New York City, thereby limiting their findings' applicability to understanding MEFs for other load zones in NYS. Similarly, the latest report by NYSERDA [3] predicts MEFs by aggregating load zones into upstate (A–E) and downstate (F–K), not providing the individual MEFs for each zone. Additionally, they utilized production simulation outputs given by Siemens for NYSERDA through their PROMOD model analyses, which included price forecasts and generation data [25]. Therefore, it is important to note that previous research has encountered limitations due to the lack of a zonal model capable of providing MEF estimates with a high degree of spatiotemporal precision. Using a comprehensive dispatch model to effectively capture the spatial and temporal variations, thereby facilitating the estimation of real-time hourly zonal MEFs, and then employing a Machine Learning (ML) model to project these MEFs into future periods can be considered the most effective approach.

Review of the literature shows that there is a need for a zonal-resolution model that accurately describes both spatial and temporal variations in MEFs, considers the physical characteristics of the electricity grid, and captures the relationship between explanatory variables like electricity price, demand, and fuel mix. This project aims to address this gap in the existing literature. It is the first study to develop a hybrid model that combines the power system model with a machine learning algorithm to estimate and forecast MEFs at a zonal level in the NYS electricity grid. The power system model allows for an accurate assessment of marginal units in different load zones at an hourly temporal resolution. It also provides quality training data for a machine learning algorithm to predict these values in the future. Our objective is to provide an assessment of generator units that respond to an incremental increase in demand across different load zones, calculate hourly zonal MEFs, and predict hour-ahead values to guide informed decision-making to reduce carbon emissions.

In the following section we provide an overview of the NYS electricity grid IN 2019 that will be used to demonstrate the functionality of the proposed model. Understanding the grid architecture, its generation, and load patterns across zones is critical for estimating and forecasting MEFs with sufficient resolution for decision-making.

2.6 Description of the NYS Electricity Grid

The NYS electricity grid is divided into eleven load zones (A-K) and has connections to four neighboring grids, including Hydro Quebec, Ontario (IESO), ISO-New England, and PJM [2]. The interconnection of these zones and neighboring grids enables the transfer of power between them, supporting a secure and reliable power supply for consumers. The NYS electricity grid delivers reliable and affordable electricity to more than 19 million residents and businesses. The grid includes generation facilities, transmission and distribution lines, and customer interconnections.

Transmission lines are high-voltage lines that transport electricity from generation facilities to distribution substations. At distribution stations, the voltage is stepped down for, distribution lines that deliver electricity to the end customers. The customer interconnections in NYS are made up of residential, commercial, industrial, and government consumers. The customers are connected to the distribution lines through transformers that step down the voltage to a level suitable for homes and businesses. To ensure the reliability of the NYS electricity grid, the New York Independent System Operator (NYISO) operates the wholesale electricity market, which coordinates the supply and demand of electricity in the state.

The generation facilities in NYS in 2019 consisted of a diverse mix of energy sources, including thermal (natural gas, kerosene, fuel oil 2, fuel oil 6, and coal), nuclear, hydro, wind, and other renewables (solar, storage, methane, refuse, or wood). The NYS generation mix for 2019 consisted of approximately 41% thermal, 32% nuclear, 22% hydro, 3% wind, and 2% other renewable [2]. This mix of generation technologies requires a robust and flexible grid system to ensure that electricity demand is met consistently and reliably.

Additionally, the generation mix is different for upstate zones (A-E) compared to downstate zones (F-K) and is referred to as the "Tale of Two Grids" by NYISO, as shown in Figure 2.2 below. Nearly 90% of the energy generated in upstate zones is from zero-emission sources, whereas in downstate zones, 70% of the electricity comes from CO_2-emitting thermal power plants [1]. The NYS grid also has an imbalance between generation and demand across different load zones. The upstate zones have more generation capacity than demand, whereas the downstate zones have higher demand than generation capacity. This leads to power flowing from upstate to downstate.

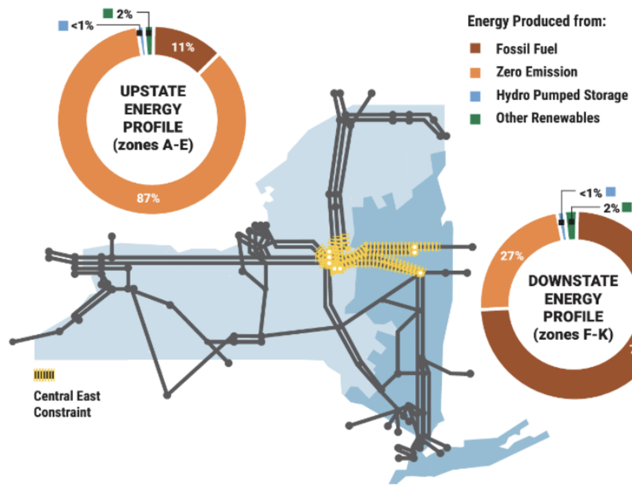

Figure 2.2: In 2019, 90% of upstate electricity was carbon-free, while 70% of downstate electricity came from CO_2-emitting thermal power plants. "Zero Emission" energy came from nuclear, wind, and hydro plants. [1]

As we can see from the above discussion, estimating MEF for the NYS electricity grid is challenging due to its complex and dynamic nature. The NYS grid has a diverse mix of energy sources that are distributed unevenly across different load zones. There is also the potential for a mismatch between generation and demand, and interaction with neighboring grids makes the estimation more challenging. Therefore, we need a model that captures these inter-zonal variations effectively and provides MEF estimates for each load zone.

CHAPTER 3

DATA AND METHODS

3.1 Introduction

This chapter describes the data and methodology used to develop the proposed hybrid model. This chapter is a helpful resource for understanding the development of the proposed model. First, we provide a brief description of the data and modeling utilized in the open-source digital NYS electricity grid model, which serves as the basis for our dispatch model. This grid model is a reliable and accurate representation of the electricity grid in New York State. Next, the modeling details of our dispatch model are discussed, which has been developed to identify marginal generators within the electricity grid. This dispatch model optimizes the allocation of generation units subject to power system constraints. In addition, we explain how MEFs, a crucial metric for quantifying the environmental impact of various power generation sources, are determined. Following this, a comprehensive explanation of the training procedure for the machine learning algorithm used in our hybrid model is provided. This algorithm is utilized to predict MEF values by considering a number of explanatory variables. The architecture and design of the algorithm, as well as the training methodology, which involves the selection and preparation of training data, are discussed. This chapter establishes the groundwork for subsequent chapters that present the results, analysis, and implications of our hybrid model within the context of the NYS electricity grid.

3.2 Overview of the Open-Source NYS Electricity Grid Model

Using publicly available data, a representational model of the electricity system in NYS is built that incorporates major physical attributes of the actual grid. [2]. This section outlines the data and modeling steps performed in [2] to construct the grid model, as the data collected and processed for this model serves as the foundation for our dispatch model to simulate the NYS grid.

3.2.1 NPCC System Dataset

The NPCC 140-bus system is a standard test case [26] of the Eastern Interconnection that comprises representations, either full or partial, of five ISOs: NYISO (full), IESO (partial), PJM ISO (partial), NE ISO (partial), and MISO (partial) [27]. The system consists of 140 buses, 233 transmission lines, and 48 generators. This network provides benchmark generation and load data to perform optimal power flow study and location of all the buses to integrate spatial relationships. Therefore, this network represents the NYS grid and keeps significant interconnections with surrounding grids, allowing NYS and interactions with other ISOs/RTOs to be modeled [2].

3.2.2 NYS Generation Data

NYISO provides hourly generation data for several fuel types [28]. However, there is no access to generator hourly profiles required for performing power flow analysis. [2] estimated generation profiles for all generators using NYISO's 2019 Load & Capacity Data Report [29]. [2] Fitted cost curves with changing fuel costs to perform optimal power flow. Readers are encouraged to refer to [2] for additional details regarding the synthesis for hourly generation profile using publicly available data for different generators including thermal, nuclear, hydro, wind, and other renewables (solar, methane, refuse, and wood).

3.2.3 NYS Load Data

NYISO's database [30] provides information on both real-time and day-ahead load data for each load zone. [2] collected real-time hourly load data to supplement the hourly generation profile discussed in the above section.

3.2.4 Interface Power Flow Data

Major interface flows recorded by NYISO in 2019 were gathered [28]. The dataset includes interface flows between internal zones as well as external trades with neighboring grids [2].

3.2.5 Locational Marginal Price Data

2019 real-time Locational Based Marginal Price (LBMP) data was taken from NYISO for each load zone and four external nodes in the neighboring grid [28]. Marginal costs for external generators were considered as the LBMP of the corresponding external node [2].

3.2.6 Network Modeling

Liu et al. [2] performed network reduction and system updates to simplify neighboring regions, capture 2019 conditions, and reduce computing complexity. Comparable network reduction was performed using the modified WARD algorithm [31]. Before performing network reduction NPCC 140-bus system was updated to reflect the load, generation conditions, and interface flows according to the 2019 data. Hydro Quebec was also added to NPCC 140-bus system. Additionally, some simplifications were made prior to the reduction to remove the transmission lines between neighboring regions such that external regions are only connected to NYS. The final reduced network after adding a transmission line, HVDC lines to reflect the controllable interfaces, and preserving a few buses is displayed in figure 3.1. More details on simplification and updates can be found in [2].

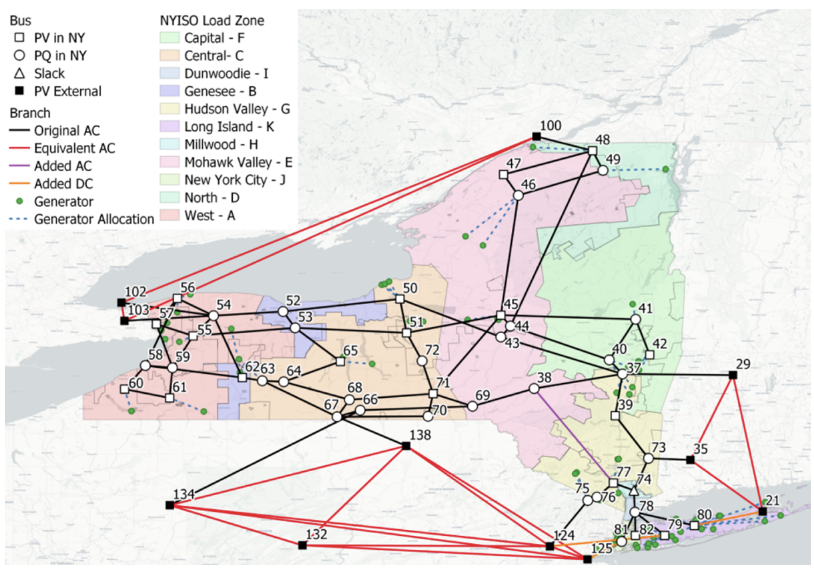

Figure 3.1: NYS electricity grid used for the DCOPF simulation model [2]

3.3 Assumptions

We have made some reasonable assumptions in order to calculate MEFs for the NYS grid at a zonal level. The first assumption is that all the nuclear generators in the system are assumed to serve the base load and run at their maximum capacity. Nuclear plants primarily serve the baseload due to a combination of technical and economic factors. From a technical perspective, it is not possible to promptly activate or deactivate nuclear power plants, and they are typically operational except for occasional maintenance periods. Economically, nuclear plants have high capital costs and low operational costs, making it important to run them at their maximum capacity to recover the initial investment. Similarly, the minimum output for hydro plants is assumed to be sixty percent of its corresponding hourly capacity, so that external generators outside NYS with lower operational costs do not compete with NYS hydro plants while performing power flow analysis.

Another assumption is that the external generators do not respond to an incremental demand increase in NYS, and therefore they are excluded from calculating the marginal emission factor. This is a fair approximation, as external generators are modeled as large synthetic units in the simulated NY grid in [2] with huge capacities and low operating costs and were being assigned as marginal units 80-90% of the time, which is impractical considering a very small demand increase within NYS.

3.4 DCOPF Simulation Model to Identify Marginal Generators

To simulate the dispatch of generators in the NYS electricity grid at an hourly temporal resolution for each load zone, an extended DC Optimal Power Flow (DCOPF) model has been built. DCOPF is a mathematical optimization technique used to determine the dispatch of generators by solving optimal power flow in an electric power system. The basic objective of DCOPF is to minimize the total cost of generating electricity while meeting the system's load demand and satisfying various operational constraints [32].

The DCOPF model simplifies the power system by using a DC approximation of the AC power flow equations, which is a linearized version of the nonlinear AC power flow equations. This simplification allows for a more efficient solution to the optimization problem [33].

The DCOPF model consists of a set of equations that describe the power system's behavior. The model takes as input the power system's topology, generator characteristics, and load demand. It then computes the optimal power generation dispatch that minimizes the total generation cost subject to constraints on the power flow, generator limits, and other operational constraints. The DCOPF model is a convex optimization problem, meaning that it has a unique global minimum. This makes it an efficient and reliable optimization model for solving the problem of economic dispatch in power systems. The mathematical equations of the DCOPF model can be formulated using linear programming or quadratic programming techniques [33].

The extended DCOPF model developed for this project utilizes the Pyomo optimization library. Additionally, functions used to define variables, prepare data, and adapted equations in our DCOPF model are taken from MATPOWER [33, 34] and PYPOWER [35]. This allows for algorithm transparency, standardization, and easy incorporation of additional constraints to the model that are not present in most standard DCOPF models, like ramping constraints, interface flow limits, and DC line constraints.

The following equations are used for the extended DCOPF model to minimize the total cost of production in eq. (3.2) subject to energy balance in eq. (3.3), transmission limits in eq. (3.4) & (3.5), DC line constraints in eq (3.8), interface flow limits in eq (3.9) & (3.10), and ramping constraints in eq. (3.11) & (3.12). Voltage magnitudes are assumed to be one, and reactive power is disregarded. Equations (3.1)-(3.10) are taken from the MATPOWER manual [36].

$$x = \begin{bmatrix} \frac{\theta}{Pg} \end{bmatrix} \tag{3.1}$$

where θ refers to voltage angles and P_g is real power injections [33].

$$\min_{\theta, Pg} \sum_{i=1}^{n_g} f_p^i(p_g^i) \tag{3.2}$$

where f_p^i refers to cost function of real power injections for each generator i, p_g^i refers to real power injections for each generator i, and n_g is the number of generators

subject to

$$g_P(\theta, P_g) = B_{bus}\theta + P_d + G_{sh} - C_g P_g = 0 \tag{3.3}$$

where B_{bus} refers to bus susceptance matrix, P_d denotes active power demand at each bus vector, G_{sh} is real power consumed by shunt elements at each bus vector, and C_g is the connection matrix.

$$h_f(\theta) = B_f\theta - F_{max} \leq 0 \tag{3.4}$$

$$h_t(\theta) = -B_f\theta - F_{max} \leq 0 \tag{3.5}$$

where B_f is the line susceptance matrix, and F_{max} denotes the maximum limit for power flow on the transmission line.

$$\theta_i^{ref} \leq \theta_i \leq \theta_i^{ref}, i \in I_{ref} \tag{3.6}$$

where θ_i^{ref} refers to voltage reference angle and I_{ref} is the set of bus indices of the reference bus

$$p_g^{i,min} \leq p_g^i \leq p_g^{i,max},\ i = 1 \ldots n_g \tag{3.7}$$

where $P_g^{i,\min}$ refers to the minimum power injection for each generator i, and $P_g^{i,\max}$ is the maximum power injection for each generator i.

$$p_t = (1 - l_1)p_f - l_0 \tag{3.8}$$

where refers to p_t refers to power at 'to' bus, p_f is power at 'from' bus, l_0 denotes constant term of the linear loss function, and l_1 is the linear term of the linear loss function.

$$f_k(\theta) = \sum_{i \in B_k} d_i p_i(\theta) \tag{3.9}$$

where p_i refers to real power flow ('from' bus $\rightarrow$ 'to' bus) in branch i, d_i is equal to 1 or -1 to indicate the direction of flow and f_k is interface flow for interface k.

$$F_k^{min} \leq f_k(\theta) \leq F_k^{max},\ V_K \in I_f \tag{3.10}$$

where $F_k^{\min}$ refers to lower bound of interface flow, $F_k^{\max}$ is upper bound of interface flow, and I_f is set indices of interfaces to be considered

$$p_g^i(t) - p_g^i(t-1) \geq RD^i(t),\ i = 1 \ldots n_g \tag{3.11}$$

$$p_g^i(t) - p_g^i(t-1) \leq RU^i(t),\ i = 1 \ldots n_g \tag{3.12}$$

where $p_g^i(t)$ refers to real power injection for generator i at time t, $p_g^i(t-1)$ is real power injection for generator i at time t-1, $RD^i(t)$ denotes ramp down limit for generator i at time t, and $RU^i(t)$ is ramp up limit for generator i at time t

It is worthwhile to provide additional information on constraints that are not included in the standard DCOPF formulation. These constraints include ramping constraints, interface limits, and limits on DC transmission lines. Ramping limit for each generator is set as the minimum of the maximum real power output and ramp rate for 60-minute reserves. Generators are constrained to ramp up or ramp down within these ramping limits as shown in eq. (11) and (12). A detailed description of interface limits and limits on DC transmission lines can be found in [33].

The data utilized to inform the extended DCOPF simulation model is sourced from reference [2], with a brief overview provided in section 3.2. This extended DCOPF model is verified by removing the ramping constraints and comparing the power produced by each generator and the voltage angle at each bus with the results of the baseline model described in [2] at random hours.

Once the baseline dispatch of generators is set based on network constraints, generation, and the load profile for each hour, marginal generators are found by rerunning the DCOPF model with an incremental increase in load for one zone at a time and comparing the results to the baseline run. This results in eleven simulations corresponding to eleven load zones. Marginal generators are those that change their generation in response to this small change in demand.

3.5 Calculating Zonal Marginal Emission Factor

Each generator has a specific carbon emission rate, which is the quantity of carbon emitted per unit of electricity generated (measured in pounds per megawatt-hour, (lb/MWh). The carbon emission rate for each fossil fuel generator in New York State is obtained from the Emissions and Generation Resource Integrated Database (eGRID) of the EPA [37]. Emissions for renewable generators such as hydro, wind, and solar are assumed to be zero.

The marginal emission factor for each load zone at a particular time is calculated by multiplying the power output of the marginal generator by its corresponding carbon emission rate for a unit change in generation. There are times that there is more than one generator on the margin, in which case the MEF is the weighted generation average as shown in equation (3.13).

Mathematically, the MEF for each zone has been calculated using the below equation, where *i* indexes the generators on the margin:

$$MEF = \frac{\sum E_i * ER_i}{E_i} \tag{3.13}$$

where E_i refers to electricity generated (MW) by a marginal generator i, and ER_i is carbon emission rate (Ib/MWh) for that marginal generator

For analysis of these MEF values, they are converted to lb/kWh by dividing them by 1000. Next, we examine the application of the machine learning model to forecast MEF values, which could then support enhanced decision-making in grid operations.

3.6 MLP Model to Forecast MEFs

Machine learning is a subfield of artificial intelligence that facilitates the ability of computer systems to automatically acquire knowledge and enhance performance via experience, without the need for explicit programming. It is a powerful tool that can be used to analyze large datasets and make predictions based on patterns and relationships in the data [38]. In the context of calculating marginal emission factors for NYS, machine learning can be used to develop predictive models based on historical data.

The Multilayer Perceptron (MLP) neural network is a powerful machine learning algorithm that has gained popularity in various domains, including time series forecasting [39]. Therefore, this research utilizes MLP to forecast hour-ahead MEFs in the NYS grid. The MLP is a type of feedforward neural network, meaning that data propagation occurs unidirectionally, starting with input layer and passing through one or many hidden layers before reaching the output layer. It consists of multiple layers of interconnected artificial neurons. The input layer receives the explanatory variables as input features. Each neuron in the input layer represents a specific feature and transmits the corresponding value to the neurons in the subsequent layers. The hidden layers, as the name suggests, are not directly observable and provide intermediate computations. Each neuron in a hidden layer takes the weighted sum of the outputs from the previous layer and applies a non-linear activation function to produce an output. This process allows the MLP to learn complex relationships in the data. Therefore, MLP model is particularly suitable for this research as it can effectively capture complex patterns within the input data [40].

The output layer represents the final prediction or output of the MLP. In this research, the output layer predicts the MEF for the following hour. The number of neurons in the output layer corresponds to the number of output variables. The activation function used in the output layer depends on the nature of the problem. For regression tasks, a non-linear activation function is commonly employed. In our case, we have used logistic activation function [40].

Training an MLP involves an iterative process called backpropagation. Initially, the weights and biases of the network are assigned random values. The training data, which consists of historical records of explanatory variables and corresponding MEF values, is used to adjust these weights and biases. The model tries to minimize a predefined loss function, such as mean squared error, by comparing its predictions with the actual MEF values. The training process is repeated for multiple iterations or epochs until the network achieves a satisfactory level of accuracy in predicting the output values [40]. The MLP model is trained for each zone within the grid, utilizing training data from DCOPF simulation model in addition to other explanatory variables such as the fuel mix of the grid, marginal generator fuel type, and electricity price from the previous hour.

Once the MLP model is trained, it can be used for forecasting the zonal MEF in the NYS grid for the next hour. The accuracy and performance of the MLP model are evaluated by comparing its predicted MEF values with the actual MEF values from the test data. This assessment helps gauge the effectiveness of the model in predicting future MEF values. Measurements such as normalized mean squared error and mean absolute percentage error are used to quantitatively evaluate the model's performance [39].

For this research, several input features were investigated to determine which were the most important for forecasting the MEF values in each zone. Table 3.1 describes the nine input features assessed for estimating hour-ahead MEF values, and table 3.2 summarizes the source of the input data.

Input	**Description**
Temperature	Temperature at 2 Meters (^0C)
Humidity	Specific Humidity at 2 Meters* (g/kg)
Precipitation	Precipitation* (mm/hour)
Solar Irradiance	Clear Sky Surface Shortwave Downward Irradiance (Wh/m^2) *
Wind speed	Wind Speed at 10 Meters* (m/s)
Fuel mix	Fraction of Electricity from Different Fuel Types in New York State
LMP	Electricity Price ($/MWh)*
Load	Electricity Load (MW)*
Marginal Generator	Marginal Generator Fuel Type*

*Collected at a zonal level (A-K)

Table 3.1: Inputs used for feature selection in machine learning training process

Considering that there is a continual flow of electricity in the NYS grid between different zones, MEF values in one zone are likely to depend on parameters from other zones as well. Therefore, it is important to consider features from neighboring zones while making predictions for any particular zone.

Features are the input variables that we provide to our machine-learning models. Each column within the dataset represents a feature. To train a suitable model, we must utilize only the most important variables. If there are too many variables, the model may learn from noise and capture irrelevant patterns. Additionally, when two or more features are highly correlated, it can lead to collinearity that can undermine the ability of the model to identify variables that are statistically significant [22]. Therefore, feature selection was performed to select the optimum set of features and remove multicollinearity from our model. Variation Inflation Factor (VIF) is a statistical tool that can determine the degree to which one independent variable is correlated with a number of other variables and is used to address multicollinearity in our model. Any independent variable having a VIF value greater than 10 indicates a high correlation and is removed as an input to the model [22]. Additionally, Pearson correlation coefficients quantify the relationship between two variables by considering the ratio of their covariance to the product of their standard deviations, giving a value within the range of +1 and -1, with 0 indicating no correlation, 1 representing complete positive correlation, and -1 denoting complete negative correlation [41]. Pearson correlation coefficients were used to identify the input features having the highest correlation with MEF values and use them as inputs to the model.

Data Collected	Source
Weather	MERRA-2 Reanalysis Dataset [42]
Fuel mix	DCOPF Simulation model
LMP	Pricing Data New York ISO [28]
Load	Load Data New York ISO [30]
Marginal Generator	DCOPF Simulation model

Table 3.2: Source of the data collected for machine learning model

CHAPTER 4

RESULTS

In this chapter, a comprehensive assessment of the diurnal, seasonal, and interzonal patterns of simulated MEFs is conducted. This analysis is essential for validating the accuracy and dependability of the MEF values, ensuring that they capture historical trends accurately. To further validate the simulated MEFs, they are compared with the findings of an NYSERDA study focused on evaluating emission factors for the NYS electricity grid. By comparing the simulated MEFs to this standard, we determine the model's ability to accurately estimate the emission values. In addition, an examination of marginal generators is conducted to identify the variances that exist between zones and the type of generators that appear on the margin most of the time. This analysis provides a crucial foundation for future decision-making processes and policy interventions aimed at maximizing the performance of the electricity grid while minimizing environmental impacts at a zonal level. In addition, the performance of the forecasting model is evaluated to determine its ability to accurately predict MEF values. Performance metrics, such as normalized root mean square error and mean absolute percentage error allow us to draw meaningful conclusions regarding the forecasting model's effectiveness. By evaluating its strengths and weaknesses, we can refine and improve the model's predictive capabilities, thereby providing stakeholders with valuable insights for environmental management.

4.1 Accuracy of the MEF Estimates

Using the above-described methodology, we have simulated the hourly MEFs for each zone in the NYISO based on the historical conditions in 2019. We then consider the simulated MEFs from the perspective of diurnal and seasonal patterns and inter-zonal variations in figures. 4.1 and 4.3.

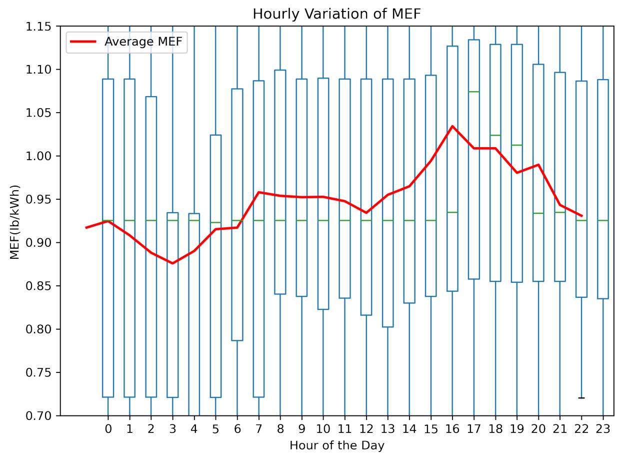

Figure 4.1: Hourly variation of MEF estimates for the 24-hour period in Zone D shows that MEFs are lower during nighttime hours when demand is low, and increase during the day as demand rises This diurnal pattern is similar for all the load zones, and Zone D is shown as an example

In Figure 4.1, we first consider the daily pattern in MEFs; both average and variability, showing lower MEFs during the nighttime hours, followed by a gradual increase over the course of the day. This pattern is a result of lower demand during the nighttime hours, when demand can be met by generators with lower emission rates, as shown in Figure 4.2.

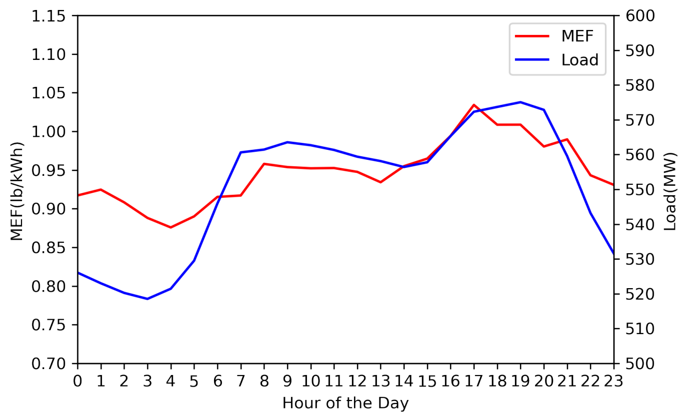

Figure 4.2: MEF and load for each hour in Zone D shows that the MEF follows load patterns over the day, aligning with the expectation that renewable resources are dispatched before thermal (fossil) generators are needed.

While the diurnal pattern of MEFs is similar across load zones, the MEF values are driven by the mix of generators within and adjacent to each zone. In Figure 4.3, a longer-term perspective shows the differences among seasonal MEF values across all zones in NYS. The findings presented in Figure 4.3 show that MEFs tend to be lower in the spring and fall seasons, and higher in winter and summer seasons. This observation aligns with the anticipated pattern of higher emissions during months characterized by extreme temperatures, which correspond to periods of increased energy demand, requiring the use of less efficient generators.

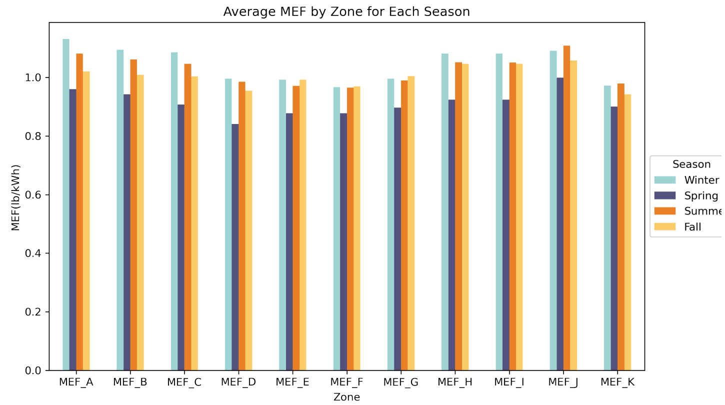

Figure 4.3: Inter-zonal and seasonal variation of MEFs shows that MEFs are generally higher during summer and winter compared to spring and fall seasons.

Examining Figure 4.3, shows that zones A, B, and C exhibit higher MEFs than other upstate zones (D and E). This can be attributed to the presence of two coal plants in each of zones A and C. In 2019, the coal plants in zones A and C were on the margins 12–14% of the time. Conversely, Zone K tends to have lower MEF estimates as natural gas plants with lower emission factors are more frequently on the margin in this area. This will be further discussed in the marginal generator analysis in the results section.

Current studies conducted by the New York State Energy Research and Development Authority (NYSERDA) use a constant MEF of 0.55 short tons of carbon dioxide per megawatt-hour of electricity produced in NYS [3], which is equivalent to 1.1 pounds of CO_2 per kWh. While the NYSERDA study does not provide zonal resolution, the aggregated MEFs estimated by the NYSERDA study (shown in Figure 4.4) align, in general, with the estimated zonal MEFs simulated by the model proposed here.

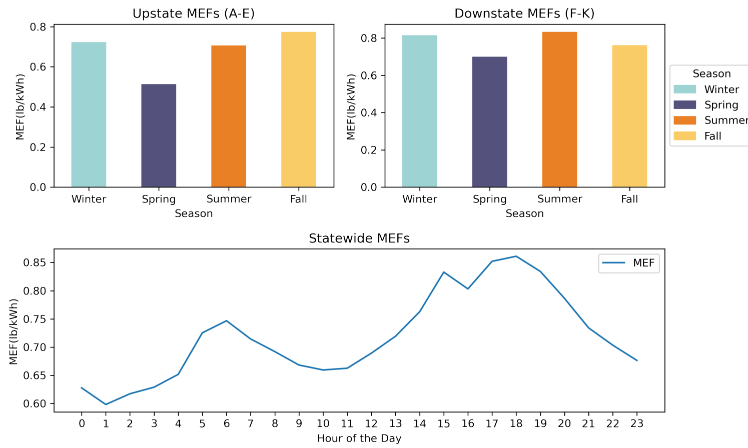

Figure 4.4: Seasonal and diurnal variations of MEFs reported by NYSERDA [3] demonstrate similar patterns as simulated by our model

Based on these comparisons, we believe that the MEF values simulated from the proposed model provide a sufficient representation of historical MEFs representing dynamic states of the system. Next, we explore the marginal generators across different zones.

4.2 Marginal Generator Analysis

Figure 4.5 shows the capacity mix and frequency of occurrence of marginal generators based on their respective fuel types for 2019 across all the load zones. The latter is obtained from our DCOPF simulation model.

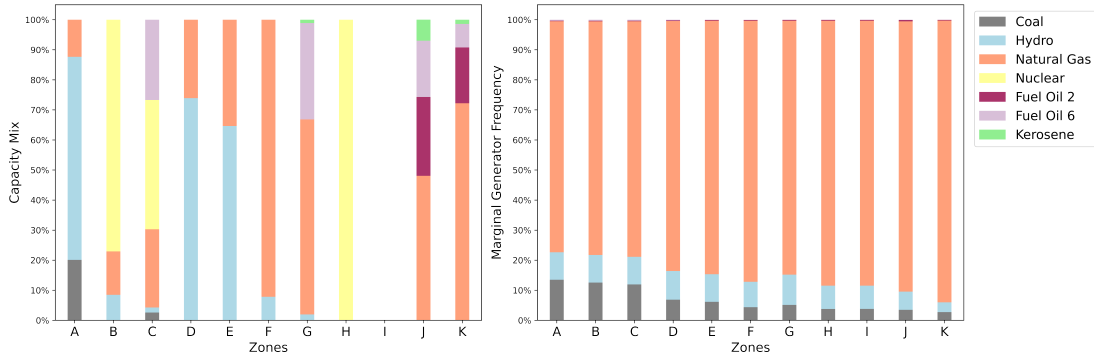

Figure 4.5: Capacity mix (left) and marginal generator frequency (right) across zones, highlighting the dominance of natural gas plants as marginal generators irrespective of the capacity mix.

As shown in figure 4.5 our simulation model indicates that natural gas plants consistently operate at the margin majority of the time across all load zones irrespective of the capacity mix, exerting a significant influence on the mean values of MEFs. The NYSERDA report [3] provides additional evidence that natural gas facilities are typically the marginal units in NYS. Natural gas plants being the marginal generators is a consequence of the demand, generation mix, and fuel cost. In NYS, demand generally exceeds the generation capacity of renewable generators such as wind, solar, and hydro, and these plants having the least cost to operate, are dispatched first; any increase in demand is met by the spare capacity of natural gas plants, making them marginal units for the majority of hours. As shown in Figure 4.6, there is a clear upward trend in the frequency of natural gas plants operating as marginal generators as we move from upstate to downstate. This is consistent with our expectations, as downstate zones have more thermal generators than upstate zones. There are two coal plants, one in Zone A and the other in Zone C for 2019, and accordingly, they appear as marginal generators most of the time in Zones A, B, and C. The power generation from these plants decreases as we move from upstate to downstate zones, as shown in Figure 4.7. Similarly, the frequency of hydro plants on the margin generally decreases from Zone A to Zone K. This is because there are more hydro plants in upstate compared to downstate.

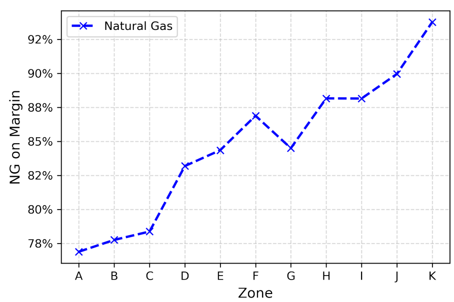

Figure 4.6: Increasing trend of marginal NG plants from upstate to downstate

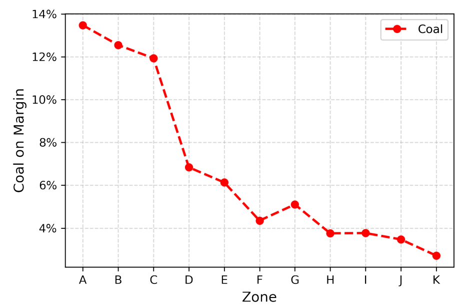

Figure 4.7: Decreasing trend of marginal coal plants from upstate to downstate

Unlike the majority of the literature that assumes a single emission rate for a particular fuel type [9, 10, 23], our simulation model takes a more holistic approach and acknowledges the fact that different thermal plants can have different emission rates; therefore, we consider annual average emission rates at a plant level (lb/MWh) [37], ignoring the variability in emission rate over the generator's dispatch curve. The variability of emission rates can be attributed to several factors, including plant design, efficiency, fuel quality, maintenance practices, and regulatory compliance. Figure 4.8 illustrates the emission rate of natural gas plants that are present on the margin across all the zones. The data indicate that marginal natural gas plants located in Zone K exhibit a relatively lower emission rate in comparison to other zones. This observation provides an explanation for the lower average marginal emission values for Zone K identified in Figure 4.3.

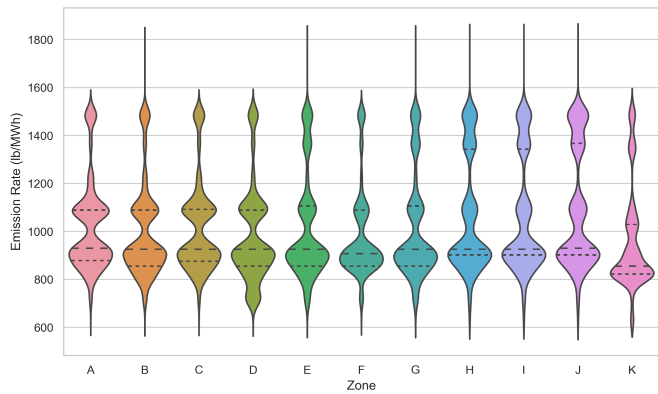

Figure 4.8: Emission rates for NG plants on margin across zones, where width of the shape represents the relative frequency of a particular emission rate on margin for hourly simulation in 2019.

4.3 Performance of the Forecasting Model

The forecasting model revealed the importance of primary variables for predicting hour-ahead MEF values. These variables included marginal generator fuel type, NYS fuel mix, historical MEF values, and electricity price. All of these input features, with the exception of electricity price, which is easily accessible, are taken directly from the output of our dispatch model, making it straightforward to collect the input features used for training the machine learning model. Appendix A provides further information on input features that are important for each zone.

The forecasting accuracy of the proposed MLP model is evaluated through the holdout cross-validation technique, which splits the hourly data for 2019 into three parts: training (80%), validation (10%), and test (10%) for all the load zones (A-K).

The Root Mean Square Error (RMSE) is a commonly used method to evaluate how accurately a model predicts numerical data [39], which can be defined as:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{4.1}$$

where n is number of test data, y_i denotes real value, and $\hat{y}_i$ is the predicted value.

The Normalized RMSE (NRMSE) measures the proportion of RMSE in relation to the range of the target variable [39].

$$\mathrm{NRMSE} = \frac{RMSE}{y_{max} - y_{min}} \tag{4.2}$$

where $y_{\max}$ is the maximum real value, and $y_{\min}$ is the minimum of real value.

The Mean Absolute Percentage Error (MAPE) is a measure that calculates the average absolute percentage errors and can be mathematically defined as:

$$\mathrm{MAPE} = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{4.3}$$

where n is number of test data, y_i denotes real value, and $\hat{y}_i$ is the predicted value.

Table 4.1 shows the performance of the model for each load zone.

Zone	NRMSE	MAPE %
A	0.06	39
B	0.11	38
C	0.12	39
D	0.12	36
E	0.09	33
F	0.12	27
G	0.07	28
H	0.13	33
I	0.13	32
J	0.13	30
K	0.12	27

Table 4.1: Performance of the MLP model for each load zone

The low normalized RMSE values observed across all zones in Table 4.1 show that the proposed model exhibits a high degree of accuracy [22, 43]. This suggests that the model is capable of capturing the overall trend present in the data. However, the MAPE values are high, indicating the challenging nature of predicting the precise magnitudes of the MEF values.

Figure 4.9 depicts an 80-hour time series for the predicted and actual values of MEFs for zone B. We can observe that the model captures the general pattern on average, however, the extreme values can not be predicted accurately. This can be attributed to imbalanced data in the target variable (hour-ahead MEFs). As shown in Figure 4.8, natural gas generators were the marginal units in the majority of NYS intervals (77-94%) in 2019, indicating that the majority of the MEFs will fall within the emission rate range of these plants. For instance, approximately 70% of the MEFs in Zone A and 88% of the MEFs in Zone K fall within the range of 0.8 to 1.5 lb/kWh (typical emission rate range for natural gas plants). Therefore, extreme values, such as zero MEFs when hydro is on margin or high MEFs when a coal plant or a highly carbon-intensive natural gas plant is on margin can be considered relatively infrequent events. As the standard algorithm for supervised machine learning gives equal weight to all MEFs, these infrequent events are given less importance during model training making it difficult to accurately forecast their magnitude using historical data.

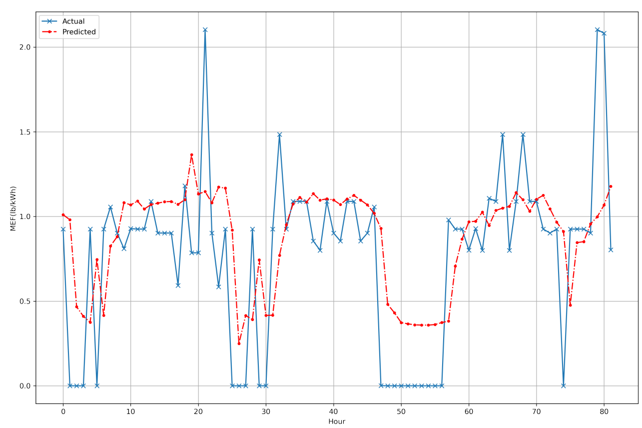

Figure 4.9: Actual vs predicted MEFs for Zone B over a random 80-hour period show the ability of the proposed model to predict patterns in MEFs.

From Figure 4.9, we can conclude that our model is capable of capturing the underlying relationships, trends, and dependencies present in the data. It is able to identify the typical patterns in MEFs that occur over time. Correctly identifying the pattern is crucial because it enables the successful implementation of time-sensitive emission reduction strategies, such as storage optimization and demand shift. For example, it would be helpful for energy storage system owners to know the exact hour when the MEFs will shift from high to low patterns, so they can start charging their batteries and stop when the MEFs are expected to rise. This can help mitigate emissions because stored energy generated by less carbon-intensive generators can be utilized during periods when a carbon-intensive generator is operating at the margin. On the other hand, it is equally important to be able to accurately determine the magnitude of MEFs, as this would result in greater confidence and more reliable data for decision-making. In section 5.2, we will discuss a few strategies that could be implemented to enhance the model's ability to accurately predict magnitudes.

CHAPTER 5
DISCUSSION AND CONCLUSIONS

The results presented in Chapter 4 show that while there is scope for additional enhancements, the findings presented in this study offer valuable insights into the MEFs with detailed spatial and temporal resolution. We observed that nighttime MEFs are lower due to decreased demand than daytime. On a longer-term perspective, spring and fall MEFs are less than summer and winter as increased energy demand during extreme temperatures places carbon-intensive generators on margin. Moreover, we found that MEFs varied between zones based on marginal generator composition and not primarily on the zonal capacity mix. In addition, we determined that natural gas generators are typically the marginal units, and we identified variables such as electricity price, fuel mix, historical MEFs, and marginal generator fuel type to help explain the pattern in future MEFs. We can effectively utilize the proposed model to not only estimate MEFs in real time but also predict their pattern a short time in the future, which can be helpful in making informed decisions to cut down on emissions. To highlight the potential for an accurate MEF estimation model, in this chapter, we discuss some key applications that are enabled and a detailed case study describing the real-world application of the findings. Following this, potential future improvements to the model are discussed, along with related areas of future research. This chapter concludes with a summary of the thesis's research contributions and findings.

5.1 Applications

MEFs are an important metric of the environmental impact of power generation. Estimating and predicting MEFs can be useful to a number of stakeholders, including consumers, utilities, and policymakers, for various purposes.

- Shifting Consumption: Commercial consumers may want to avoid using energy-intensive appliances or devices during periods of high emission factors. They can instead delay the use of energy until the marginal emission is predicted to be lower.
- Optimizing Storage: By knowing the expected marginal emissions in the next hour, buildings can optimize their energy storage. For example, commercial real estate managers may want to focus on charging their energy storage systems during periods of low marginal emissions and utilizing this clean energy when emissions in the grid are higher, thereby offsetting some emissions.
- Evaluating Benefits: Policymakers can use the model to compare the benefits of different renewable energy technologies considering the social cost of carbon. This particular application is illustrated with a stylized example in subsection 5.1.1.

5.1.1 Evaluating benefits of different renewable energy technologies

Using the MEF estimates derived from our dispatch model, policymakers and investors can evaluate the benefit derived from different renewable energy technologies at a zonal level. The purpose of this case study is to determine the zonal financial benefits of integrating renewable energy into the NYS electricity grid under various scenarios. As this application considers generation mix changes in 2040 in accordance with the climate change impact phase II report [44], the dispatch model will be better able to capture these changes than the MLP model, which is more effective for forecasting short-term patterns based on historical data. Consequently, we utilize the dispatch model of our proposed algorithm to estimate the reduction in emissions at a zonal level based on various renewable energy integration scenarios. This study focuses on Zone A, but it can be easily extended to other zones as well in order to understand the variation in emission reduction between zones. Through this case study, we aim to determine which renewable energy technology would be most beneficial in order to provide investors and policymakers with valuable guidance for prioritizing the development of one renewable energy over another.

To determine the reduction in emissions for Zone A with the new set of marginal generators, we rerun our dispatch model with the updated generation mix. This emission reduction is converted into tangible benefits by designating a social cost to carbon emissions offset, and this profit is then compared across various renewable energy scenarios in order to determine the option that is most beneficial. While rerunning the DCOPF model with the modified generation mix, the load profile, operating costs, and transmission architecture remain unchanged from the initial run. Behind-the-meter solar and storage capacity are not considered for this study and can be included in future work. Table 5.1 displays the allocation of 17,761 MW of wind and 19,631 MW of utility solar to each zone [44]. Zonal allocation of wind and solar is further disaggregated to PV buses in the dispatch model and the bus allocation is taken from [45]. The time series trajectories for these simulated wind and solar sites based on wind speed, incident solar radiation, and ambient air temperature are obtained from [45].

Zone	**Wind (MW)**	**Solar (MW)**
A	2692	5748
B	390	656
C	1923	3585
D	1935	-
E	1821	2268
F	-	4661
G	-	1636
H	-	-
I	-	-
J	6391	-
K	2609	77

Table 5.1: Zonal utility wind and solar allocation assume renewable resources will be prioritized for development to achieve the CLCPA 100% renewable resource goal for 2040.

The New York State Department of Environmental Conservation (NYSDEC) [46] has developed guidelines for assessing the costs associated with greenhouse gas emissions from grids. These expenses are determined by applying a value that is based on the damages caused by carbon. This strategy makes use of Integrated Assessment Models (IAMs) to convert an incremental increase in emissions into a change in atmospheric concentrations of greenhouse gases, a change in the climate that results from this change, and finally, the economic repercussions that arise from this change. For this study, we use a central discount rate of 2% and take the social cost value of carbon for future years, as mentioned in the NYSDEC report. The social cost of carbon is not reported by NYSDEC for 2019 and for this study, we take an approximate value based on the trend of future values.

MEF values were calculated using the DCOPF model for different years to quantify uncertainty (2019, 2018, and 2017) and under four different scenarios as follows:

- Scenario 1: No action scenario
- Scenario 2: Integrating wind and solar generation as per the climate change impact phase II report
- Scenario 3: Integrating only solar generation as per the climate change impact phase II report
- Scenario 4: Integrating only wind generation as per the climate change impact phase II report

Year	**Scenario**	**Average Annual MEF (lb/kWh)**	**MEF Reduction (lb/kWh)**
2019	No Action	1.06	
	Solar +Wind	0.47	0.59
	Solar	0.83	0.23
	Wind	0.63	0.43
2018	No Action	1.11	
	Solar +Wind	0.51	0.6
	Solar	0.93	0.18
	Wind	0.7	0.41
2017	No Action	1.05	
	Solar +Wind	0.43	0.62
	Solar	0.78	0.27
	Wind	0.6	0.45

Table 5.2: Annual average MEF and reduction under different scenarios for Zone A

As we can see from Table 5.2, MEF reduction in Zone A after integrating renewable energy is fairly consistent across different years, and therefore we have taken the average value to calculate the avoided emissions and associated benefits based on the social cost of carbon. Below in Table 5.3, we summarize the net present value for each scenario. Detailed calculations for these results can be found in Appendix B.

Scenario	**Present Benefit ($/kW)**
Solar + Wind	5,221
Solar	1,914
Wind	3,742

Table 5.3: Present benefit for different renewable energy technologies

The calculations presented earlier allow us to draw the following conclusions, subject of course to the assumptions that we have a 2% discount rate, NYSDEC reported social cost of carbon, and our model's suggested emission reduction range: (1) Installing both wind farms and solar units will give the highest benefits, (2) Installing wind farms will provide more value compared to solar units. Policymakers and investors can utilize this information to evaluate the benefits of various renewable technologies in specific zones. This can aid in their decision-making process to prioritize one technology over another.

5.2 Future Directions

To the best of our understanding, this study is the first to estimate MEFs at a zonal level within the NYS grid. In forthcoming endeavors, the machine learning model can be improved to accurately determine the magnitude and variance of MEFs. Since we believe that the model's inability to predict extreme values may be due to an imbalance in the target variable where majority of the MEFs fall within the typical emission rate range of natural gas plants (0.8 to 1.5 lb/kWh), it would be beneficial to investigate techniques such as applying a relevance function to MEFs that introduce a larger bias toward under-represented MEFs that fall outside this range [47]. Synthetically generating the under-represented data or undersampling the majority data using techniques such as SMOTE (Synthetic Minority Over-sampling Technique) for Regression [48] or SMOGN [49] are additional methods for creating a balance in the training dataset. Furthermore, the variability in emission rate for natural gas generators can make it difficult to predict the exact magnitude of MEFs. This issue can be resolved by incorporating publicly available plant-level data, such as technology used (combined cycle plants have lower emissions than combustion turbine plants), age of the plant (older plants typically have higher emissions than newer plants), and overall efficiency levels. Another possible direction could be to increase the forecasting horizon of the machine learning model. Having the ability to predict MEFs a day in advance would prove advantageous for early decision-making. Furthermore, instead of MLP, other machine learning methods like DeepAR [50], which combines deep learning and autoregressive parts, could be used to do probabilistic forecasting instead of point forecasting and get a distribution of possible values for future MEF estimates for better risk management. Finally, the dispatch model can be enhanced by including binary dispatch decision variables to account for the scheduled maintenance and outages of generating units.

5.3 Conclusion

This study estimates and forecasts marginal emission factors at the zone level for New York State. Having knowledge of these hourly zonal MEFs is important to make informed emission reduction decisions both from the supply and demand sides. As the energy system is dynamic, MEFs can vary from place to place and from time to time, emphasizing the need for a model that can provide estimates with high spatial and temporal resolution. This analysis indicates that MEFs are generally low during off-peak hours (11 p.m.–7 a.m.) and elevated during on-peak hours (7 a.m.–11 p.m.). In addition, average MEF values are greater in the summer and winter, when more energy is required to maintain a comfortable temperature, than in the spring and fall, when less energy is required. It is further noticed that natural gas plants are on margin for most of the operating hours (80%-90%) and this percentage increases as we move from upstate to downstate. This shows that, in addition to integrating renewable energy, improving the efficiency of natural gas plants can help reduce MEFs, specifically in downstate zones. Our forecasting model is capable of projecting the pattern of future MEF values and can play an important role in optimizing energy activities by integrating it with a digital energy management platform. Economic benefits resulting from emission reduction, such as the environmental value assigned to energy generated from distributed energy resources under the Value of Distributed Energy Resources (VDER) methodology [13], will accelerate the adoption of cleaner practices and help us combat climate change.

APPENDIX A

INPUT FEATURES SELECTED BY ZONE

The table below provides information on the top four input features selected after feature engineering for each load zone. One important thing to note here is that these features are from the zone under investigation as well as the neighboring zones, highlighting the spatial dependence of MEF estimates.

Target Variable	**Input Features**
MEF_A(t+1)	Hydro Marginal Generator_B(t)
	LMP_C(t)
	NYS Hydro Mix(t)
	MEF_C(t-1)
MEF_B(t+1)	Hydro Marginal Generator_A(t)
	LMP_C(t)
	MEF_C(t-1)
	NYS Hydro Mix(t)
MEF_C(t+1)	Hydro Marginal Generator_C(t)
	MEF_B(t-1)+MEF_B(t-2)
	LMP_B(t)
	NYS Nuclear Mix(t)
MEF_D(t+1)	Hydro Marginal Generator_D(t)
	LMP_C(t)
	MEF_E(t-1)
	NYS Hydro Mix(t)

MEF_E(t+1)	Hydro Marginal Generator_E(t)
	MEF_E(t-1)
	NG Marginal Generator_D(t)
	MEF_E(t-2)
MEF_F(t+1)	Hydro Marginal Generator_F(t)
	MEF_E(t-1)
	LMP_E(t)
	NG Marginal Generator_F(t)
MEF_G(t+1)	Hydro Marginal Generator_G(t)
	MEF_H(t-1)
	LMP_H(t)
	NYS Coal Mix(t)
MEF_H(t+1)	Hydro Marginal Generator_H(t)
	MEF_G(t-1)
	LMP_I(t)
	Hydro Marginal Generator_G(t)
MEF_I(t+1)	Hydro Marginal Generator_I(t)
	MEF_I(t-1)
	NYS Nuclear Mix(t)
	Hydro Marginal Generator_J(t)
MEF_J(t+1)	MEF_K(t-1)
	LMP_I(t)
	LMP_K(t)
	MEF_J(t-2)
MEF_K(t+1)	NYS Nuclear Mix(t)

LMP_I(t)
LMP_K(t)
MEF_K(t-1)

Table A.1: Input/output variables in MLP (displayed as "name_zone(timestep)")

APPENDIX B

CASE STUDY CALCULATIONS

The table below provides detailed calculations for the present benefit of different types of renewable integration.

B.1 Scenario 2 (Solar + Wind):

Year	US Social Cost of Carbon ($/mt)	Annual CO_2 Emission Reduction (mt/kW)	Annual Benefits ($/kW)
2019	120	2.38	286.09
2020	121	2.38	288.47
2021	123	2.38	293.24
2022	124	2.38	295.63
2023	126	2.38	300.39
2024	128	2.38	305.16
2025	129	2.38	307.55
2026	131	2.38	312.31
2027	132	2.38	314.7
2028	134	2.38	319.47
2029	136	2.38	324.23
2030	137	2.38	326.62
2031	139	2.38	331.39
2032	141	2.38	336.16

2033	142	2.38	338.54
2034	144	2.38	343.31
2035	146	2.38	348.08
2036	147	2.38	350.46
2037	149	2.38	355.23
2038	151	2.38	360.0
Cumulative Benefit			6437.01
Present Benefit			5220.87

Table B.1: Scenario 2 (Solar + Wind)

B.2 Scenario 3 (Solar):

Year	US Social Cost of Carbon ($/mt)	Annual CO_2 Emission Reduction (mt/kW)	Annual Benefits ($/kW)
2019	120	0.87	104.9
2020	121	0.87	105.77
2021	123	0.87	107.52
2022	124	0.87	108.4
2023	126	0.87	110.14
2024	128	0.87	111.89
2025	129	0.87	112.77
2026	131	0.87	114.52
2027	132	0.87	115.39
2028	134	0.87	117.14
2029	136	0.87	118.89
2030	137	0.87	119.76
2031	139	0.87	121.51
2032	141	0.87	123.26
2033	142	0.87	124.13
2034	144	0.87	125.88
2035	146	0.87	127.63
2036	147	0.87	128.50
2037	149	0.87	130.25
2038	151	0.87	132.0

Cumulative Benefit	2360.24
Present Benefit	1914.32

Table B.2: Scenario 3 (Solar)

B.3 Scenario 4 (Wind):

Year	US Social Cost of Carbon ($/mt)	Annual CO_2 Emission Reduction (mt/kW)	Annual Benefits ($/kW)
2019	120	1.71	205.03
2020	121	1.71	206.74
2021	123	1.71	210.16
2022	124	1.71	211.87
2023	126	1.71	215.28
2024	128	1.71	218.70
2025	129	1.71	220.41
2026	131	1.71	223.83
2027	132	1.71	225.53
2028	134	1.71	228.95
2029	136	1.71	232.37
2030	137	1.71	234.08
2031	139	1.71	237.49
2032	141	1.71	240.91
2033	142	1.71	242.62
2034	144	1.71	246.04
2035	146	1.71	249.45
2036	147	1.71	251.16
2037	149	1.71	254.58
2038	151	1.71	258.0

Cumulative Benefit	4613.19
Present Benefit	3741.62

Table B.3: Scenario 4 (Wind)

BIBLIOGRAPHY

[1] "Power trends 2019," 2019. Retrieved May 7, 2023, from https://www.nyiso.com/documents/20142/2223020/2019-Power-Trends-Report.pdf/0e8d65ee-820c-a718-452c-6c59b2d4818b.

[2] M. V. Liu, B. Yuan, Z. Wang, J. A. Sward, K. M. Zhang, and C. L. Anderson, "An open source representation for the nys electric grid to support power grid and market transition studies," *IEEE Transactions on Power Systems*, vol. 38, no. 4, pp. 3293–3303, 2023.

[3] New York State Energy Research and Development Authority (NYSERDA), *Projected Emission Factors for New York State Grid Electricity*. NYSERDA, Albany, NY, 2022. nyserda.ny.gov/publications.

[4] Intergovernmental Panel on Climate Change, "The ipcc sixth assessment report on climate change impacts," *Population and Development Review*, vol. 48, pp. 629–633, June 2022.

[5] "Toward a clean energy future: A strategic outlook 2022-2025," 2022.

[6] United States Environmental Protection Agency (USEPA), "Greenhouse Gas Emissions in the United States." Retrieved February 13, 2023, from https://www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions, 2021.

[7] "U.S. Energy Information Administration - EIA - Independent Statistics And Analysis." https://www.eia.gov/state/?sid=NYtabs-4, June 2023.

[8] A. Finenko and L. Cheah, "Temporal co2 emissions associated with electricity generation: Case study of singapore," *Energy Policy*, vol. 93, pp. 70–79, 2016.

[9] C. Wang, Y. Wang, C. J. Miller, and J. Lin, "Estimating hourly marginal emission in real time for pjm market area using a machine learning approach," in *2016 IEEE Power and Energy Society General Meeting (PESGM)*, pp. 1–5, 2016.

[10] A. J. Pimm, J. Palczewski, E. R. Barbour, and T. T. Cockerill, "Using electricity storage to reduce greenhouse gas emissions," *Applied Energy*, vol. 282, p. 116199, 2021.

[11] 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. PMID: 22486733.

[12] Z. Zheng, F. Han, F. Li, and J. Zhu, "Assessment of marginal emissions factor in power systems under ramp-rate constraints," *CSEE Journal of Power and Energy Systems*, vol. 1, no. 4, pp. 37–49, 2015.

[13] "NYSERDA." [online]. Available at https://www.nyserda.ny.gov/All-Programs/NY-Sun/Contractors/Value-of-Distributed-Energy-Resources.

[14] A. Hawkes, "Estimating marginal co2 emissions rates for national electricity systems," *Energy Policy*, vol. 38, no. 10, pp. 5977–5987, 2010. The socio-economic transition towards a hydrogen economy - findings from European research, with regular papers.

[15] N. Gilbraith and S. E. Powers, "Residential demand response reduces air pollutant emissions on peak electricity demand days in new york city," *Energy Policy*, vol. 59, pp. 459–469, 2013.

[16] C. N. Smith, "Energy savings and marginal emissions factor approach to emissions reductions from household appliance efficiency improvements," 2016.

[17] V. Raichur, D. S. Callaway, and S. J. Skerlos, "Estimating emissions from electricity generation using electricity dispatch models: The importance of system operating constraints," *Journal of Industrial Ecology*, vol. 20, pp. 42–53, 2016.

[18] V. Olkkonen and S. Syri, "Spatial and temporal variations of marginal electricity generation: the case of the finnish, nordic, and european energy systems up to 2030," *Journal of Cleaner Production*, vol. 126, pp. 515–525, 2016.

[19] T. A. Deetjen and I. L. Azevedo, "Reduced-order dispatch model for simulating marginal emissions factors for the united states power sector," *Environmental Science & Technology*, vol. 53, no. 17, pp. 10506–10513, 2019. PMID: 31436968.

[20] M. Li, T. M. Smith, Y. Yang, and E. J. Wilson, "Marginal emission factors considering renewables: A case study of the u.s. midcontinent independent system operator (miso) system," *Environmental Science & Technology*, vol. 51, no. 19, pp. 11215–11223, 2017. PMID: 28863474.

[21] M. P. S. Thind, E. J. Wilson, I. L. Azevedo, and J. D. Marshall, "Marginal emissions factors for electricity generation in the midcontinent iso," *Environmental Science & Technology*, vol. 51, no. 24, pp. 14445–14452, 2017. PMID: 29152978.

[22] K. Leerbeck, P. Bacher, R. G. Junker, G. Goranovi*f*á, O. Corradi, R. Ebrahimy, A. Tveit, and H. Madsen, "Short-term forecasting of co2 emission intensity in power grids by machine learning," *Applied Energy*, vol. 277, p. 115527, 2020.

[23] D. Maji, R. K. Sitaraman, and P. Shenoy, "Dacf: Day-ahead carbon intensity forecasting of power grids using machine learning," in *Proceedings of the Thirteenth ACM International Conference on Future Energy Systems*, e-Energy '22, (New York, NY, USA), p. 188–192, Association for Computing Machinery, 2022.

[24] B. Howard, M. Waite, and V. Modi, "Current and near-term ghg emissions factors from electricity production for new york state and new york city," *Applied Energy*, vol. 187, pp. 255–271, 2017.

[25] "Zero-emissions electric grid in new york by 2040 study." https://www.nyserda.ny.gov/-/media/Project/Nyserda/Files/Publications/NY-Power-Grid/Appendix-E.pdf (accessed Jul. 04, 2023), 2023.

[26] J. Chow and G. Rogers, "User manual for power system toolbox," *version*, vol. 3, pp. 1991–2008, 1991.

[27] W. Ju, J. Qi, and K. Sun, "Simulation and analysis of cascading failures on an npcc power system test bed," in *2015 IEEE Power & Energy Society General Meeting*, pp. 1–5, IEEE, 2015.

[28] "Energy market & operational data - nyiso." https://www.nyiso.com/energy-market-operational-data.

[29] New York Independent System Operator, "Nyiso gold book—load and capacity data." [online], 2019. Available at https://www.nyiso.com/documents/20142/2226333/2019-Gold-Book-Final-Public.pdf/a3e8d99f-7164-2b24-e81d-b2c245f67904?t=1556215322968.

[30] "Load data - nyiso." Retrieved from https://www.nyiso.com/load-data.

[31] D. Shi, D. L. Shawhan, N. Li, D. J. Tylavsky, J. T. Taber, R. D. Zimmerman, and W. D. Schulze, "Optimal generation investment planning: Pt. 1: network equivalents," in *2012 North American Power Symposium (NAPS)*, pp. 1–6, IEEE, 2012.

[32] J. Liu, H. Zhang, W. Qiao, and L. Qu, "Dc (optimal) power flow-based models for simulation and mitigation of overload cascading failures," in *2019 North American Power Symposium (NAPS)*, pp. 1–5, 2019.

[33] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas, "Matpower: Steady-state operations, planning, and analysis tools for power systems research and education," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 12–19, 2011.

[34] R. D. Zimmerman and C. E. Murillo-Sanchez, "MATPOWER (version 7.1) [software]." https://matpower.org, 2020.

[35] R. Lincoln, "Pypower." https://github.com/rwl/PYPOWER, 2023.

[36] R. D. Zimmerman and C. E. Murillo-Sanchez, "MATPOWER User's Manual, Version 7.1." https://matpower.org/docs/MATPOWER-manual-7.1.pdf, 2020.

[37] United States Environmental Protection Agency (EPA), "Historical egrid data." Retrieved January 16, 2023, from https://www.epa.gov/egrid/historical-egrid-data, 2023.

[38] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255–260, 2015.

[39] A. K. Nanda, N. S, S. Gupta, A. L. M. Saleth, R. S, and S. Kiran, "Multi-layer perceptron's neural network with optimization algorithm for greenhouse gas forecasting systems," *Environmental Challenges*, vol. 11, p. 100708, 2023.

[40] M.-C. Popescu, V. E. Balas, L. Perescu-Popescu, and N. Mastorakis, "Multilayer perceptron and neural networks," *WSEAS Transactions on Circuits and Systems*, vol. 8, no. 7, pp. 579–588, 2009.

[41] H. Zhang, F. Xiao, C. Zhang, and R. Li, "A multi-agent system based coordinated multi-objective optimal load scheduling strategy using marginal emission factors for building cluster demand response," *Energy and Buildings*, vol. 281, p. 112765, 2023.

[42] P. Stackhouse, "NASA POWER — Data Access Viewer." https://power.larc.nasa.gov/data-access-viewer/.

[43] S. Mujeeb and N. Javaid, "Deep learning-based carbon emissions forecasting and renewable energy's impact quantification," *IET Renewable Power Generation*, vol. 17, no. 4, pp. 873–884, 2023.

[44] P. Hibbard, "Climate change impact phase ii: An assessment of climate change," 2020.

[45] L. Anderson, V. Liu, A. Gupta, and K. Doering, "A spatiotemporal analysis of new york state grid transition under the clcpa energy strategy," tech. rep., University of Hawaii at Manoa, 2023.

[46] K. Hochul, "Nysdec value of carbon guidance 2022 update." New York State Department of Environmental Conservation. Retrieved May 20, 2023, from https://www.dec.ny.gov/docs/administration$_pdf/vocguid22.pdf$, 2022.

[47] R. Ribeiro and N. Moniz, "Imbalanced regression and extreme value prediction," *Machine Learning*, vol. 109, p. 1803–1835, 2020.

[48] L. Torgo, R. P. Ribeiro, B. Pfahringer, and P. Branco, "Smote for regression," in *Portuguese conference on artificial intelligence*, pp. 378–389, Springer, 2013.

[49] P. Branco, L. Torgo, and R. P. Ribeiro, "Smogn: a pre-processing approach for imbalanced regression," in *First international workshop on learning with imbalanced domains: Theory and applications*, pp. 36–50, PMLR, 2017.

[50] D. Salinas, V. Flunkert, J. Gasthaus, and T. Januschowski, "Deepar: Probabilistic forecasting with autoregressive recurrent networks," *International Journal of Forecasting*, vol. 36, no. 3, pp. 1181–1191, 2020.