

THREE ESSAYS ON GREENHOUSE GAS EMISSIONS, ENERGY USE AND CLIMATE CHANGE

by

JAMESON AUGUSTIN

(Under the Direction of Berna Karali)

ABSTRACT

This dissertation examines interconnected aspects of greenhouse gas emissions, energy use, and climate change impacts under the overarching theme of informing emissions mitigation policy priorities in the United States. The first chapter explores causal relationships between economic growth, disaggregated energy consumption across transportation, residential, commercial, and industrial sectors, renewable energy use, and CO_2 emissions using state-level data from 1997-2020. Applying panel vector error correction models and fully modified ordinary least squares estimation, findings reveal complex short-run interactions and significant long-run impacts of all energy sectors on emissions, with the transportation and industrial sectors exerting the largest effects. Evidence supports the Environmental Kuznets Curve hypothesis, and renewable energy demonstrates emissions reduction potential. The second chapter evaluates the causal impact of state-level net metering policies on residential greenhouse gas emissions over 1990-2020 using the Callaway and Sant'Anna (2021) estimator to address potential biases in staggered policy adoption. Results indicate modest but statistically significant emissions reductions that intensify over time, with substantial heterogeneity across policy designs, political contexts, and underlying mechanisms of solar PV adoption and grid interaction. The final chapter leverages remote sensing data and machine learning models to predict county-level rice yields and associated methane emissions across six

major U.S. rice-producing states from 2008-2022. XGBoost and Explainable Boosting Machines accurately forecast yields as early as April-June, with soil properties emerging as key predictors. Exploring yield-emissions trade-offs reveals a positive correlation between yield improvement and methane reduction. Collectively, the studies advance integrated empirical assessments of historic relationships between economic activity, energy systems, and greenhouse gas emissions across scales. Findings directly inform prioritizing policy portfolios blending incentives, mandates, and market reforms—from targeted strategies in high-impact sectors and net metering enhancements to agricultural extension programs—to balance continued U.S. prosperity with climate resilience and global leadership. Demonstrating interlinkages across sectoral systems, gas regulations, and regional climate impacts cements the urgency and efficacy of coordinated federal action in partnership with state and local initiatives.

INDEX WORDS: [Greenhouse Gas Emissions, Energy Use, Net Metering, Yield Prediction, Climate Change]

THREE ESSAYS ON GREENHOUSE GAS EMISSIONS, ENERGY USE AND CLIMATE
CHANGE

by

JAMESON AUGUSTIN

Bachelor of Agribusiness, Zamorano University, Honduras, 2019

Master of Agribusiness, Texas A&M-Commerce, 2021

A Dissertation Submitted to the Graduate Faculty of the
University of Georgia in Partial Fulfillment of the Requirements for the Degree.

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2024

©2024

Jameson Augustin

All Rights Reserved

THREE ESSAYS ON GREENHOUSE GAS EMISSIONS, ENERGY USE AND CLIMATE
CHANGE

by

JAMESON AUGUSTIN

Major Professor: Berna Karali

Committee: Gopinath Munisamy

Mateusz Filipski

Susana Ferreira

Electronic Version Approved:

Ron Walcott

Dean of the Graduate School

The University of Georgia

December 2024

DEDICATION

I dedicate this dissertation to my wife Alexia Augustin, and my parents for their unwavering support, especially my mom, Saintanise Bazil, who has been very supportive since my early education, challenging me to always reach for the stars.

ACKNOWLEDGMENTS

This dissertation would not be possible without the unconditional support of my advisor Dr. Berna Karali. Her thoughtful and timely feedback guided me through every step. I am also thankful for my committee members Dr. Gopinath Munisamy, Dr. Mateusz Filipski and Dr. Susana Ferreira for their invaluable contribution throughout this journey.

CONTENTS

Acknowledgments	v
List of Figures	vii
List of Tables	ix
1 Introduction	1
2 Causal Links between Greenhouse Gas Emissions, Energy Consumption and Economic Growth in the U.S.: A Sectoral Analysis	5
2.1 Abstract	6
2.2 Introduction	7
2.3 Literature Review	10
2.4 Data	13
2.5 Econometric Methods and Results	17
2.6 Conclusions	34
3 Measuring the Impact of Net Metering Policy on Residential GHG Emissions in the U.S.: A Difference-in-Difference Analysis	36
3.1 Abstract	37
3.2 Introduction	38
3.3 U.S. Net Metering Policy Overview	41
3.4 Literature Review	44
3.5 Methodology and Empirical Strategy	45
3.6 Data	53

3.7	Results	59
3.8	Discussions	78
3.9	Conclusions	81
4	Leveraging Remote Sensing and Machine Learning to Predict U.S. County-level Rice Yield: The Role of GHG Emissions	83
4.1	Abstract	84
4.2	Introduction	85
4.3	Literature Review	91
4.4	Data	94
4.5	Methods	100
4.6	Results	109
4.7	Discussions	131
4.8	Conclusions and Policy Implications	133
5	Conclusions	136
	Bibliography	139
	Appendices	150
A	CHAPTER 2	150
B	CHAPTER 3	153
B.1	A Note on Residential GHG Emissions	153
C	CHAPTER 4	154

LIST OF FIGURES

2.1	Patterns of Emissions and Energy Consumption across States (1997-2020)	14
2.2	Causal relationships based on VEC Granger causality tests for the augmented model.	28
2.3	Generalized Impulse Response Functions for the Augmented Model.	29
3.1	Net Metering Policy Adoption Year by State	54
3.2	Average Residential GHG Emissions per Capita by State (MT; 1990-2020).	55
3.3	Average Residential GHG Emissions across Years (MMT; 1990-2020).	56
3.4	Average Residential Energy Use per Capita by State (Thousands BTU; 1990-2020).	57
3.5	Impact of Net Metering Adoption on Residential GHG Emissions Relative to Adoption Year.	61
3.6	Heterogeneity between States of More versus Less Favorable Net Metering Policies.	63
3.7	Heterogeneity by Policy Traits.	65
3.8	Heterogeneity between Democratic versus Republican States.	66
3.9	Cumulative Effects of the Net Metering Policy.	69
3.10	Event Study Results Accounting for Policy Diffusion.	72
3.11	Placebo Tests: Event Studies across Specifications.	76
4.1	States and Counties (Darker Areas) where Rice is Grown.	94
4.2	Comparison of Average Yield and Rice-based Methane Emissions by County.	98
4.3	Panel of Spatiotemporal Plots.	111
4.4	Feature Importance for the EBM and XGBoost Models.	117
4.5	Comparison of Actual vs Predicted Yield over Space by Model.	119
4.6	Yield-Emissions Trade-off Plot.	128
4.7	PDPs for Common Important Features across Yield and Emission Models.	130

A.1	Causal relationships based on VEC Granger causality tests for the Base Model. . .	150
A.2	Generalized Impulse Response Functions for the Base Model.	151
C.1	Matrix of Plots for Univariate Analysis.	155
C.2	Correlation Heatmap.	157
C.3	Boxplot-based Outlier Detection.	159
C.4	Data Drift Tests.	160

LIST OF TABLES

2.1	Summary of Variables.	14
2.2	Descriptive Statistics Summary for Transformed Variables.	15
2.3	Pairwise Correlation Coefficients with Significance Levels.	16
2.4	Panel unit root test results.	19
2.5	Cointegration Test Results.	21
2.6	Unrestricted Cointegration Rank Tests for the Base Model.	25
2.7	Unrestricted Cointegration Rank Tests for the Augmented Model.	26
2.8	Short-run Granger Causality/Block Exogeneity Wald Tests Summary for the Augmented Model.	27
2.9	FMOLS Results for augmented model.	32
3.1	Net Metering Policy Characteristics Summary.	51
3.2	Variable Description and Summary Statistics.	54
3.3	Balancing checks.	58
3.4	Impact of Net Metering on Residential GHG Emissions.	60
3.5	Overall ATTs and Signs of the Diffusion Variable across Specifications.	71
3.6	Placebo Tests: Overall ATTs across Specifications.	75
3.7	Understanding Mechanisms: DiD Estimates.	79
4.1	Summary of Data Sources and Variables.	96
4.2	Summary Statistics of Variables with Units and Resolution.	100
4.3	Comparison of Model Performance Across Different Settings.	115
4.4	Model Performance with Out-of-Time Validation.	121
4.5	Out-of-Time Prediction with Rolling Window Approach.	122

4.6	Model Performance for Early vs Late Season.	123
4.7	Model Performance for Yield and Emission Models.	124
4.8	EBM-based Feature Importance: Yield vs. Emissions Models.	125
A.1	Short-run Granger Causality/Block Exogeneity Wald Tests Summary for the Base Model.	150
A.2	FMOLS Results for the Base Model.	152

CHAPTER 1

INTRODUCTION

The United States bears unique responsibilities and wields unrivaled capacity in the global fight against climate change. As the world's largest historical emitter accounting for over one-quarter of cumulative CO_2 emissions (Ritchie, 2019), the nation faces mounting pressure to drastically curtail its outsized carbon footprint. Failure to bend the emissions curve risks jeopardizing both the 1.5°C pathway of the Paris Agreement goal and U.S. credibility in international climate diplomacy.¹ With per capita emissions triple the world average and GDP exceeding \$25 trillion, the U.S. also possesses the financial and technological means to pioneer a net-zero transition that decouples growth from GHG emissions. Recent climate disasters across the country, from record California wildfires to Gulf Coast hurricanes, not only expose the staggering human and economic costs of inaction but also the potential for transformative climate leadership rooted in national self-interest. This dissertation seeks to advance rigorous, policy-oriented analysis of U.S. emissions drivers, interdependencies, and mitigation pathways across key sectors to accelerate economy-wide decarbonization commensurate with the scale of the climate crisis.

Despite substantial cross-state variation in energy mixes, policies, and emissions profiles, prior U.S. focused scholarship often remains confined to national-level, sectorally-aggregated analyses of the emissions-energy-growth nexus. The first chapter addresses this gap by employing a vector error correction model and fully modified OLS to uncover the dynamic causal relationships between economic growth, energy consumption patterns disaggregated across residential, commercial, industrial, transportation, and renewable sectors, and CO_2 emissions at the state level over 1997-2020. Granger causality and impulse response results demonstrate the outsized influence of the transportation and industrial sectors on overall emissions, the mitigating impact of renewables,

¹The 1.5°C pathway refers to the global goal set by the Paris Agreement in December 2015 to limit the increase in average global temperatures to 1.5 degrees Celsius above pre-industrial levels.

and an Environmental Kuznets Curve effect of income growth. The findings argue for differentiated sector-specific emissions reduction policies attuned to heterogeneous energy demand drivers and decarbonization pathways.

Clean energy resource deployment on the state and local scales remains integral to national emissions reductions, given electricity and associated emissions are regulated primarily by state policies. Consequently, Chapter 2 exploits state-level variation in net metering policies, foundational to distributed solar adoption, to rigorously quantify the average impact on residential emissions over 1990-2020. Leveraging a staggered difference-in-differences estimator robust to heterogeneous treatment effects, together with Smith et al. (2021) database of 5 key net metering design traits, the causal analysis reveals modest (0.3-0.6%) but progressively increasing emissions reductions in the 3-5 year period post-implementation. The striking divergence across policy favorability levels, spillovers, partisan orientation, and individual design features demonstrates the crucial importance of incentive structures. Surprisingly, Republican-leaning states exhibit stronger effects, likely due to lower solar baselines. The results illuminate the complex policy landscapes governing state clean energy transitions and offer generalizable insights for optimizing net-metering frameworks.

Agriculture and associated methane emissions have remained a comparatively neglected component of U.S. climate policy. Addressing this lacuna, Chapter 3 harnesses explainable machine learning and remote sensing to jointly model and map rice yields and methane emissions across 67 counties in six major rice-producing states over 2008-2022. Combining satellite-derived vegetation indices, climate reanalysis, and soil variables in multiple black-box and glass-box machine learning frameworks, the analysis reveals the complex agro-climatic and management determinants of rice productivity. Model skill at early-season yield prediction holds promise for guiding precision agriculture. Intriguingly, the multi-objective genetic algorithm identifies a positive association between yield gains and methane mitigation, indicating synergies through sustainable intensification. The findings demonstrate the untapped potential for data-driven optimization of climate-smart practices.

Collectively, the essays present a thematically unified analysis of the entangled impacts of energy, economic structures, and policies on U.S. GHG emissions across electricity, transportation, and agricultural systems. Integrating methods spanning time-series econometrics, quasi-experimental causal inference, and machine learning, the studies not only reveal the preeminence of fossil-oriented infrastructure and policy regimes in driving emissions growth but also the manifold decarbonization opportunities from renewable energy expansions, electricity market reforms, and precision agriculture. The recurring focus on heterogeneity—across states, economic sectors, policy attributes, and agroecological regions—underscores the necessity of sectorally-differentiated, contextually-tailored mitigation approaches in lieu of blunt, one-size-fits-all solutions.

Ultimately, this dissertation seeks to both deepen scholarly understanding of U.S. emissions patterns and characteristics and directly inform policy strategies commensurate to achieving net-zero emissions. Prioritizing sector-specific interventions, such as vehicle electrification, market-based industrial carbon pricing, optimized net metering for distributed renewables, and financial and technical assistance for lower-emission rice cultivation practices emerge as prime policy levers. Equally crucially, the cross-cutting findings highlight the catalytic potential of integrated federal climate policies, such as an economy-wide carbon price or national clean energy standard, to harmonize state-level actions, prevent emissions leakage, and accelerate decarbonization through aligned incentives and predictable investment signals. The transportation and industrial sectors emissions predominance further reinforces the need for increased infrastructure spending, along with Research, Development and Demonstration (RD&D) efforts to scale zero-carbon technological solutions.

This dissertation arrives at a pivotal juncture for U.S. climate action, as cascading extreme weather events, intensifying geopolitical tensions around fossil fuels, and landmark federal investments converge to stimulate economy-wide transitions. While recent policies like the 2021 Infrastructure Investment and Jobs Act and the 2022 Inflation Reduction Act represent historic down payments in clean energy tax credits, low-carbon technology incentives, and climate resilience,

they likely prove insufficient to achieve the halving of emissions by 2030 required for the 1.5°C pathway of the Paris Agreement. As the world's second-largest emitter grappling with existential climate risks, the U.S. must drastically increase decarbonization across all sectors this decade to secure a prosperous, equitable, and sustainable future. In elucidating policy-relevant insights into energy-emissions-economy dynamics at subnational scales, this dissertation aspires to both spur more granular, impacts-oriented scholarship and illuminate evidence-based strategies for bold, immediate, and sustained federal climate leadership. With the window for limiting warming rapidly closing, embracing comprehensive mitigation frameworks supported by rigorous integrated policy assessments remains paramount. The U.S., and indeed the world, can scarcely afford to wait.

CHAPTER 2

CAUSAL LINKS BETWEEN GREENHOUSE GAS EMISSIONS, ENERGY CONSUMPTION AND ECONOMIC GROWTH IN THE U.S.: A SECTORAL ANALYSIS¹

¹J. Augustin & B. Karali. To be submitted to Energy Economics.

2.1 Abstract

This study examines the dynamic causal relationships between sectoral energy consumption, renewable energy use, economic growth, and carbon dioxide (CO_2) emissions in the United States from 1997 to 2020 using state-level data. We apply a panel vector error correction model (VECM) and fully modified ordinary least squares (FMOLS) method to analyze both short-run and long-run linkages among these variables. Our approach extends existing literature through the disaggregation of energy consumption into residential, commercial, industrial, and transportation sectors, which offers a more nuanced understanding of the emissions-energy-economy nexus. Short-run Granger causality tests reveal complex interactions, with the residential and transportation sectors as key influencers of other energy sectors, emissions, and economic activity. Long-run FMOLS estimates show that all energy sectors significantly impact emissions, with the transportation sector exerting the largest effect, followed by the industrial sector. We find evidence to support the environmental Kuznets curve (EKC) hypothesis, which suggests that economic growth leads to decreased emissions at higher income levels. Renewable energy use demonstrates a significant negative impact on emissions, which highlights its importance in mitigation efforts. Impulse response functions further elucidate the dynamic relationships among variables, and reveal sector-specific shocks and their propagation through the system. Our findings highlight the need for targeted strategies in high-impact sectors such as the transportation and industrial sectors, continued support for renewable energy deployment, and consideration of sector-specific dynamics in the design of comprehensive climate policies.

2.2 Introduction

Global climate change poses one of the most significant threats to human civilization in the 21st century. The rapid rise in global temperatures, driven largely by increasing concentrations of greenhouse gases (GHGs) like carbon dioxide (CO_2) in the atmosphere, is already impacting weather patterns, sea levels, ecosystems, and human communities around the world (Bishnoi et al., 2022). In 2021, total U.S. GHG emissions were 6.34 billion metric tons of CO_2 equivalent, with CO_2 accounting for approximately 80% of this total. Over 70% of U.S. GHG emissions come from burning fossil fuels across the energy, transportation, commercial, residential, and industrial sectors (Environmental Protection Agency, 2023).

As the second-largest global emitter, accounting for at least 13% of cumulative global CO_2 emissions since 1751, the United States bears a significant responsibility for driving climate change (Liu & Yuan, 2023). Achieving substantial, sustained emissions reductions in the United States is critical to limiting global temperature rise and preventing the most catastrophic climate change impacts. Understanding the key drivers of U.S. emissions growth across different economic sectors is essential for policymakers to design effective tools and pathways to curb emissions.

Fundamentally, GHG emissions are linked to economic activity and energy consumption, as carbon-intensive fossil fuels like coal, oil, and natural gas meet over 80% of U.S. energy demand (Environmental Protection Agency, 2023). On the global level, rising affluence and energy consumption, both resulting from and driving economic growth, accelerate GHG emissions (Zaman & el Moemen, 2017). The same holds in the United States, where the sectors that form the backbone of economic activity—transportation, electricity, industry, commercial enterprises, and residences—also produce the vast majority of national emissions.

In the United States over the past half-century, bursts of economic growth have tended to increase energy demand and boost emissions, while periods of stagnation and recession have suppressed emissions. For example, U.S. GDP growth averaged over 7% annually between 1950

and 1973 amid booming post-war industrial activity and consumption, coinciding with a steady rise in energy use and CO_2 emissions (Dehdar et al., 2023). Conversely, the 2008 Global Financial Crisis and the Great Recession brought a decline in the U.S. GDP, energy use, and emissions. In 2021, a 5.7% surge in U.S. GDP drove CO_2 emissions from fossil fuel combustion up 6.6% over 2020 levels, confirming the positive relationship between economic activity, energy, and emissions (Environmental Protection Agency, 2023).

While the connections between energy, emissions, and growth are well established at the national level, analyzing these relationships at the sectoral level reveals a more complex picture. Previous studies have explored the emissions-energy-economy nexus through three interconnected strands (Ozturk & Acaravci, 2010):

1. The environment-income relationship centered around tests of the environmental Kuznets curve (EKC) hypothesis, which states that emissions initially rise but eventually fall with income growth after some turning point.
2. The energy-growth linkage centered around the role of energy as a vital input to economic production and growth.
3. The combined dynamic interactions between economic output, energy use, and environmental pollutants.

The third integrated approach encapsulates and expands upon the first two by introducing energy as a central mediator linking economic growth and emissions over time. However, most studies utilizing this framework focus on national or cross-country analyses rather than disaggregated sectoral analyses within the United States.

Understanding these sectoral relationships is crucial, as the main sources of U.S. emissions differ widely in their intensities. Electricity and transportation together produced over half of the national CO_2 emissions in 2021, reflecting their heavy reliance on fossil fuel combustion (Environmental Protection Agency, 2023). Meanwhile, the industrial sector accounted for 23% of

emissions through direct fossil fuel use and electricity consumption in manufacturing processes, edging out the combined commercial and residential sectors at 13% (Environmental Protection Agency, 2023). Agricultural emissions comprised the remaining 10% share (Environmental Protection Agency, 2023). These proportions shifted between 1990 and 2021, with transportation emissions growing while the power sector's share declined amid the transition from coal to natural gas and renewables (Environmental Protection Agency, 2023).

Capturing these complex, evolving sectoral dynamics is necessary for targeted policy interventions to curb emissions. If income growth drives energy consumption and emissions uniformly across all sectors, economy-wide measures like carbon taxes or energy efficiency standards may suffice to mitigate emissions. However, if relationships diverge between sectors, tailoring policies to specific emissions-intensive activities may prove more effective. For example, the rapid growth of zero-carbon renewable energy in the utility sector is decoupling electricity generation from fossil fuel combustion and emissions, while electric vehicles (EVs) powered by low-carbon electricity offer a pathway to reduce transportation emissions.

Despite the importance of sectoral analysis, studies examining the emissions-energy-income nexus in the United States remain surprisingly scant. Existing research often focuses on national aggregates or individual sectors in isolation (Marrero, 2010; Acaravci & Ozturk, 2010; Menyah & Wolde-Rufael, 2010; Pao & Tsai, 2010; Hamit-Haggar, 2012; Paramati et al., 2017; Acheampong, 2018; Salari et al., 2021), lacking a connected sectoral analysis of how economic output induces greater energy consumption and emissions across different sectors.

This paper bridges this research gap by analyzing dynamic causal relationships between sectoral economic growth, energy consumption patterns, and CO_2 emissions in the United States from 1997 to 2020. It provides a comprehensive sectoral analysis of the emissions-energy-economy nexus in the United States, offering novel insights into both short-run and long-run dynamics among these variables through time series econometric techniques, including panel vector error correction modeling and fully modified ordinary least squares estimation.

Our findings reveal intricate relationships among sectoral energy use, emissions, and economic growth in the United States. Short-run Granger causality tests indicate that transportation energy use strongly influences CO_2 emissions, other energy sectors, and economic growth, while residential energy use impacts commercial and industrial sectors. The impulse response analysis shows that CO_2 emissions are most sensitive to shocks in transportation and industrial energy use. Our long-run FMOLS estimates confirm that all energy sectors positively impact emissions, with transportation having the largest effect, followed by the industrial sector. Importantly, we find evidence supporting the EKC hypothesis, suggesting that economic growth eventually leads to decreased emissions at higher income levels. The use of renewable energy demonstrates a significant negative impact on emissions, which highlights its importance in mitigation efforts.

The remainder of the paper is structured as follows: Section 3 reviews relevant literature on emissions, energy consumption, and economic growth. Section 4 describes our data sources and provides summary statistics. Section 5 outlines our econometric methods and presents results from Granger causality tests, impulse response functions, and FMOLS estimation. Section 6 concludes with an emphasis on policy implications.

2.3 Literature Review

Greenhouse gas emissions, energy consumption, and economic growth are intricately linked in complex, multi-directional relationships often studied through the lens of the EKC hypothesis. This hypothesis posits an inverted U-shaped relationship where economic development initially increases environmental degradation and then decreases it after reaching a turning point (Acaravci & Ozturk, 2010; Apergis & Payne, 2010; Menyah & Wolde-Rufael, 2010; Ozturk & Acaravci, 2010; Marrero, 2010). However, empirical findings frequently contradict or fail to universally support the EKC hypothesis across regions, pollution types, and time. More granular, sectoral analyses may clarify inconsistencies found at the national aggregate level (Acaravci & Ozturk, 2010; Begum et al., 2015). Causality testing helps determine lead-lag linkages between emissions, energy, and GDP as a step

toward informed policy. For example, bidirectional Granger causality implies the ability of economic and climate policies to restrict emissions without sacrificing growth (Apergis & Payne, 2010; Pao & Tsai, 2010; Wang et al., 2011; Begum et al., 2015). Yet unidirectional causality from GDP to emissions may suggest policy tradeoffs between economic development and environmental goals (Menyah & Wolde-Rufael, 2010; Ozturk & Acaravci, 2010; Begum et al., 2015). Similarly, the causation between energy and emissions highlights opportunities for mitigation through improved efficiency, technological change, or shifts between renewable and fossil fuel sources (Menyah & Wolde-Rufael, 2010; Wang et al., 2011).

Methodologically, previous literatures have utilized multivariate time series, panel data, or quantile regressions to study these relationships within nations including China, Brazil, Russia, Turkey, Malaysia, Canada, and the U.S. (Acaravci & Ozturk, 2010; Ozturk & Acaravci, 2010; Marrero, 2010; Pao & Tsai, 2010; Pao et al., 2011a; Begum et al., 2015; Paramati et al., 2017; Zaman & el Moemen, 2017; Salari et al., 2021; Liu & Yuan, 2023; Dehdar et al., 2023). Vector autoregressive (VAR) model, vector error correction model (VECM), and Granger causality tests determine short- and long-run dynamics (Apergis & Payne, 2010; Wang et al., 2011; Begum et al., 2015; Jian et al., 2019; Liu & Yuan, 2023). Previous studies utilizing panel cointegration techniques demonstrate the existence of a long-run equilibrium relationship between emissions, energy consumption, and GDP across countries (Acaravci & Ozturk, 2010; Apergis & Payne, 2010; Pao & Tsai, 2010; Wang et al., 2011; Begum et al., 2015). Estimated long-run elasticities demonstrate the substantial influence of energy consumption on emissions and the potential mitigating impact of GDP growth in some cases (Apergis & Payne, 2010; Begum et al., 2015). However, evidence supporting the EKC hypothesis remains mixed regarding the existence of an inverted U-shaped effect of income on pollution (Acaravci & Ozturk, 2010; Pao & Tsai, 2010, 2011b; Begum et al., 2015).

Critically assessing causal directions proves useful for policy even amidst conflicting empirical results across studies. Many studies find evidence of short- and long-run feedback

from energy consumption to emissions, highlighting the continued relevance of energy efficiency measures (Apergis & Payne, 2010; Wang et al., 2011; Pao et al., 2011a; Begum et al., 2015). Meanwhile, linkages from economic growth demonstrate the necessity of sustainable technological change and development policy to avoid restrictive tradeoffs on emissions (Menyah & Wolde-Rufael, 2010; Ozturk & Acaravci, 2010; Wang et al., 2011; Begum et al., 2015). Comparing developed and developing countries also highlights differential policy needs and challenges (Paramati et al., 2017). Advanced economies with service-oriented GDP may weaken energy-growth connections and environmental pressures relative to industrialized nations (Zaman & el Moemen, 2017). This suggests sector-specific examinations to illuminate policy opportunities within diverse economic structures.

While broad trends emerge across studies that use aggregated sectors, tracing nuanced causality for emissions, energy, and income often yields conflicting results between studies. Disaggregated, dynamic analyses, therefore, offer analytical traction amidst ambiguity in macro studies (Salari et al., 2021; Liu & Yuan, 2023; Dehdar et al., 2023). Data limitations may explain discrepancies, as longer time series strengthen robustness and sector-specific metrics enhance relevance over aggregated indices (Liu & Yuan, 2023). Quantile regressions also allow the examination of changing effects across development levels akin to the EKC, corroborating emblematic non-linear emissions patterns (Liu & Yuan, 2023; Dehdar et al., 2023). Recent U.S. state-level studies apply these techniques to determine policy-relevant causal influences from economic growth, fossil fuel, and renewable consumption on emissions across income distributions (Salari et al., 2021; Liu & Yuan, 2023; Dehdar et al., 2023). Our sector-specific study for the United States aims to provide illumination to the ambiguity resulting from aggregated analyses. Comparing evidence across macro studies for China, Europe, Russia, and Canada suggests a lack of definitive consensus regarding the direction of causality and the presence of non-linear inverted U-shaped income dynamics (Marrero, 2010). However, examinations in Canada's industrial sector demonstrate differentiated short- and long-run impacts of income, energy consumption, and policy

efforts (Hamit-Haggar, 2012). Previous state-level research in the U.S. corroborates heterogeneity in emissions determinants and mitigation opportunities between renewable integration, efficiency improvements, and fostering transitions to cleaner economic growth pathways (Salari et al., 2021; Liu & Yuan, 2023; Dehdar et al., 2023).

Our study addresses key gaps in the literature by providing a comprehensive, sector-specific analysis of the emissions-energy-economy nexus in the United States. By disaggregating energy consumption across residential, commercial, industrial, and transportation sectors, we capture heterogeneous impacts often overlooked in aggregate analyses. Our methodological approach, combining panel VECM, Granger causality, impulse response functions (IRFs), and FMOLS estimation, allows for a nuanced examination of both short-run and long-run dynamics. This granular, state-level analysis offers timely insights into emissions drivers and mitigation strategies across different economic sectors. By doing so, we contribute to the ongoing debate on the validity of the EKC hypothesis and provide policymakers with targeted, sector-specific evidence to inform climate change interventions while considering diverse economic dynamics.

2.4 Data

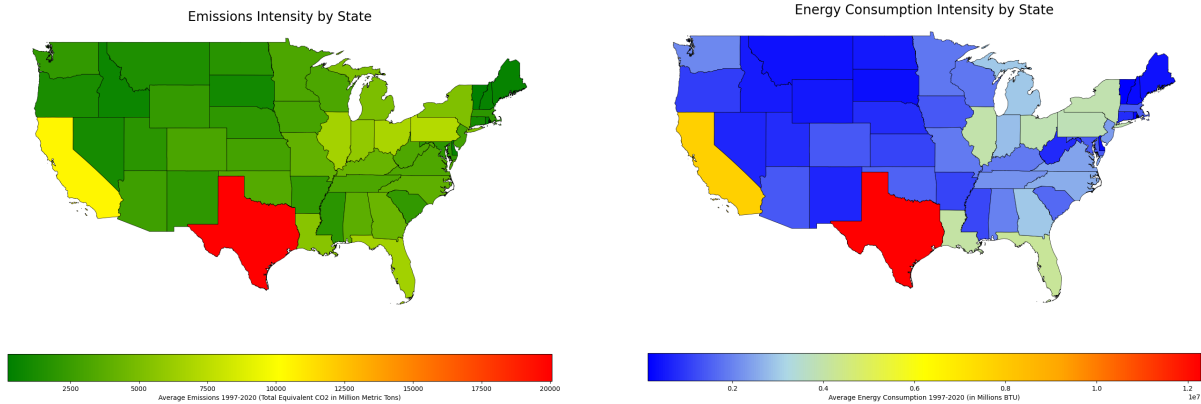
We use panel data from 48 U.S. states, Alaska, and Hawaii for the period 1997-2020. Table 2.1 presents data sources and descriptions of the variables. Total equivalent CO_2 emissions (in million metric tons) are obtained from the U.S. Environmental Protection Agency (EPA), while all other variables are sourced from the U.S. Energy Information Administration (EIA). The retail electricity price, expressed in dollars per million BTU, serves as a proxy for energy prices.

Figure 2.1 presents average variations in energy use and total emissions across states over the 24 years. Texas and California are the top CO_2 emitters and also the top energy consumers, as expected. In general, states with higher emissions tend to consume more energy. On the contrary, states with low emissions, such as Maine, are at the bottom of the spectrum for total energy consumption.

Table 2.1: Summary of Variables.

Variable	Description	Unit	Data Source
RGDP	Real GDP	Millions of (2012) USD	EIA
NONRENEW	Total Nonrenewable Energy Use	Million BTU	EIA
POP	Total Population	Thousands	EIA
HDD	Heating Degree Days	Degree Days	EIA
CDD	Cooling Degree Days	Degree Days	EIA
RENEW	Total Renewable Energy Use	Million BTU	EIA
RESID	Residential Energy Use	Million BTU	EIA
COMM	Commercial Energy Use	Million BTU	EIA
INDUST	Industrial Energy Use	Million BTU	EIA
TRANSP	Transportation Energy Use	Million BTU	EIA
CO2	Total CO ₂ Emissions	Million Metric Tons (MMT)	EPA
PRC	Retail Electricity Price	Dollars per Million BTU	EIA

Notes: The terms *Consumption* and *Use* are interchangeably used in this document.



(a) Average Carbon Dioxide Emissions by State

(b) Average Energy Consumption by State

Figure 2.1: Patterns of Emissions and Energy Consumption across States (1997-2020)

To facilitate comparison across states while accounting for size differences, all variables except for HDD, CDD, and energy price, are divided by the state population to express them per capita except for energy price and then logged (natural logarithm). The log transformation allows for the interpretation of the estimated coefficients as elasticities².

²Additionally, HDD is divided by a factor of 1,000 to facilitate its interpretation.

Table 2.2 displays the descriptive statistics for the variables used in the analysis. The sample consists of 1,200 state-year observations. Most variables show evidence of right-skewed distributions, particularly in energy consumption and CO2 emissions per capita. This skewness suggests that while most states have moderate levels of energy use and emissions, a few states have significantly higher values, potentially due to differences in industrial composition or energy policies. Real GDP per capita appears more symmetrically distributed, indicating more even economic development across states. Sectoral energy use patterns vary, with industrial energy use displaying the most skewed distribution, likely reflecting the uneven distribution of energy-intensive industries across states.

Table 2.2: Descriptive Statistics Summary for Transformed Variables.

Variables	Mean	Std. Dev.	Min	Max
HDD	5.164	2.286	0.000	11.702
lnRGDPpc	-3.036	0.194	-3.508	-2.520
lnNONRENEWpc	-1.078	0.372	-1.831	0.179
lnRENEWpc	-3.704	0.933	-6.419	-1.485
lnRESIDpc	-2.638	0.222	-3.822	-2.244
lnCOMMpc	-2.817	0.220	-3.608	-2.062
lnINDUSTpc	-2.338	0.804	-4.258	-0.382
lnTRANSPpc	-2.353	0.287	-3.095	-0.908
lnCO2pc	-10.583	0.614	-11.694	-8.559
lnPRC	2.727	0.357	1.625	3.653
Observations	1,200			

Table 2.3 presents the pairwise correlation coefficients between the key variables, along with their statistical significance levels. Heating and cooling degree days exhibit a strong negative correlation (-0.867), as expected, given their inverse relationship. For this reason, the latter is dropped from the analysis. Real GDP per capita shows a positive and statistically significant correlation with the overall energy price (0.358) and a negative correlation with the natural logarithm of energy consumption per capita (-0.425). The correlation between the natural logarithm of real GDP per capita and its squared term is 0.402, indicating a potential non-linear relationship between

economic growth and energy consumption. The sectoral energy use variables display positive and significant correlations with the natural logarithm of energy consumption per capita, ranging from 0.283 to 0.949. For this reason, in the further sections, we consider two models: one with the total nonrenewable and renewable energy categories and another with the sectoral energy categories in addition to renewable energy.

Table 2.3: Pairwise Correlation Coefficients with Significance Levels.

	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)
HDD (a)	1.00 (0.00)	-0.87*** (0.00)	-0.08* (0.01)	0.26*** (0.00)	-0.34*** (0.00)	0.24*** (0.00)	0.28*** (0.00)	0.42*** (0.00)	0.48*** (0.00)	0.18*** (0.00)	0.15*** (0.00)
CDD (b)		1.00	0.09** (0.00)	-0.21*** (0.00)	0.16*** (0.00)	-0.07* (0.02)	-0.20*** (0.00)	-0.42*** (0.00)	-0.38*** (0.00)	-0.04 (0.15)	0.08* (0.01)
PRC (c)			1.00	0.36*** (0.00)	0.19*** (0.00)	-0.42*** (0.00)	0.06* (0.05)	-0.31*** (0.00)	-0.25*** (0.00)	-0.47*** (0.00)	-0.29*** (0.00)
RGDPpc (d)				1.00	0.40*** (0.00)	0.02 (0.47)	-0.10** (0.00)	-0.16*** (0.00)	0.24*** (0.00)	-0.10** (0.00)	0.05 (0.08)
RGDP2pc (e)					1.00	-0.31*** (0.00)	-0.32*** (0.00)	-0.20*** (0.00)	-0.12** (0.00)	-0.31*** (0.00)	-0.38*** (0.00)
NONRENEWpc (f)						1.00	0.28*** (0.00)	0.49*** (0.00)	0.62*** (0.00)	0.95*** (0.00)	0.86*** (0.00)
RENEWpc (g)							1.00	0.17*** (0.00)	0.10** (0.00)	0.32*** (0.00)	0.24*** (0.00)
RESIDpc (h)								1.00	0.70*** (0.00)	0.42*** (0.00)	0.20*** (0.00)
COMMpc (i)									1.00	0.46*** (0.00)	0.49*** (0.00)
INDUSTpc (j)										1.00	0.77*** (0.00)
TRANSPpc (k)											1.00
CO2pc (l)	0.29*** (0.00)	-0.10* (0.00)	-0.43*** (0.00)	-0.12** (0.00)	-0.44*** (0.00)	0.87*** (0.00)	0.22*** (0.00)	0.48*** (0.00)	0.62*** (0.00)	0.84*** (0.00)	0.75*** (0.00)

Notes: *** p<0.01, ** p<0.05, * p<0.1. All variables, except for HDD, CDD, and PRC, are per capita.

2.5 Econometric Methods and Results

2.5.1 Panel Unit Root Tests

Panel unit root tests have become increasingly common in econometric analysis, as they provide a powerful framework for examining the stationarity properties of variables in a panel data framework. These tests are crucial for understanding the dynamic behavior of economic variables and have implications for model specification, estimation, and inference (Maddala & Wu, 1999; Hadri, 2000; Breitung, 2001; Im et al., 2003; Levin et al., 2002).

One of the seminal contributions in this area is the work of Levin et al. (2002), who developed a panel unit root test that accounts for cross-section-specific deterministic trends, heterogeneous autoregressive structures, and serially correlated errors across cross-sectional units. However, their test assumes a common autoregressive parameter across cross-sectional units under the alternative hypothesis, which may be restrictive in practice. To overcome this limitation, Im et al. (2003) proposed a more flexible panel unit root test that allows for heterogeneous autoregressive coefficients. Their test statistic is constructed by averaging individual augmented Dickey-Fuller (ADF) test statistics, which enables capturing a wider range of dynamic behaviors across cross-sectional units.

While these tests have been widely applied, they are not without limitations. In particular, including cross-section-specific trends can lead to a significant loss of power due to the bias correction employed in the testing procedures. To address this issue, Breitung (2001) developed a panel unit root test that does not rely on bias correction factors and has been shown to exhibit superior power and size properties compared to the tests of Levin et al. (2002) and Im et al. (2003). An alternative approach to panel unit root testing is the Fisher-type test proposed by Maddala & Wu (1999). This test combines p-values from individual unit root tests and offers several advantages, such as the ability to handle unbalanced panels and varying lag lengths across cross-sectional units. However, the p-values need to be obtained through simulation techniques, which can be computationally intensive. Extending the well-known Kwiatkowski-Phillips-Schmidt-Shin (KPSS)

test (Kwiatkowski et al., 1992) from the time series domain to the panel data framework, Hadri (2000) developed a residual-based Lagrange multiplier test for stationarity. This test provides a complementary perspective to the aforementioned panel unit root tests, as it allows for testing the null hypothesis of stationarity against the alternative of a unit root.

The results in Table 2.4 provide strong evidence, across at least half of the tests for each series and regardless of including a trend, that all variables except for HDD are non-stationary in levels but become stationary after taking the first differences. Thus, these variables are integrated of order one, or $I(1)$. Considering the real GDP per capita ($\ln\text{RGDPpc}$) and its squared term ($\ln\text{RGDPpcSQ}$), the Levin, Lin, and Chu (LLC) test rejects the null hypothesis of a unit root in levels at the 1% significance level, both with and without a trend. However, the Breitung test fails to reject the null in levels. For the first differenced series, all tests strongly reject the null of non-stationarity, confirming that the GDP variables are $I(1)$. For nonrenewable energy consumption ($\ln\text{NONRENEWpc}$), renewable energy use ($\ln\text{RENEWpc}$), and all sectoral energy use variables, the tests generally fail to reject the null of a unit root in levels (with some exceptions when a trend is included) but strongly reject the null in the first differences. This indicates that these variables are also $I(1)$. The CO_2 emissions variable ($\ln\text{CO2pc}$) and energy price ($\ln\text{PRC}$) show similar patterns, with most tests failing to reject the null of a unit root in levels but rejecting it in the first differences. The heating degree days variable (HDD) appears stationary in both levels and first differences based on most tests.

Contrary to the previous tests explained, the null hypothesis of the Hadri test is stationarity. Table 2.4 shows that we reject its null hypothesis for most variables in levels but fail to reject it in the first differences, further confirming the $I(1)$ nature of the series. In sum, these extensive panel unit root tests provide consistent evidence that the variables under consideration, except for heating degree days, are non-stationary in levels but stationary in first differences. In the next section, we investigate the existence of cointegration among the variables.

Table 2.4: Panel unit root test results.

	Levels		First Differences		Levels		First Differences	
	No trend	Trend	No trend	Trend	No trend	Trend	No trend	Trend
lnRDGDpc					lnNONRENEWpc			
Levin, Lin, and Chu	-9.72***	-4.27***	-6.31***	-3.69***	1.06	-4.32***	-13.18***	-9.71***
Breitung	7.18	6.51	-10.21***	-1.00	3.74	4.23	-17.66***	-2.72**
IPS	-5.41***	-2.64**	-11.57***	-13.12***	5.53	-6.76***	-16.15***	-16.51***
Fisher-ADF	103.96	63.57	166.94***	122.41*	39.54	85.54	211.68***	128.19**
Fisher-PP	248.30***	85.01	447.83***	429.94***	46.00	137.02**	932.33***	788.80***
Hadri [!]	82.66***	41.38***	6.83***	4.40***	79.89***	24.89***	-1.43	-2.38
lnRGDP2pc					lnRENEWpc			
Levin, Lin, and Chu	-10.00***	-4.32***	-6.59***	-4.36***	-1.14	-4.07***	-12.78***	-10.89***
Breitung	10.30	7.00	-9.59***	-2.22**	3.21	-1.33*	-14.02***	-13.76***
IPS	-5.83***	-2.16**	-11.30***	-13.17***	1.93	-7.30***	-16.36***	-16.70***
Fisher-ADF	109.20	60.38	159.95***	133.36**	113.27	354.97***	431.41***	251.74***
Fisher-PP	299.81***	84.33	434.52***	436.67***	117.39	167.43***	1066.57***	882.03***
Hadri [!]	88.10***	43.96***	8.50***	5.58***	87.75***	32.38***	0.24	4.18***
lnRESIDpc					lnCOMMpc			
Levin, Lin, and Chu	-1.63*	-7.44***	-20.42***	-16.87***	1.81	-4.29***	-9.33***	-8.76***
Breitung	-6.90***	-4.91***	-17.58***	-13.47***	-1.16	5.74	-14.53***	-7.31***
IPS	-3.18***	-9.91***	-20.26***	-20.44***	1.29	-5.42***	-16.88***	-18.24***
Fisher-ADF	43.86	71.15	326.95***	353.40***	122.25*	120.55*	225.63***	208.06***
Fisher-PP	137.75**	213.16***	1925.63***	1900.18***	125.64*	133.52**	1050.24***	1072.57***
Hadri [!]	50.52***	25.60***	-4.19	-5.82	53.07***	34.18***	0.49	-3.63
lnINDUSTpc					lnTRANSPpc			
Levin, Lin, and Chu	-4.34***	-4.89***	-14.91***	-11.75***	3.57	-1.06	-1.79*	2.52
Breitung	5.65	-2.48**	-16.57***	-13.35***	-1.69*	9.76	-14.21***	5.38
IPS	0.92	-5.92***	-17.81***	-18.12***	3.28	-4.91***	-13.59***	-14.28***
Fisher-ADF	102.95	134.30**	237.50***	156.63***	40.60	71.38	149.32***	90.32
Fisher-PP	86.80	110.09	1174.97***	1012.50***	75.83	95.17	722.72***	603.88***
Hadri [!]	79.92***	31.92***	-2.94	-1.88	57.08***	23.39***	1.51*	-0.24
lnCO2pc					HDD			
Levin, Lin, and Chu	7.13	-2.88**	-10.97***	-9.29***	-18.97***	-16.10***	-23.30***	-18.39***
Breitung	8.57	5.57	-15.82***	-3.34**	-14.32***	-13.66***	-13.03***	-9.31***
IPS	11.82	-5.54***	-16.64***	-17.42***	-15.65***	-16.49***	-21.00***	-20.71***
Fisher-ADF	16.67	47.33	151.47***	128.46**	246.41***	204.55***	800.67***	658.07***
Fisher-PP	42.41	112.88	1024.51***	963.79***	774.22***	615.06***	2674.33***	2334.11***
Hadri [!]	88.21***	32.46***	0.58	-2.88	-0.72	-1.53	-6.44	-6.51
lnPRC								
Levin, Lin, and Chu	-10.14***	1.63	-11.92***	-16.29***				
Breitung	1.13	5.63	-15.58***	-11.32***				
IPS	-1.89**	6.29	-14.61***	-16.96***				
Fisher-ADF	97.59	2.46	92.16	173.29***				
Fisher-PP	85.08	2.52	631.05***	740.19***				
Hadri [!]	71.70***	66.91***	7.16***	-0.29				

Notes: ***, **, and * represent significance at 1%, 5%, and 10% respectively. (!) Null hypothesis, the series is stationary. Akaike Information Criterion (AIC) is used for the choice of the lag length. The highest allowed number of lags is restricted to three.

2.5.2 Panel Cointegration Tests

The goal of cointegration tests is to assess whether the variables in a system exhibit a long-run equilibrium relationship. Table 2.5 shows the cointegration results for the variables in the base model in column (1) and those for the augmented model in column (2) across three specifications. The base model includes all the control variables used in addition to total nonrenewable energy use and renewable energy use, whereas the augmented model includes all the variables in the base model, but with total nonrenewable energy use broken down into the four different sectors (residential, commercial, industrial, and transportation).

The Kao test (1999), is an extension of the Dickey-Fuller test for cointegration in a panel data framework. It tests the null hypothesis of no cointegration by considering the residuals from a panel data regression, assuming homogeneity among the cross-sections and not allowing for individual effects to vary across panel units. This test involves estimating a common autoregressive parameter across different cross-sections, applying a residual-based approach to test for unit roots in the residuals of the estimated long-run relationships. The results in Table 2.5 show that the modified Dickey-Fuller t-statistic and the Dickey-Fuller t-statistic for both model specifications are statistically significant (p-values < 0.01), suggesting strong evidence against the null hypothesis of no cointegration among the panel units and therefore, implying a long-run equilibrium relationship between the variables considered.

The Pedroni test (1999), extends the Kao test by allowing for heterogeneity across different units in the panel. This test is particularly flexible in accounting for different dynamics and fixed effects across individual units, thus providing a more robust framework for testing cointegration in the presence of heterogeneity. The results from the Pedroni test, such as the modified Phillips-Perron t-statistic and the Phillips-Perron t-statistic, are also significant at the 1% level. This strengthens the evidence of a cointegrating relationship. The negative values in the Phillips-Perron t-statistic, which

Table 2.5: Cointegration Test Results.

Test Name	(1)	(2)
Kao Test		
Modified Dickey–Fuller t	3.245*** (0.001)	3.886*** (0.000)
Dickey–Fuller t	4.078*** (0.000)	4.984*** (0.000)
Pedroni Test		
Modified Phillips–Perron t	5.511*** (0.000)	4.175*** (0.000)
Phillips–Perron t	-7.695*** (0.000)	-7.087*** (0.000)
Westerlund Test		
Variance ratio	-2.532** (0.006)	-1.917* (0.028)

Notes: Column (1) introduces total nonrenewable and renewable energy categories, while column (2) includes all energy categories except for total nonrenewable energy use. The null hypotheses for the Kao and Pedroni tests is that there is no cointegration across the panels, with the Kao test assuming homogeneity and the Pedroni test allowing for panel-specific heterogeneity. The Westerlund test's null hypothesis is no cointegration among some or all panels, with a focus on error correction mechanisms. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors in parentheses.

are statistically significant, further confirm the presence of cointegration, suggesting that deviations from the long-run equilibrium are stationary.³

Westerlund (2005) proposed a test that emphasizes error correction mechanisms in panel data settings. This approach is distinct because it tests for the presence of a cointegrating relationship by examining the error correction term in the panel regression framework. The test is designed to detect the existence of cointegration without requiring pre-testing for unit roots, accommodating individual effects and cross-sectional dependence. The Westerlund test's variance ratio statistics, although showing a lower level of significance (p -values < 0.05 and < 0.1), still provide evidence supporting cointegration. The results suggest that short-run deviations from equilibrium are adjusted towards the long-run equilibrium path, indicating the presence of error-correction dynamics.

³The standard Phillips-Perron test tends to produce a negative statistic due to its unit root testing design, where large negative values suggest rejection of the null hypothesis. This sign difference, however, does not alter the interpretation of the results.

The significant results across all three tests provide robust evidence that a long-run equilibrium relationship exists among the variables across the panels, accounting for both homogeneous and heterogeneous effects. Despite their differences in approach and assumptions, all three tests consistently support the presence of cointegration in the panel data. This finding proves that the variables move together over time towards a common equilibrium state despite short-term fluctuations and provides us with a strong foundation for implementing panel vector error correction tests in the next section.

2.5.3 Panel VECM

Drawing from the existing literature on the relationship between CO_2 emissions, energy consumption, and economic growth (Pao et al., 2011a; Acaravci & Ozturk, 2010), the canonical model used to investigate the determinants of total CO_2 emissions can be expressed as:

$$\begin{aligned} \ln CO2pc_{s,t} = & \eta_1 \ln NONRENEWpc_{s,t} + \eta_2 \ln RENEWpc_{s,t} \\ & + \eta_3 HDD_{s,t} + \eta_4 \ln PRC_{s,t} + \eta_5 \ln RGDPpc_{s,t} \\ & + \eta_6 \ln RGDPpcSQ_{s,t} + \lambda_s + \theta_t + \epsilon_{s,t} \end{aligned} \quad (2.1)$$

where the state index, denoted by s , ranges from 1 to 50, while the time period, denoted by t , spans from 1 to 24. The dependent and explanatory variables are as described above. $\ln RGDPpcSQ$ is the squared form of $\ln RGDPpc$. State and year-specific fixed effects are denoted by λ_s and θ_t . The error term is $\epsilon_{s,t}$. It is important to note that this canonical model is a part of the VECM framework and is estimated together with other equations in the system.

The inclusion of both GDP and its squared term allows for the possibility of an inverted U-shaped relationship between economic growth and CO_2 emissions, as suggested by the EKC

hypothesis.⁴ By incorporating additional control variables such as heating degree day and energy prices, often ignored in most previous studies, our model aims to provide a more comprehensive analysis of the factors influencing CO_2 emissions across states and over time. In the second model, we break down the nonrenewable energy use variable into four different sectors: commercial, residential, industrial, and transportation. This extension allows us to analyze the specific impacts of each sector's energy use on total CO_2 emissions:

$$\begin{aligned}
\ln CO_2 pc_{s,t} = & \eta_1 \ln NONRENEW pc_{s,t} + \eta_2 \ln RESID pc_{s,t} + \eta_3 \ln COMM pc_{s,t} + \eta_4 \ln INDUST pc_{s,t} \\
& + \eta_5 \ln TRANSP pc_{s,t} + \eta_6 \ln RENEW pc_{s,t} + \eta_7 HDD_{s,t} + \eta_8 \ln PRC_{s,t} \\
& + \eta_9 \ln RGDP pc_{s,t} + \eta_{10} \ln RGDP pc SQ_{s,t} + \lambda_s + \theta_t + \epsilon_{s,t}
\end{aligned} \tag{2.2}$$

Even though Equations 2.1 and 2.2 are the main focus of our analysis, to capture the dynamic short-run and long-run equilibrium relationships among these variables, we employ the following panel VECM framework:

$$\begin{bmatrix} \Delta \ln CO_2 pc_{s,t} \\ \Delta \ln NONRENEW pc_{s,t} \\ \Delta \ln RENEW pc_{s,t} \\ \Delta HDD_{s,t} \\ \Delta \ln PRC_{s,t} \\ \Delta \ln RGDP pc_{s,t} \\ \Delta \ln RGDP pc SQ_{s,t} \end{bmatrix} = \sum_{j=1}^p \begin{bmatrix} \eta_{11j} & \eta_{12j} & \eta_{13j} & \eta_{14j} & \eta_{15j} & \eta_{16j} & \eta_{17j} \\ \eta_{21j} & \eta_{22j} & \eta_{23j} & \eta_{24j} & \eta_{25j} & \eta_{26j} & \eta_{27j} \\ \eta_{31j} & \eta_{32j} & \eta_{33j} & \eta_{34j} & \eta_{35j} & \eta_{36j} & \eta_{37j} \\ \eta_{41j} & \eta_{42j} & \eta_{43j} & \eta_{44j} & \eta_{45j} & \eta_{46j} & \eta_{47j} \\ \eta_{51j} & \eta_{52j} & \eta_{53j} & \eta_{54j} & \eta_{55j} & \eta_{56j} & \eta_{57j} \\ \eta_{61j} & \eta_{62j} & \eta_{63j} & \eta_{64j} & \eta_{65j} & \eta_{66j} & \eta_{67j} \\ \eta_{71j} & \eta_{72j} & \eta_{73j} & \eta_{74j} & \eta_{75j} & \eta_{76j} & \eta_{77j} \end{bmatrix} \begin{bmatrix} \Delta \ln CO_2 pc_{s,t-j} \\ \Delta \ln NONRENEW pc_{s,t-j} \\ \Delta \ln RENEW pc_{s,t-j} \\ \Delta HDD_{s,t-j} \\ \Delta \ln PRC_{s,t-j} \\ \Delta \ln RGDP pc_{s,t-j} \\ \Delta \ln RGDP pc SQ_{s,t-j} \end{bmatrix} + \begin{bmatrix} \eta_{81} \\ \eta_{82} \\ \eta_{83} \\ \eta_{84} \\ \eta_{85} \\ \eta_{86} \\ \eta_{87} \end{bmatrix} ECT_{s,t-1} + \begin{bmatrix} \lambda_s \\ \lambda_s \\ \lambda_s \\ \lambda_s \\ \lambda_s \\ \lambda_s \\ \lambda_s \end{bmatrix} + \begin{bmatrix} \theta_t \\ \theta_t \\ \theta_t \\ \theta_t \\ \theta_t \\ \theta_t \\ \theta_t \end{bmatrix} + \begin{bmatrix} \epsilon_{1s,t} \\ \epsilon_{2s,t} \\ \epsilon_{3s,t} \\ \epsilon_{4s,t} \\ \epsilon_{5s,t} \\ \epsilon_{6s,t} \\ \epsilon_{7s,t} \end{bmatrix} \tag{2.3}$$

The vector on the left-hand side contains the first-differenced dependent variables. The first vector on the right-hand side represents the state-specific intercepts for each equation in the system. The second component on the right-hand side is a matrix that represents the short-run adjustment dynamics, capturing the impact of lagged changes in each variable on the current changes in the dependent variables. The optimal lag length, p , is determined by the Schwarz information criteria

⁴In this case, The RGDP variable is expected to have a positive coefficient, and its squared version is expected to have a negative coefficient.

(SIC). Each element in the coefficient matrix, denoted as η_{ij} , represents the short-run coefficient that measures the impact of the j -th lagged change in one variable on the current change in another variable. For example, η_{12} represents the short-run impact of a change in energy use two periods ago on the current change in total CO_2 emissions, while η_{21} represents the short-run impact of a change in total CO_2 emissions in the previous period on the current change in energy use.

The 1×7 vector, ECT, contains the error correction terms and the associated parameter vector captures the speed of adjustment of each variable to deviations from the long-run equilibrium relationship. The last vector represents the idiosyncratic error terms for each equation in the system, capturing the unexplained variations or shocks specific to each equation and state s at time t . Equation 2.2 can be represented similarly in a VECM framework which is not included for brevity.

Cointegration Rank Identification

To determine the number of cointegrating relations among the variables, we employ the Johansen cointegration test (Johansen & Juselius, 1990), which is based on the trace and maximum eigenvalue statistics. The results of the unrestricted cointegration rank tests are presented in Table 2.6.

The trace test evaluates the null hypothesis that there are at most r cointegrating relations against the alternative hypothesis that there are more than r cointegrating relations, where r is a value between 0 and $k - 1$, and k is the number of endogenous variables in the system. The test is performed sequentially, starting with $r = 0$ and increasing r by one until the null hypothesis cannot be rejected. The trace test statistic is computed as $LR_{tr}(r|k) = -T \sum_{i=r+1}^k \log(1 - \hat{\lambda}_i)$, where T is the sample size, $\hat{\lambda}_i$ is the i -th largest estimated eigenvalue of the Π matrix, k is the number of endogenous variables, and r is the hypothesized number of cointegrating relations. The Π matrix is the matrix of long-run coefficients in the VECM representation of the VAR model. It can be decomposed as $\Pi = \alpha\beta'$, where α represents the speed of adjustment to equilibrium, and β is the matrix of cointegrating vectors. The trace test statistic follows a non-standard distribution, and the critical values are obtained through simulation. If the test statistic exceeds the critical value at

a given significance level, the null hypothesis is rejected, indicating the presence of more than r cointegrating relations. The trace test results in Table 2.6 indicate the presence of three cointegrating equations at the 5% significance level.

The maximum eigenvalue test assesses the null hypothesis of r cointegrating relations against the alternative of $r + 1$ cointegrating relations. The test statistic is given by $LR_{max}(r|r + 1) = -T \log(1 - \hat{\lambda}_{r+1})$, where T is the sample size and $\hat{\lambda}_{r+1}$ is the $(r+1)$ -th largest estimated eigenvalue of the Π matrix. Similar to the trace test, the maximum eigenvalue test is conducted sequentially, starting with $r = 0$ and increasing r until the null hypothesis cannot be rejected. The test statistic follows a non-standard distribution, and the critical values are obtained through simulation. If the test statistic exceeds the critical value at a given significance level, the null hypothesis is rejected, indicating the presence of $r + 1$ cointegrating relations. The maximum eigenvalue test differs from the trace test in terms of the alternative hypothesis. While the trace test evaluates the null hypothesis against the alternative of more than r cointegrating relations, the maximum eigenvalue test assesses the null hypothesis against the specific alternative of $r + 1$ cointegrating relations.

Table 2.6: Unrestricted Cointegration Rank Tests for the Base Model.

Hypothesized No. of CE(s)	Eigenvalue	Trace Test Statistic	Max-Eigen Test Statistic
None	0.230	613.060***	287.257***
At most 1	0.184	325.803***	223.206***
At most 2	0.056	102.597***	63.040***
At most 3	0.029	39.556*	32.915***
At most 4	0.004	6.641	4.414
At most 5	0.002	2.227	1.796
At most 6	0.000	0.431	0.431

Notes: Probabilities are MacKinnon-Haug-Michelis (1999) p-values. Significance levels are indicated as *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Considering the economic theory and the interpretability of the cointegrating equations, as well as the strongest significance level (1%), we proceed with the assumption of three cointegrating relations among the variables in our model (Table 2.6). We repeat the same analysis for the model

proposed in Equation 2.2 in Table 2.7. Similarly, we find the existence of three cointegrating relationships among the variables at the 1% significance level.

Table 2.7: Unrestricted Cointegration Rank Tests for the Augmented Model.

Hypothesized No. of CE(s)	Eigenvalue	Trace Test Statistic	Max-Eigen Test Statistic
None	0.268	651.184***	343.174***
At most 1	0.090	308.010***	103.966***
At most 2	0.073	204.044***	83.869***
At most 3	0.039	120.175*	43.849*
At most 4	0.030	76.325	33.337
At most 5	0.020	42.989	22.701
At most 6	0.012	20.287	13.814
At most 7	0.004	6.474	4.050
At most 8	0.002	2.424	1.737
At most 9	0.001	0.687	0.687

Notes: Probabilities are MacKinnon-Haug-Michelis (1999) p-values. Significance levels are indicated as *** p<0.01, ** p<0.05, * p<0.1.

Granger Causality Tests

From this section onward, we only report the results of the augmented model (Equation 2.2) in the main paper. Readers may refer to the appendix section to find all the counterpart results for the base model (Equation 2.1).

We use Granger causality tests in a panel VECM context to estimate the short-run causality paths (Pesaran et al., 1999). Specifically, our Granger causality results are based on a post-estimation approach from the panel VECM results, which is more suitable in this case given the presence of cointegration. In contrast to traditional Granger causality tests performed in a VAR framework,⁵ the VECM-based Granger causality test accounts for both short-term dynamics and long-term equilibrium relationships through the error correction term. This allows us to explore short-run predictive causality while respecting the long-run cointegrating relationships.

⁵In a VAR model, Granger causality is typically assessed by testing whether past values of a variable improve the prediction of another variable without accounting for any long-term equilibrium relationship.

Table 2.8 presents the results of short-run Granger causality tests based on the panel VECM for the extended model. Each row tests the null hypothesis that the variable in the row does not Granger cause the variable in the column. Rejection of the null implies evidence of Granger causality. Figure 2.2 visualizes the causal relationships.

Table 2.8: Short-run Granger Causality/Block Exogeneity Wald Tests Summary for the Augmented Model.

From	To									
	$\Delta \ln CO_2 pc$	$\Delta \ln RESID pc$	$\Delta \ln COMM pc$	$\Delta \ln INDUST pc$	$\Delta \ln TRANSP pc$	$\Delta \ln RENEW pc$	ΔHDD	$\Delta \ln PRC$	$\Delta \ln RGDP pc$	$\Delta \ln RGDP^2 pc$
$\Delta \ln CO_2 pc$	-	4.426**	0.030	2.701	0.382	0.844	2.475	0.215	0.065	0.125
$\Delta \ln RESID pc$	1.023	-	15.324***	5.172**	2.373	2.584	6.349**	0.370	8.309***	7.316***
$\Delta \ln COMM pc$	0.017	0.162	-	12.194***	0.580	4.283**	2.256	0.555	1.514	1.123
$\Delta \ln INDUST pc$	1.129	1.473	1.338	-	0.383	0.089	4.185**	8.632***	0.324	0.353
$\Delta \ln TRANSP pc$	7.848***	6.936***	11.516***	1.304	-	1.578	3.633*	38.040***	21.436***	24.178***
$\Delta \ln RENEW pc$	4.292**	0.692	0.086	3.523*	2.570	-	0.446	20.197***	1.588	1.836
ΔHDD	2.957*	6.263**	1.777	8.286***	5.632**	2.342	-	0.648	8.658***	7.828***
$\Delta \ln PRC$	0.856	1.607	0.161	20.971***	0.717	8.853***	2.551	-	0.303	0.433
$\Delta \ln RGDP pc$	0.512	0.586	0.852	1.665	0.197	0.333	0.999	10.294**	-	23.122***
$\Delta \ln RGDP^2 pc$	0.500	0.856	0.607	1.691	0.151	0.164	0.518	7.912**	2.843*	-

Notes: The values in the table report Chi-square statistics from VEC Granger causality/block exogeneity tests with one degree of freedom for each analysis. These Chi-square statistics are from Wald-style exclusion tests on the lagged difference terms. Significance levels are marked as $p < 0.1$, $p < 0.05$, and $p < 0.01$.

The short-run Granger causality tests based on the panel VECM reveal intricate relationships among sectoral energy use, renewable adoption, emissions, prices, weather, and economic growth. We find that changes in residential energy use Granger cause changes in commercial and industrial energy use, heating degree days, GDP, and GDP squared, suggesting that residential energy consumption plays a key role in influencing other energy sectors, weather-related energy demand, and economic activity. Similarly, transportation energy use changes strongly Granger cause changes in CO_2 emissions, residential and commercial energy use, energy prices, GDP, and GDP squared, highlighting the central role of the transportation sector in shaping emissions, energy dynamics, prices, and economic growth. These findings are consistent with the work of Hamit-Hagggar (2012), who found unidirectional short-run causality from energy consumption and economic growth to greenhouse gas emissions in Canadian industrial sectors.

Furthermore, our results indicate that changes in renewable energy use Granger cause changes in CO_2 emissions, industrial energy use, and energy prices. This finding aligns with

the work of Salari et al. (2021), who showed that renewable energy consumption has a negative relationship with CO_2 emissions in the United States at the state level, and the work of Marrero (2010), who found that shifting the energy mix from coal to renewable and nuclear sources has the largest impact on reducing emissions in European Union countries.

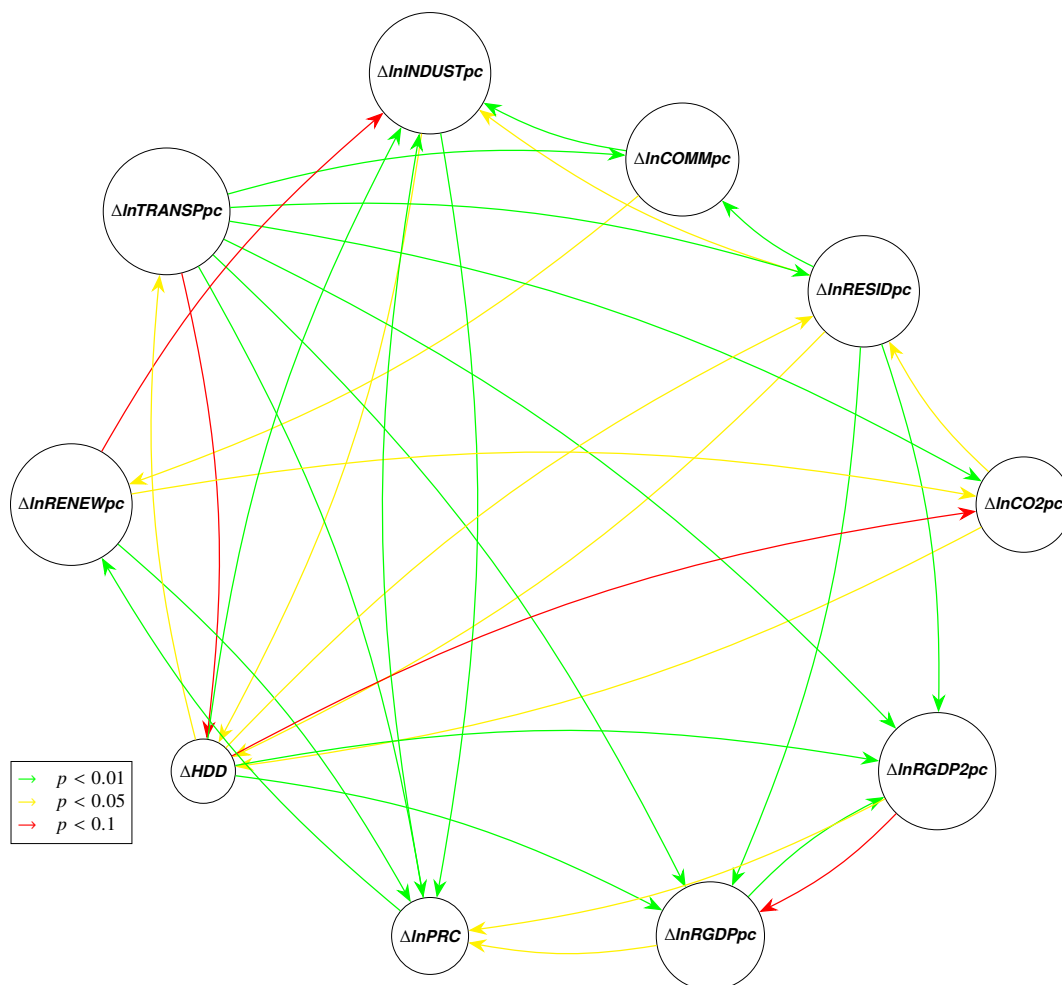


Figure 2.2: Causal relationships based on VEC Granger causality tests for the augmented model.

Notes: The arrows represent the direction of causality, and the colors indicate the level of statistical significance: green for $p < 0.01$, yellow for $p < 0.05$, and red for $p < 0.1$.

Notably, we find that GDP growth Granger causes changes in energy prices but does not directly drive changes in sectoral energy use, emissions, or renewable adoption in the short run, suggesting that economic growth influences energy prices but is more of an outcome than a direct driver of changes in energy use and emissions at the sectoral level. This result contrasts with the

findings of Wang et al. (2011), who found bidirectional short-run causality between CO_2 emissions and energy consumption, as well as between energy consumption and economic growth, in China, and with the results of Pao & Tsai (2010), who found bidirectional causality between energy and emissions, and unidirectional causality from emissions and energy to output in BRIC countries.

Generalized Impulse Response Functions

Generalized impulse response functions (GIRFs) examine the dynamic responses of variables in a system to shocks in other variables. Unlike orthogonalized impulse responses, GIRFs do not depend on the ordering of variables in the VAR model (Pesaran & Shin, 1998). Figure 2.3 presents the GIRFs for the augmented model, tracing out the response paths of each variable to one standard deviation shock in the other variables over a 10-year horizon.

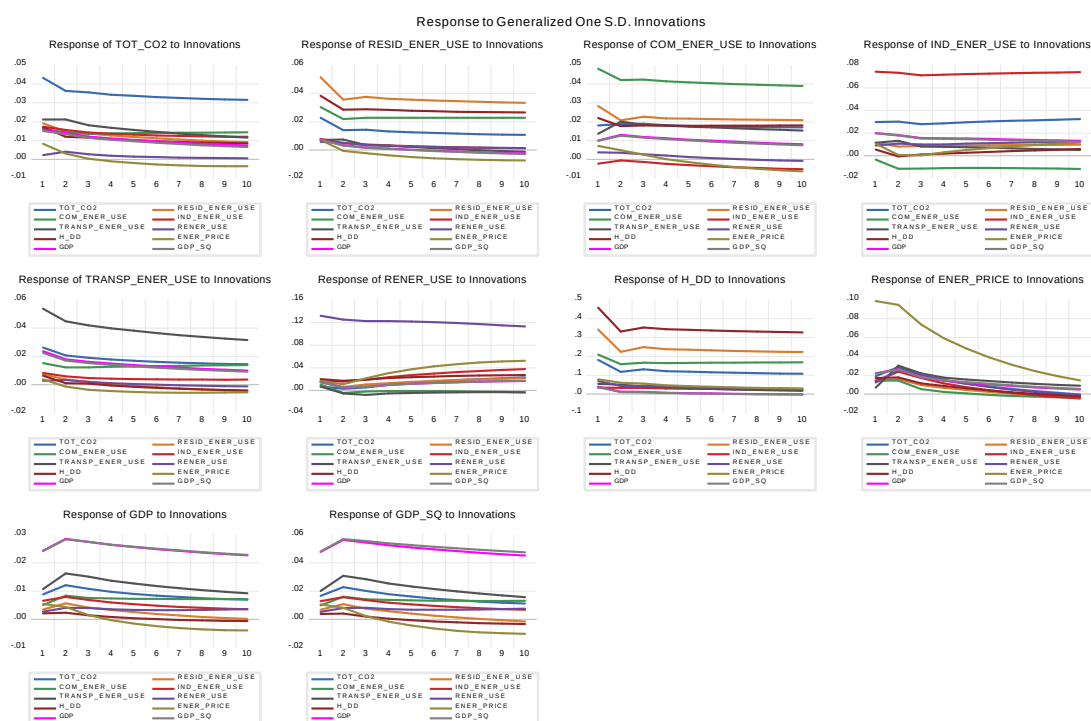


Figure 2.3: Generalized Impulse Response Functions for the Augmented Model.

The impulse response analysis reveals several critical relationships in the energy-emissions-GDP system. Most notably, CO_2 emissions show the strongest response to transportation and industrial energy use shocks, which again highlights these sectors' outsized influence on emission patterns. Renewable energy shocks generate only modest negative responses in emissions. This suggests limited immediate potential for emissions reduction through renewable adoption alone. In terms of sectoral dynamics, transportation energy use emerges as particularly influential and generates significant positive responses across residential and commercial energy consumption while also strongly affecting GDP. This emphasizes the transportation sector's central role in both energy systems and economic activity. Weather shocks, measured by heating degree days, trigger substantial responses in residential and commercial energy use as the climate tends to be sensitive to these sectors.

The economic growth responses reveal important policy implications—while GDP shows persistent positive responses to transportation and industrial energy shocks, it demonstrates relatively weak responses to renewable energy adoption. Energy price shocks generally induce responses that shift from positive to negative over time across multiple variables, which suggests that price mechanisms may have complex dynamic effects that evolve differently in the short versus long run.

2.5.4 Fully Modified Ordinary Least Squares (FMOLS)

The FMOLS is a cointegration technique developed by Pedroni (2001) for estimating long-run relationships in heterogeneous panels. The key advantage of this approach is that it allows researchers to pool the long-run information across panel units while permitting short-run dynamics and fixed effects to vary across these panel units. This is particularly useful when working with macroeconomic panels, where, in our case, states may exhibit similar long-run behavior but differ in their short-term responses and characteristics. FMOLS produces asymptotically unbiased estimators and standard normal distributions, enabling inference on common long-run relationships that are robust to short-run heterogeneity.

Table 2.9 presents the FMOLS estimates for the long-run determinants of total CO_2 emissions. The results indicate that energy use in all sectors considered has a significant positive effect on emissions. A 1% increase in energy use in the residential, commercial, industrial, and transportation sectors is associated with a 0.138%, 0.105%, 0.249%, and 0.328% increase in emissions, respectively, holding other factors constant. The transportation sector appears to have the largest impact on emissions, followed by the industrial sector. This finding is consistent with the work of Hamit-Haggar (2012), who found a strong positive long-run relationship between energy consumption and emissions for Canadian industries in the aggregate, and the work of Marrero (2010), who found that the industry and transport sectors have reduced emissions in the most developed EU countries, likely due to efficiency improvements.

To put our findings into perspective, using the average state-by-year population of 6.07 million residents over the study period and the average state-by-year emissions of 138.1 million metric tons (MMT), the per capita emissions amount to approximately 22.75 metric tons per person per year.⁶ A 1% increase in emissions per capita per state per year would represent an additional 0.2275 metric tons of CO_2 per person per year, totaling approximately 1.382 MMT of CO_2 emissions annually on average per state.⁷ This means a 1% increase in industrial energy use per capita, given the estimated elasticity of 0.249, would lead to a 0.249% increase in emissions per capita. This corresponds to an increase of approximately 0.0567 metric tons of CO_2 per person per year, totaling about 0.344 MMT of additional emissions annually on average per state.⁸ Similarly, a 1% increase in transportation energy use per capita, with an estimated elasticity of 0.328, would result in a 0.328% increase in emissions per capita. This equates to an increase of approximately 0.0746 metric tons of CO_2 per person per year, totaling about 0.453 MMT of additional emissions annually on average per state.⁹

⁶Calculated as $\frac{138.1 \text{ MMT}}{6.07 \text{ million people}} \approx 22.75$ metric tons per person per year.

⁷Calculated as $0.01 \times 22.75 \text{ metric tons per person} \times 6.07 \text{ million people} = 1.382 \text{ MMT}$.

⁸Calculated as $0.00249 \times 22.75 \text{ metric tons per person} \times 6.07 \text{ million people} = 0.344 \text{ MMT}$.

⁹Calculated as $0.00328 \times 22.75 \text{ metric tons per person} \times 6.07 \text{ million people} = 0.453 \text{ MMT}$.

Table 2.9: FMOLS Results for augmented model.

Variable	Estimate
lnRESIDpc	0.1375** (0.0544)
lnCOMMpc	0.1051** (0.0419)
lnINDUSTpc	0.2494*** (0.0171)
lnTRANSPpc	0.3284*** (0.0351)
lnRENEWpc	-0.0456*** (0.0078)
lnPRC	-0.0157 (0.0127)
HDD	0.0280*** (0.0080)
lnRGDPpc	0.8213*** (0.1429)
lnRGDP2pc	-0.4502*** (0.0621)

Notes: The table presents the estimates and standard errors for the variables impacting $\ln CO2pc$, derived from panel FMOLS regression. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

In contrast, the use of renewable energy has a significant negative effect, implying that a transition to renewables can help mitigate emissions in the long run. A 1% increase in renewable share is associated with a 0.046% decrease in emissions. The effect of energy prices is negative but not statistically significant, suggesting that price changes may have limited long-run effects on emissions after accounting for sectoral energy use and renewable share. Heating degree days have a small but significant positive effect on emissions. This indicates that colder temperatures, requiring more heating, tend to increase emissions in the long run.

Importantly, the results confirm the presence of an inverted U-shaped relationship between GDP per capita and emissions. The positive coefficient on $\ln RGDPpc$ (0.821) and negative coefficient on $\ln RGPD2pc$ (-0.450) imply that emissions initially rise with economic growth but eventually peak and decline at higher income levels, consistent with the EKC hypothesis. This finding suggests that emissions initially rise with economic growth but eventually peak and decline at higher income levels. This result aligns with the work of Apergis & Payne (2010), who found evidence of an inverted U-shaped relationship between emissions and GDP for a panel of Commonwealth of Independent States countries, and Pao & Tsai (2010), who found a similar relationship for BRIC countries. However, it contrasts with the findings of Marrero (2010), who rejected the EKC hypothesis for a panel of European Union countries after controlling for energy factors and convergence, and Wang et al. (2011), who found a U-shaped relationship between economic growth and CO_2 emissions in China.

Comparing our base and augmented models reveals the importance of disaggregating energy consumption into sectoral components. While the base model, which includes total nonrenewable and renewable energy use, provides a broad understanding of the relationships among the selected variables, the augmented model offers a more detailed picture by capturing the heterogeneous impacts of different energy sectors. The FMOLS estimates for the base model (Table A.2) show that total energy use has a positive and significant effect on emissions, while renewable energy use has a negative and significant effect, consistent with the augmented model results. However, the

augmented model allows us to identify the relative magnitudes of the impacts of different energy sectors on emissions, with the transportation and industrial sectors having the largest effects. This sectoral disaggregation provides valuable insights for policymakers in designing targeted emissions reduction strategies, as emphasized by Hamit-Haggar (2012) and Marrero (2010).

2.6 Conclusions

This study investigates the dynamic causal relationships between sectoral energy consumption, renewable energy use, economic growth, and CO_2 emissions in the United States from 1997 to 2020. By employing VECM and FMOLS estimation, we uncover both short-run and long-run linkages among these variables. Our approach extends the existing literature by disaggregating energy consumption into residential, commercial, industrial, and transportation sectors, providing a more nuanced understanding of the emissions-energy-economy nexus.

Our findings have significant policy implications for practitioners and policymakers in the United States. First, the long-run FMOLS estimates reveal that the transportation sector has the largest impact on emissions, followed by the industrial sector. This is further supported by the impulse response analysis, which shows that CO_2 emissions respond most strongly to shocks in transportation and industrial energy use. Policymakers could prioritize emissions reduction strategies in these sectors, such as promoting electric vehicles, improving fuel efficiency standards, and encouraging the adoption of cleaner technologies in industry. Investing in public transportation infrastructure and incentivizing using low-carbon transportation modes can also help curb emissions growth in the long run.

Second, our results highlight the importance of renewable energy in mitigating emissions. The FMOLS estimates show that an increase in the share of renewable energy leads to a significant decrease in emissions. The impulse response functions also indicate that renewable energy shocks result in a slight negative response in CO_2 emissions, suggesting the potential for emissions reduction. Policymakers could continue to support deploying renewable energy technologies

through tax incentives, subsidies, and regulations. Encouraging the integration of renewables into the grid and removing barriers to their adoption can help accelerate the transition to a low-carbon energy system.

Third, the short-run Granger causality tests indicate that the residential sector is key in influencing other energy sectors, weather-related energy demand, and economic activity. The impulse response analysis further reveals that residential energy use responds strongly to its own shocks, heating degree day shocks, and commercial energy use shocks. This suggests that policies targeting energy efficiency improvements in the residential sector, such as building codes, appliance standards, and weatherization programs, can have spillover effects on other sectors and contribute to emissions reductions.

Fourth, the evidence for the EKC hypothesis in our long-run estimates implies that economic growth can eventually lead to decreased emissions, but this turning point may occur at higher income levels. The impulse response functions show that economic growth responds positively to shocks in transportation, industrial, commercial, and residential energy use, as well as emissions. Policymakers may focus on decoupling economic growth from emissions by promoting sustainable development practices, such as green infrastructure investment, resource efficiency, and adopting circular economy principles.

Finally, our study emphasizes the importance of considering the heterogeneous impacts of different energy sectors on emissions and economic growth when designing climate policies. While the transportation and industrial sectors are the largest contributors to emissions, the short-run Granger causality tests and impulse response analysis reveal that the residential and renewable energy sectors also play significant roles in shaping emissions and economic dynamics. Policymakers could consider adopting a comprehensive and balanced approach that targets multiple sectors simultaneously, while considering their specific characteristics and potential trade-offs.

CHAPTER 3

MEASURING THE IMPACT OF NET METERING POLICY ON RESIDENTIAL GHG EMISSIONS IN THE U.S.: A DIFFERENCE-IN-DIFFERENCE ANALYSIS¹

¹J. Augustin, B. Karali, S. Ferreira & M. Filipski. To be submitted to the American Economic Review.

3.1 Abstract

This study investigates the effects of net metering policies on residential greenhouse gas emissions across U.S. states from 1990 to 2020. Leveraging the staggered implementation of these policies, we employ the Callaway and Sant’Anna (2021) estimator to address potential biases in traditional two-way fixed effects models with heterogeneous treatment effects. Our findings reveal that net metering policies lead to modest but statistically significant reductions in residential GHG emissions, with effects intensifying over time to reach a 0.6% decrease five years post-adoption compared to the adoption year baseline. These results hold even when accounting for policy spillovers to neighboring states. We observe substantial heterogeneity in policy impacts across various dimensions. Contrary to expectations, states with less favorable policies (i.e., those with a below-average score) —from electricity consumers’ perspectives— show significant emissions reductions, while those with more favorable policies exhibit insignificant effects. We argue that this is likely due to each trait having a different weight in different states. Analysis of specific policy traits indicates that generous compensation schemes, higher system size caps, and customer ownership of renewable energy credits drive larger emissions reductions. Republican-leaning states demonstrate more consistent emissions reductions compared to Democratic-leaning states. Our examination of cumulative effects reveals persistent, albeit small, reductions in emissions across years of policy implementation. Exploration of underlying mechanisms indicates that increased residential solar PV adoption and enhanced grid interaction drive these reductions.

3.2 Introduction

Recent extreme weather events, from devastating wildfires in California (Legislative Analyst Office, 2022) to hurricanes Helene and Milton, provide alarming evidence of intensifying climate impacts. Decarbonizing the power sector, which contributes almost a quarter of economy-wide emissions, is central to achieving a significant overall reduction in emissions. Renewables like solar photovoltaic distributed generation present viable technological pathways to mitigate residential carbon emissions while ensuring affordable and equitable clean energy access (Chang et al., 2022). But realizing their emissions reduction potential requires optimized policy mechanisms and market frameworks, aspects which remain understudied.

Distributed solar photovoltaics have surged in adoption over recent years as an avenue for households to advance sustainability objectives while benefiting through self-generation. Rooftop solar also presents a technological pathway for unlocking demand flexibility and grid services through smart integration with storage and load controls. By 2022, cumulative small-scale solar deployment in the United States reached over 3 million installations with 22 GW capacity, up from 405 MW in 2010, demonstrating substantial progress (REN21 Secretariat, 2020). Yet realizing the emissions reduction and grid modernization potential requires policy mechanisms optimizing integration and returns. With high upfront capital costs (Feldman et al., 2015), net metering policy financial feasibility relies on grant subsidies and revenue uplifts from interstate energy trade, which depends on net metering credit and compensation structures specified by state programs. Additionally, the emissions offset also follows directly from the solar energy displacing marginal fossil generation. Therefore, state net metering policy design variations can constrain emission mitigation by impacting solar adoption rates and self-consumption incentives.

First adopted in the United States in the late 1970s, net metering enables distributed solar owners to receive credit for excess generation fed into the grid, typically valued at the retail volumetric electricity rate (\$/kWh) (Smith et al., 2021). System owners offset their net consumption

from the grid each billing cycle, incentivizing self-supply for resilience and bill management (Poullikkas et al., 2013). By facilitating renewable energy self-consumption and sale at prevailing rates (Schelly et al., 2017), net metering serves as a crucial incentive for adopting renewable energy. Despite its preeminence within state policy portfolios, the interlinkages between net metering policy design, solar adoption and emissions remain empirically underexplored, especially amidst ongoing state-level policy updates. The purpose of this study is to evaluate the causal impact of net metering policy adoption across U.S. states on residential CO_2 emissions.

Quantitative evaluations of distributed solar and net metering efficacy have followed several strands. Optimization studies employ cost-benefit analysis and multi-objective programming to ascertain system configurations and policy structures minimizing expenses or emissions, often in specific contexts like the Netherlands or Canada (Delgado et al., 2018; Hashemi et al., 2023). Such work provides techno-economic guidance but lacks generalizability. Recent empirical assessments have leveraged state-level solar capacity and electricity data to study impacts of net metering policy design on net metering policy outcomes such as electricity sold back to the grid, but limitations around identification assumptions persist (Smith et al., 2021). For instance, unobserved factors like pro-environmental preferences simultaneously influence policy adoption and outcomes. Still no studies have established causality using quasi-experimental techniques measuring emissions changes attributable to net metering policy enactment. Furthermore, much scholarship focuses solely on solar adoption rather than emissions, or sidesteps net metering policy heterogeneity to consider binary adoption status. Addressing these gaps through national-level causal analysis while accounting for policy detail can offer replicable insights.

To the best of our knowledge, this study is the first one to analyze the impact of net metering on residential greenhouse gas (GHG) emissions in the United States. It contributes to the interdisciplinary scholarship and policy discourse on distributed solar support mechanisms by evaluating the causal impact of net metering policy adoption across U.S. states on residential CO_2 emissions over 1990-2020. It also assembles a novel panel dataset tracking net metering

policy enactment dates, credit rates, system size caps, and program size caps across states over time, juxtaposed against state-level residential emissions. Leveraging temporal and cross-state policy variation, the analysis deploys a staggered difference-in-differences approach to isolate the emissions mitigation attributable to net metering policies. The research design compares emissions trajectories across early and late adopting states before and after policy implementation, netting out time-invariant differences.

Our analysis reveals several key insights into the impact of net metering policies on residential GHG emissions. We find that these policies contribute to a gradual but significant decrease in emissions, with the effect strengthening over a five-year time window (up to 0.6%). Our results indicate complex patterns of policy effectiveness across different state characteristics. We uncover that certain policy design elements, such as compensation rates, system size limits, and consumer ownership of renewable energy credits play crucial roles in the magnitude of emissions reductions. Republican-leaning states demonstrate more consistent emissions reductions compared to Democratic-leaning states, likely due to lower baseline of emissions. Moreover, we identify positive spillover effects, which suggest that the benefits of net metering policies extend beyond state borders, likely due to shared electrical transmission lines. Our examination of cumulative effects (i.e., stacked over time) and underlying mechanisms demonstrates that these policies drive emissions reductions primarily through increased adoption of residential solar PV systems and greater energy sold back to the grid.

The remainder of the paper is organized as follows: Section 3 provides a comprehensive overview of U.S. net metering policy. Section 4 presents the literature review, synthesizing existing research on net metering policies and their impacts. Section 5 outlines our econometric methods, including our difference-in-differences approach and empirical strategy. Section 6 describes our data sources, variable descriptions, and summary statistics. Section 7 presents the results of our analysis. Section 8 discusses the implications of our findings, contextualizing them within the

broader landscape of climate policy. Finally, Section 9 concludes by summarizing key insights and suggesting directions for future research.

3.3 U.S. Net Metering Policy Overview

Within the United States, state and local governments shoulder much of the regulatory authority over energy policy and electricity markets despite lacking an integrated federal strategy (Hashemi et al., 2023). Consequently, clean energy transitions have followed a fragmented approach, with Renewable Portfolio Standards in over half of states catalyzing utility-scale renewable procurement but variable action on distributed energy resources (Kramarz et al., 2021). Retail electricity rate-setting falls under the jurisdiction of public utility commissions, subject to emerging pressures from declining costs of distributed technologies, changing peak demand patterns and calls for equitable rate structures (Sioshansi & Pfaffenberger, 2006). This complex multi-layered governance poses challenges for coordinated policy action towards energy affordability, reliability, and sustainability goals. It also creates opportunities for states to experiment with innovative policy subcomponents that can be evaluated and potentially adopted across different states. Our study seizes upon such regulatory heterogeneity in net metering policy design among states to elicit new evidence on the levers for emissions mitigation.

Net metering is a billing mechanism that credits utility customers with rooftop solar or other distributed renewable generation for any excess electricity they feed into the grid that offsets their own electricity consumption (Rehman et al., 2020). It functions through a bidirectional meter that can spin forwards to record electricity drawn from the grid and spin backwards when solar or other customer-sited renewable generation is exporting power back into the grid (Poullikkas et al., 2013). Customers are thus only billed for their “net” energy usage from the utility based on the net directional flow recorded by the meter over the monthly billing period (Rehman et al., 2020). Net metering provides renewable system owners an economic return for on-site generation that reduces their utility-supplied electricity needs. It is a key incentive driving substantial growth of distributed

solar photovoltaics (PV) in particular (Smith et al., 2021). As of 2020, 38 states, Washington D.C. and Puerto Rico have implemented some form of net metering policy (Mitova & Kahsar, 2022). The federal Energy Policy Act of 2005 required all public utilities to make net metering available and for states to consider implementing net metering standards (Smith et al., 2021). While it spurred wider net metering policy adoption, states differed markedly in their eventual policies. As Smith et al. (2021) document in their review, state net metering programs vary across a number of key design elements like system size limits, aggregate enrollment caps, compensation rates, and renewable energy credit treatment. System size limits constrain the generation capacity eligible for net metering incentives on a per customer basis, ranging from just 10 kW in Utah to 5 MW in Massachusetts. Program caps limit total net metered distributed generation as a percentage of peak utility demand, from 0.5% in Oklahoma to 20% in Utah. Compensation for excess generation also varies widely from avoided cost to the full retail electricity rate (Smith et al., 2021).

Not only do policies differ greatly across states, Schelly et al. (2017) found that interconnection requirements and net metering incentives for distributed renewable generation are implemented inconsistently across utilities even within states. Clear, accurate information for customers on utility net metering programs was sometimes lacking. And consistent with Smith et al. (2021)'s findings on state policy variability, they determined compensation rates under utility net metering programs were frequently opaque or difficult to discern from publicly available policy documents. Their review reveals the convoluted policy implementation resulting from the absence of unified federal standards for regulating customer renewable energy interconnections. States and utilities opted to interpret broad federal net metering mandates differently based on local priorities and conditions (Schelly et al., 2017). Although the vast majority of states had established net metering standards as of 2015 based on federal law, 7 states were still lacking a statewide net metering policy (Schelly et al., 2017). Forsyth et al. (2002) similarly highlight wide variability even among early state adopters of net metering on factors like system capacity limits, aggregate capacity limits, ownership of renewable energy credits for generation, types of eligible customer generators, interconnection charges, and the

permitted length and terms of contracts. As Rehman et al. (2020) discuss, some states have adopted more advanced “smart” grid and metering infrastructure to enable time-of-use and dynamic pricing under net metering that better aligns compensation for excess distributed renewable generation with its time-varying value to the grid. However, the general nationwide policy trend they note is for traditional volumetric, non-time-differentiated net energy metering.

Rehman et al. (2020) emphasize additional complexities that cloud the net metering landscape in the U.S. For one, policies sometimes differ for customer-generators depending on whether their utility provider is privately or publicly owned. Requirements around interconnection processes and disputes as well as any charges levied on distributed generators for applications or upgrades can also vary (Rehman et al., 2020). On top of this, a major source of uncertainty in the future of net metering stems from controversial ongoing policy changes targeting the compensation incentives at the heart of net metering in many states. For instance, Mitova & Kahsar (2022) discuss a number of states that have recently moved to cap, reform, or eliminate retail rate compensation for excess distributed renewable generation. This trend towards reducing net metering incentives threatens the financial proposition behind further customer adoption of rooftop solar and other on-site generation. However, Mitova & Kahsar (2022) suggest time-of-use rate design reforms could motivate increased self-consumption behavior to help preserve solar value even absent the stronger economic subsidies traditionally provided under retail rate net metering. While federal law prompted a surge of net metering policy adoption across nearly all U.S. states, a range of factors have led to highly inconsistent implementation within and across states. The level of incentives provided under utility and state net metering standards depends on size limits, caps, ownership rights for renewable attributes, types of technologies eligible, connection terms, contract lengths, compensation rates, billing cycles, and more.² On top of inherently variable policy design, changes targeting key incentive parameters like retail rate compensation further undermine uniformity and stability of net metering’s economic signals. This complex and shifting landscape of net metering

²We provide a detailed explanation about the policy traits in the *Methodology* section.

policy in the United States poses significant modeling challenges but also useful heterogeneity to leverage in empirically evaluating the causal impacts of net metering design decisions on distributed renewable technology adoption and emissions outcomes.

3.4 Literature Review

Net metering policies have faced challenges from utility companies who view distributed generation as a threat to their traditional business model (Rehman et al., 2020; Szmolyan, 2020). There is ongoing debate about whether it leads to uneconomic bypass of utility-provided electricity (Ros & Sai, 2023). Ros & Sai (2023) empirically estimate an elastic demand for rooftop solar with respect to system costs, household income, retail electricity rates, and compensation rates under net metering. Their analysis finds that full retail rate compensation results in substantial uneconomic bypass, supporting reforms that better reflect avoided utility costs. Brown & Sappington (2017) develop an economic optimization model to characterize factors determining optimal distributed generation compensation under net metering or feed-in tariff policies.³ They demonstrate that efficiency impacts depend heavily on utility system parameters, while distributional effects vary based on interaction between policy-induced retail rate and compensation rate changes.

Studies outside the United States provide additional insights into net metering policy design. Iliopoulos et al. (2020) examine European Union policies, comparing implementation across Belgium, Italy, Cyprus and Greece through case studies that reveal variations in generator limits, grid requirements, and compensation mechanisms. Delgado et al. (2018) model prosumer buildings that generate excess electricity for conversion into heat, analyzing cost and emissions performance under different economic and climactic conditions in Netherlands and Finland. Hashemi et al. (2023) estimate that the marginal cost of GHG abatement through residential solar net metering is more than five times Canada's scheduled economy-wide carbon price for 2030, suggesting current retail

³Feed-in Tariff (FIT) policies are regulatory mechanisms that provide long-term contracts to renewable energy producers, typically offering a guaranteed price for the electricity they supply to the grid, often at rates higher than retail electricity prices to encourage renewable energy adoption.

rate structures yield costly and inefficient emissions reductions. However, Brown & Sappington (2017) demonstrate that factors like grid management and emissions externality costs can shift optimal compensation rates. More stringent decarbonization goals may require greater distributed renewable penetration beyond centralized capacity expansions (Trevino-Martinez et al., 2022), which necessitates careful attention to these design elements that shape net metering’s emissions and economic impacts.

3.5 Methodology and Empirical Strategy

3.5.1 Identification Strategy

To identify the causal effect of net metering policies on residential CO_2 emissions, we leverage the staggered adoption of these policies across U.S. states. Our empirical approach is based on a difference-in-differences (DiD) framework that compares changes in emissions between states that implemented net metering (treated states) and states that have not yet adopted such policies (control states), before and after policy adoption, while controlling for potential confounding factors.

The key identifying assumption is that, conditional on covariates and fixed effects, residential CO_2 emissions in treated states would have evolved in parallel to emissions in control states in the absence of net metering policies. In other words, we assume that the adoption of net metering is not correlated with other time-varying state-specific factors that also affect emissions. The plausibility of this *parallel trends assumption* is assessed through visual inspection of pre-treatment trends and event study specifications in subsequent sections.

3.5.2 Two-Way Fixed Effects (TWFE) Specifications

Static Version.

As a baseline empirical specification, we estimate a TWFE DiD model, which is commonly employed in policy evaluation studies with staggered treatment timing:

$$\log(CO2_{st}) = \beta D_{st} + \gamma \mathbf{X}_{st} + \delta_s + \lambda_t + \varepsilon_{st} \quad (3.1)$$

where $\log(CO2_{st})$ denotes the logarithm of per capita residential CO_2 emissions in state s in year t , D_{st} is a binary indicator equal to 1 if state s has implemented a net metering policy by year t and 0 otherwise, $\mathbf{X}_{st}$ is a vector of time-varying state-level control variables, δ_s and λ_t represent state and year fixed effects, respectively, and ε_{st} is the idiosyncratic error term. The coefficient of interest, β , captures the average treatment effect (ATT) of net metering policy adoption on residential CO_2 emissions.

The state fixed effects, δ_s , control for any time-invariant differences across states that may be correlated with both the adoption of net metering policies and CO_2 emissions. For instance, some states may have more environmentally conscious populations or a history of progressive energy policies, which could influence both the likelihood of adopting net metering and the level of emissions. Moreover, this also controls for different geographies and solar generation potential. The inclusion of state fixed effects ensures that we are not attributing these pre-existing differences to the causal effect of net metering.

Year fixed effects, λ_t , account for common shocks and trends that affect all states uniformly over time. Examples include federal policies such as tax credits for renewable energy, changes in the price of natural gas or other fuels, and macro-economic conditions that impact energy demand. By including year fixed effects, we control for these time-varying factors and isolate the impact of net metering policies.

The vector of time-varying controls, $\mathbf{X}_{st}$, includes relevant state-level variables that could influence residential CO_2 emissions. We control for per capita residential energy use, as states with higher energy consumption are expected to have higher emissions, all else equal. The total renewable energy per capita is included to capture the level of clean energy utilization in the state, with the expectation that a higher amount of renewable energy consumption is associated with lower CO_2 emissions due to the lower carbon intensity of clean energy sources. Residential electricity prices are

incorporated to account for the potential impact of energy costs on consumption and emissions, as higher prices may encourage conservation and the adoption of more efficient technologies. Median income per capita is included to control for economic factors that may influence energy use and emissions, with the hypothesis that higher-income states may have lower per capita emissions due to the adoption of cleaner technologies and more energy-efficient buildings. Finally, we control for heating degree days, which measure the energy demand for heating based on outdoor temperatures, as states with higher heating needs are expected to have higher residential CO_2 emissions. To account for potential serial correlation in emissions within states over time, we cluster the standard errors at the state level, following the common practice in the DiD literature with state-level policies (Bertrand et al., 2004).

Event Studies.

To investigate the temporal patterns of policy impacts and validate our identification strategy, we extend our baseline model to incorporate dynamic treatment effects. This approach allows us to examine pre-policy trends and assess the evolution of net metering's impact on CO_2 emissions over time. We estimate the following event study specification:

$$\log(CO2_{st}) = \sum_{k=-K, k \neq -1}^L \beta_k D_{s,t}^k + \gamma \mathbf{X}_{st} + \delta_s + \lambda_t + \varepsilon_{st} \quad (3.2)$$

In this equation, $D_{s,t}^k$ represents a series of binary indicators denoting the time relative to policy implementation in state s . Specifically, k indicates the number of years before or after the adoption of net metering, with $k = 0$ representing the year of policy enactment. For the TWFE OLS case, we exclude the indicator for $k = -1$ to serve as the reference period, against which all other coefficients are compared.

3.5.3 Staggered Difference-in-Differences with Heterogeneous Treatment Effects

Although the TWFE specification is a natural starting point, recent econometric literature has highlighted its limitations in settings with staggered treatment adoption and heterogeneous treatment effects (Goodman-Bacon, 2021; Callaway & Sant’Anna, 2021; L. Sun & Abraham, 2021). In particular, the TWFE estimator uses already-treated states as controls for later-treated states, which can lead to biased estimates of the ATT if the treatment effects vary across states or over time.⁴

To address the concerns with the TWFE estimator in settings with staggered treatment adoption and heterogeneous treatment effects, we employ the estimator proposed by Callaway & Sant’Anna (2021). This estimator compares the outcomes of treated states to those of not-yet-treated states, in our case, avoiding the problematic comparisons between already-treated and newly-treated states that can bias the TWFE estimates. The Callaway and Sant’Anna (2021) estimator is based on the concept of group-time average treatment effects, denoted as $ATT(g, t)$, which represent the average treatment effect for the group of states (i.e., G) first treated in period g ($G_g = 1$), measured in period t :

$$ATT(g, t) = \mathbb{E}[Y_t(g) - Y_t(0) | G_g = 1], \quad \text{for } t \geq g \quad (3.3)$$

where $Y_t(g)$ and $Y_t(0)$ are the outcomes with and without treatment, respectively.

To estimate the $ATT(g, t)$ parameters using not-yet-treated states as the comparison group, the preferred setting of this estimator employs a doubly robust approach that combines inverse probability weighting (IPW) and outcome regression. The doubly robust (DR) estimand for $ATT(g, t)$ when using not-yet-treated states as the comparison group is given by:

⁴Note that the control group consists of “not yet treated” states, which are states that have not yet received the treatment at a given point in time, but will receive it later. The critique of TWFE refers to instances where “already treated” states—those that have received the treatment earlier—are used as controls for later-treated states. This can lead to biased estimates when treatment effects are heterogeneous over time or across units. Our approach avoids this issue by using only “not yet treated” states as controls at each time point.

$$ATT_{dr}^{ny}(g, t) = \mathbb{E} \left[\left(\frac{G_g}{\mathbb{E}[G_g]} - \frac{\frac{p_{g,t}(X)(1-D_t)}{1-p_{g,t}(X)}}{\mathbb{E} \left[\frac{p_{g,t}(X)(1-D_t)}{1-p_{g,t}(X)} \right]} \right) (Y_t - Y_{g-1} - m_{g,t}^{ny}(X)) \right] \quad (3.4)$$

where $m_{g,t}^{ny}(X) = \mathbb{E}[Y_t - Y_{g-1} \mid X, D_t = 0, G_g = 0]$ is the expected change in the outcome between periods $g - 1$ and t for states with covariates X that are not yet treated by time t . More explicitly, the DR estimator for $ATT(g, t)$ is given by:

$$\widehat{ATT}_{dr}^{ny}(g, t) = \frac{1}{n_g} \sum_{i: G_i = g} \left(\frac{Y_{it} - Y_{i,g-1} - \hat{m}_{gt}^{ny}(X_i)}{\hat{p}_{g,t}(X_i)} - \frac{(1 - D_{it})(1 - G_{ig})}{\hat{p}_{g,t}(X_i)} (Y_{it} - Y_{i,g-1} - \hat{m}_{gt}^{ny}(X_i)) \right) \quad (3.5)$$

where n_g is the number of states in group g , $\hat{m}_{gt}^{ny}(X_i)$ is an estimate of the expected change in the outcome between periods $g - 1$ and t for state i with covariates X_i , conditional on not being treated up to time t , and $\hat{p}_{g,t}(X_i)$ is an estimate of the generalized propensity score (i.e., the probability of being treated at time g conditional on covariates X_i and not being treated up to time t).

The DR estimator combines the IPW and outcome regression approaches to achieve robustness to misspecification of either the propensity score model or the outcome regression model. By using the not-yet-treated states as the comparison group and properly adjusting for covariates, this estimator provides a consistent estimate of the group-time average treatment effects under the assumption of parallel trends conditional on covariates.

The group-time average treatment effects, $ATT(g, t)$, can be aggregated in various ways to obtain more interpretable summary measures of the impact of net metering policies. For instance, we can estimate event study parameters that capture the dynamic effects of the policy relative to the time of adoption:

$$\theta_e = \sum_g \omega_{g,e} ATT(g, g + e) \quad (3.6)$$

where e indexes the periods relative to the policy adoption date (e.g., $e = 0$ represents the year of adoption, $e = 1$ represents one year after adoption, etc.), and $\omega_{g,e}$ are weights that sum to one across groups for each event time e .

We can also compute an overall ATT parameter that combines all post-treatment periods:

$$\theta_{post} = \sum_g \omega_g \sum_{t \geq g} ATT(g, t) \quad (3.7)$$

where ω_g are weights that sum to one across groups. In addition to these aggregate measures, we examine the $ATT(g, t)$ parameters for each group and time period separately to uncover potential heterogeneity in the effects of net metering across states and over time.

3.5.4 Investigating Heterogeneity

In order to gain a deeper understanding of how policy design features influence the effectiveness of net metering in reducing emissions, we explore potential heterogeneity in the policy impacts across states with different net metering regulations. Specifically, we estimate separate models for subsamples of states based on the following policy characteristics identified by Smith et al. (2021) and summarized in Table 3.1. Smith et al. (2021) used the following approach to create dummy variables for the following policy features:

The *system size caps (kW)* values were divided into tertiles across all states. States with values in the first tertile (1000 kW and larger) were assigned a score of one, while states in the second or third tertiles received a score of zero. Similarly, *program size caps (% peak of utility)* were divided into tertiles. States with values in the first tertile (100% or unspecified cap) were given a score of one, while those in the second or third tertiles were assigned zero. *Excess electricity compensation* was coded as a binary measure. States that compensated excess generation at the retail rate received a score of one, while all other compensation rates (e.g., avoided cost, market rate) were scored zero. *Ownership of Renewable Energy Credits (RECs)* was also coded as a binary measure. States that assigned REC ownership to the consumer received a score of one. States where

RECs were owned by the utility or shared between customers and utilities were scored zero. The *number of eligible technologies* eligible for net metering was divided into tertiles. States with values in the first tertile (9 technologies or more) were assigned a score of one, while those in the second or third tertiles received a zero.

Table 3.1: Net Metering Policy Characteristics Summary.

Characteristic	Definition	Criteria for Score of 1	State Count
System Size Caps (kW)	Maximum size of individual systems allowed for net metering policies	System size values in first tertile (1000 kW and larger)	17
Program Size Caps (% peak)	Limit on sum of net metering systems allowed in state, calculated as percent of peak utility demand	Program size values in first tertile (100% or unspecified cap)	22
Excess Electricity Compensation	Compensation rate awarded to customers for net excess electricity generation	Retail rate compensation (vs. avoided cost, market rate, or other)	27
Ownership of RECs	Person or entity to which Renewable Energy Credits (RECs) are awarded by net metering policies	Customer ownership of RECs (vs. utility ownership or shared)	21
Number of Eligible Technologies	Number of renewable energy technologies eligible for net metering policies (e.g., solar PV, wind, biomass)	Number of technologies in first tertile (9 or more technologies)	15

Notes: The state count represents the sum of states where the dummies are one.

This heterogeneity analysis based on policy design features allows us to identify which specific aspects of net metering regulations are most critical for achieving residential CO_2 emissions reductions. This information is valuable for policymakers seeking to optimize the design of net metering programs and maximize their environmental benefits.

Furthermore, we investigate whether the impacts of net metering vary depending on states' political leanings. Specifically, we examine if the effect of net metering policies on residential CO_2 emissions differs between states that lean Democratic versus those that lean Republican, based on the state election outcomes closest to but before the adoption year.⁵ Political ideology could influence the uptake of net metering and the extent to which the policy displaces fossil fuel generation, as it

⁵In this case, we create two subsets from the original data where each state belongs to a particular subset given their political leanings.

may be correlated with state-level attitudes towards renewable energy, environmental regulations, and climate change mitigation efforts.

3.5.5 Inference and Robustness Checks

To ensure the robustness of our findings, we conduct several sensitivity analyses. First, we estimate models with alternative sets of control variables and fixed effects, such as region-by-year fixed effects to flexibly account for time-varying regional shocks.⁶ Second, we test for cross-state spillover effects by including indicators for the presence of net metering policies in neighboring states.⁷ Third, we conduct placebo tests that push back adoption time by five years for a random set of states and re-estimate our models, which helps to assess whether our results could be driven by chance or by underlying trends unrelated to net metering.

3.5.6 Examining Potential Mechanisms

To investigate the mechanisms through which net metering policies influence residential CO_2 emissions, we focus on three key metrics of renewable energy integration: energy sold back to the grid, installed capacity, and customer count. These metrics, primarily driven by rooftop solar installations, serve as intermediate outcomes. By examining the impact of net metering policies on these measures, we can assess whether these policies achieve emissions reductions by encouraging the deployment and utilization of clean distributed energy resources. Significant increases in these metrics following the implementation of net metering would suggest that the policy effectively promotes the transition to cleaner energy sources, thereby contributing to lower residential CO_2 emissions.

⁶The results for this part are included in the Baseline section.

⁷Our approach accounts for both potential policy diffusion and policy spillover effects. The diffusion of net metering policies refers to how states learn from and adopt similar policies implemented by their neighbors. Spillover effects, in contrast, occur through regional market mechanisms: when one state adopts net metering, the expanded market attracts more solar companies to the region, potentially reducing installation costs through economies of scale even in states without such policies. Additionally, given the interconnected nature of electricity grids, increased renewable generation in one state can displace fossil fuel generation across state lines, leading to emissions reductions beyond state borders.

3.6 Data

3.6.1 Variable Descriptions and Descriptive Analysis

We use a dataset compiled from various sources, including the Environmental Protection Agency (EPA), Energy Information Administration (EIA), National Oceanic and Atmospheric Administration (NOAA), Census Bureau, and Bureau of Labor Statistics (BLS). The dataset spans from 1990 to 2020 and covers all 50 U.S. states. Table 3.2 provides the list of the variables collected to estimate the impact of net metering policy adoption on residential GHG emissions. The last four columns of the table also presents summary statistics for each variable across states and years. On average, states have annual residential greenhouse gas emissions of 7.44 million metric tons (MMT) and residential energy use of 409,130 million British thermal units (MBTU).⁸ The average state population is approximately 6 million, with a median income of \$66,488 (adjusted to 2022 dollars). Heating degree days (HDD) and cooling degree days (CDD) vary substantially across states, with mean values of 5,099 and 1,159, respectively.⁹ Renewable energy consumption averages 37.23 thousand BTU per state, while the average residential electricity price is \$18.68 per MBTU.

Due to the high correlation between heating degree days (HDD) and cooling degree days (CDD), we only include HDD in our analysis. We divide it by 1,000 to interpret the associated coefficient as the effect of a 1,000-unit increase in HDD on residential GHG emissions later on.¹⁰ Furthermore, we use the total population variable to create per capita measures for residential energy use, median income, and renewable energy consumption. By expressing these variables in per capita terms, we account for differences in state sizes and population levels. Finally, we take the

⁸An overview of how residential GHG emissions are calculated and how they differ from other sectors, as well as its relevance for our study, is provided in the appendix.

⁹Heating and cooling degree days measure how much the daily mean temperature falls below (HDD) or exceeds (CDD) 65°F, which indicate heating or cooling needs.

¹⁰This variable is originally expressed in thousands because yearly data is obtained by summing daily degree days across a given year.

natural logarithm of the per capita variables (except for HDD) to standardize the scale of the data and facilitate the interpretation of the estimates as percentage changes in residential GHG emissions.

Table 3.2: Variable Description and Summary Statistics.

Variable	Unit	Data Source	Mean	Std. Dev.	Min	Max
CO ₂ Residential Emissions*	Metric Tons	EPA	1.38	0.64	0.01	3.89
Residential Energy Use*	Thousand BTU	EIA	72.98	12.70	21.89	106.07
Heating Degree Days	Thousand Degree Days	NOAA	5.21	2.30	0.00	11.70
Cooling Degree Days	Thousand Degree Days	NOAA	1.14	0.96	0.00	5.21
Median Income*	USD (2022)	Census	66,164.80	11,338.84	38,600.00	108,900.00
Renewable Energy Use*	Thousand BTU	EIA	37.23	40.51	1.63	226.61
Retail Electricity Price	USD per Thousand BTU	EIA	18.68	9.11	8.04	99.63

Notes: *: expressed at the per capita level. The averages reported here are calculated using state-year-level data. These data are obtained for all 48 states, in addition to Washington D.C. and Hawaii for the period 1990-2020.

Figure 3.1 illustrates the year of net metering policy adoption for each state, as obtained from Smith et al. (2021). The earliest adopters were Idaho (1980), Arizona (1981), Massachusetts (1982), and Wisconsin (1982). As of 2020, only three states – Alabama, South Dakota, and Tennessee – had not implemented net metering policies. We focus on the post-1990 period since there's no state-level emissions data available before 1990. Hence, states that adopted net metering before 1990 are dropped from the analysis, resulting in a sample of 38 states and 1,178 observations.

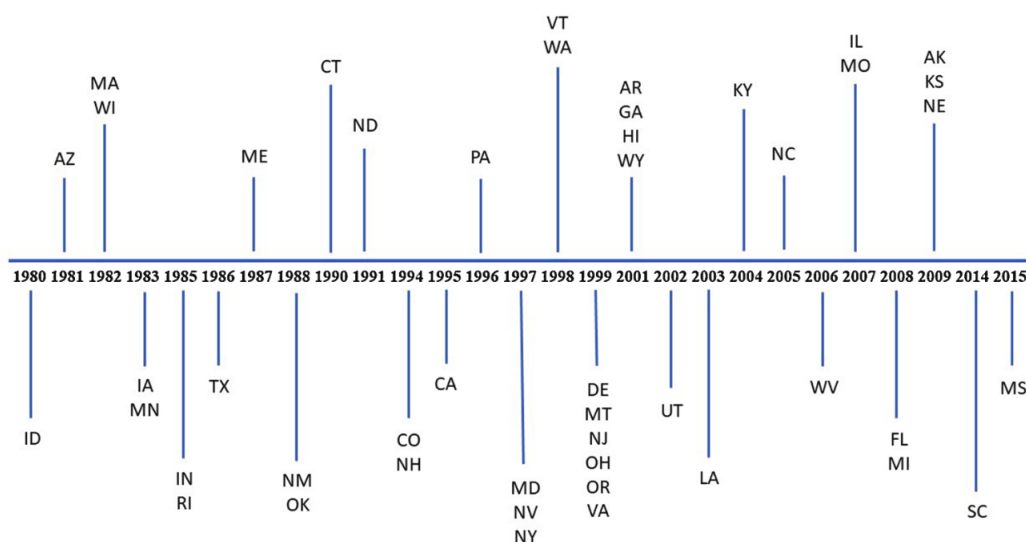


Figure 3.1: Net Metering Policy Adoption Year by State
Source: Smith et al. (2021)

Figure 3.2 reveals substantial variation in per capita residential GHG emissions across U.S. states from 1990 to 2020. Alaska leads with the highest per capita emissions at 2.58 MT, followed by Maine (2.55 MT), Vermont (2.38 MT), Michigan (2.30 MT), and Connecticut (2.23 MT). In contrast, Hawaii has the lowest per capita emissions (0.08 MT), followed by Florida (0.15 MT), Arizona (0.43 MT), South Carolina (0.57 MT), and Texas (0.63 MT). Notably, some populous states with high total emissions, such as California and New York, do not appear among the highest per capita emitters. This suggests that urbanization and population density may enable more energy-efficient living arrangements.

Figure 3.2: Average Residential GHG Emissions per Capita by State (MT; 1990-2020).

Figure 3.3 shows the average residential GHG emissions across all states for each year from 1990 to 2020. Emissions fluctuate over time, with notable peaks in 1996 (7.96 MMT), 2003 (7.85 MMT), and 2019 (7.66 MMT), and low points in 2012 (6.12 MMT) and 2016 (6.54 MMT). These fluctuations were driven by three main factors: energy price volatility marked by sharp increases in the early 2000s, extreme weather events such as severe winters in 1996 and 2003 and an unusually warm winter in 2012, and broader energy transition trends including the gradual shift from coal to natural gas and renewables post-2005. We also plot a dynamic map of residential

GHG emissions by state for each year from 1990 to 2020.¹¹ This dynamic map provides a visual representation of the temporal and spatial variation in emissions across the United States. This substantial variation in residential GHG emissions across states and time stems from several key factors. Regional differences in electricity generation sources significantly influence emission levels. States in the Midwest and Appalachia traditionally rely more heavily on coal-fired power plants, while coastal states often utilize more natural gas and renewables. Climate patterns also play a crucial role, as states with extreme temperatures experience higher energy demands for heating (Northeast) or cooling (South). Additionally, variations arise from differences in population density and urbanization, with urban areas potentially benefiting from shared infrastructure and energy efficiency, while rural areas may depend more on individual heating systems. Economic factors further contribute to this variation, as higher-income regions may have larger homes with more appliances, yet simultaneously possess greater capacity to invest in energy-efficient technologies.

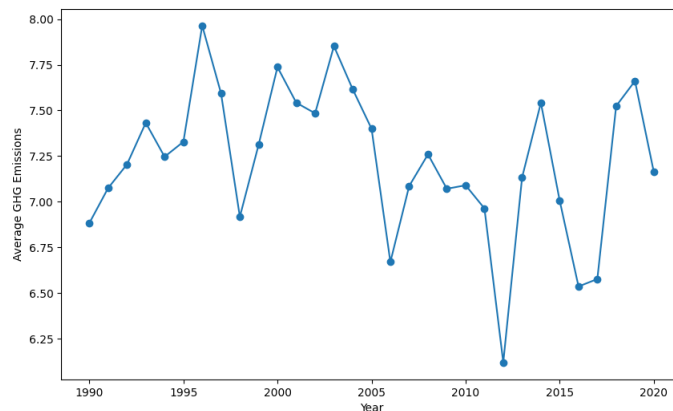


Figure 3.3: Average Residential GHG Emissions across Years (MMT; 1990-2020).

Figure 3.4 presents the average residential energy use per capita by state from 1990 to 2020. North Dakota leads with the highest per capita consumption at 93,890 BTU, followed by Missouri (87,020 BTU), Nebraska (86,260 BTU), Kentucky (85,070 BTU), and Tennessee (84,720 BTU). These high consumption levels reflect primarily cold weather demands in North Dakota and the central states, combined with substantial cooling needs in states like Tennessee and Kentucky. At

¹¹The following is the link to the YouTube video: <https://youtu.be/wCMsXkvCimU>

the other end of the spectrum, Hawaii shows the lowest per capita consumption at 25,030 BTU, followed by California (40,470 BTU), New Mexico (55,680 BTU), Utah (58,540 BTU), and New York (58,610 BTU). Hawaii's low consumption stems from its mild climate requiring minimal heating or cooling, while California and New York's relatively low per capita usage suggests the efficiency benefits of their stringent energy policies.

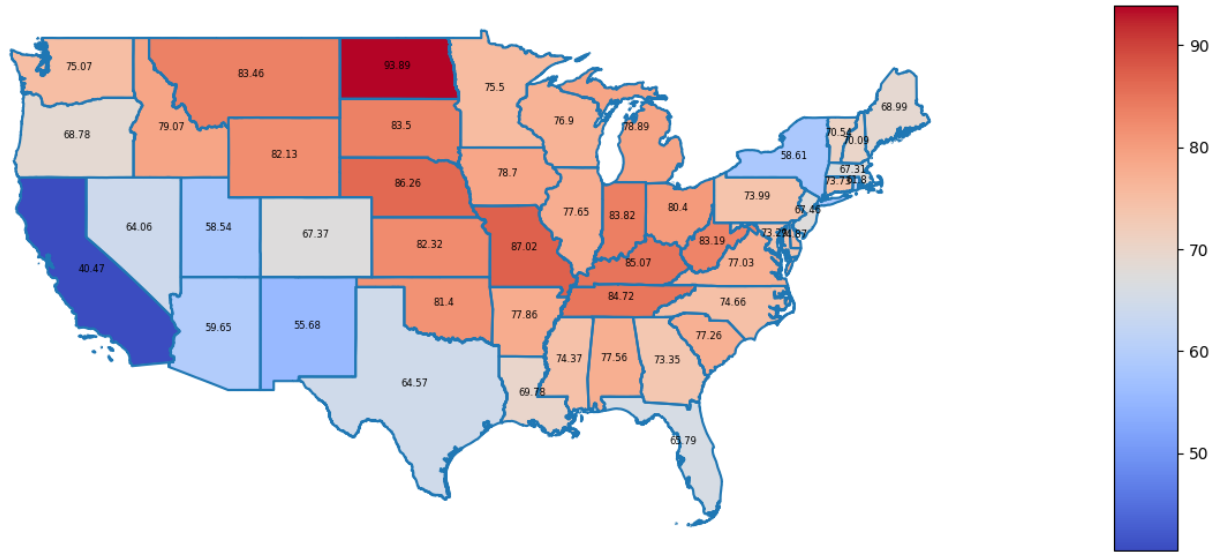


Figure 3.4: Average Residential Energy Use per Capita by State (Thousands BTU; 1990-2020).

3.6.2 Balancing Checks

To determine whether states that have not yet adopted net metering policies (i.e., not yet treated group) are a suitable comparison group for states that have implemented such policies (i.e., treated group), we conduct balancing tests on key state-level characteristics. These tests assess whether, prior to adopting net metering, the states that eventually implement the policy are comparable to those that have not. Any systematic difference could potentially confound the estimated treatment effect.

As in Bielen (2024), the balancing tests involve regressing the treatment indicator on each control variable while controlling for state and year fixed effects. The treatment indicator equals one

for states that have adopted net metering policies and zero for states that have not yet adopted such policies at a given point in time. The coefficients from these regressions are presented in Table 3.3.

Table 3.3: Balancing checks.

Variable	Policy Adoption Coef.	R^2
Heating Degree Days	-0.000 (0.989) [1.000]	0.668
log (Residential Energy Use) pc	0.260 (0.415) [1.000]	0.669
log (Adjusted Median Income) pc	0.339 (0.140) [0.837]	0.671
log (Renewable Energy Use) pc	-0.063 (0.378) [1.000]	0.670
log (Retail Electricity Price)	-0.011 (0.972) [1.000]	0.668
Observations	1178	

Notes: The second column shows the result from a separate simple linear regression with the response shown in the column label. Coefficients are followed by their normal p-values in parentheses and Bonferroni corrected p-values in square brackets, each on separate lines. The control group is “Not yet treated”. All regressions include state and year fixed effects. Standard errors (not reported here) are clustered at the state level in the regressions. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

The results of the balancing tests show no statistically significant differences between the treated and not yet treated states for any of the examined state-level characteristics. Moreover, when adjusting for multiple hypothesis testing using the Bonferroni correction, the p-values remain far from any meaningful significance level.

The absence of systematic differences between the treated and not yet treated states suggests that the not yet treated states are indeed a suitable comparison group. This comparability between the treated and not yet treated states strengthens the credibility of the identification strategy. It suggests that any observed differences in residential CO_2 emissions between the two groups after

the policy adoption can be more confidently attributed to the causal effect of net metering policies rather than preexisting differences in state-level characteristics.

3.7 Results

3.7.1 Overall ATTs and Event Studies

Overall Estimates.

Table 3.4 presents the impact of net metering adoption on residential GHG emissions under different model specifications based on Equation 3.1. Across all the TWFE specifications (columns 1-4), the estimates are biased towards zero and not statistically significant. Our preferred specification (column 5), however, reveals a statistically significant reduction in residential GHG emissions per capita of approximately 0.823%, on average, following net metering policy adoption ($p < 0.01$). The notable contrast between the CS estimate (-0.823%) and the TWFE estimates (ranging from 0.010% to 0.026%) stems from the CS estimator's ability to account for heterogeneous treatment effects across states and time, as well as its avoidance of using already-treated states as controls for later-treated states. Unlike the TWFE estimator, which can dilute treatment effects by including states at different stages of net metering policy adoption, the CS estimator uses untreated or not-yet-treated states as controls and employs propensity score matching to ensure comparability, yielding a clearer, more accurate estimate of net metering policy's impact on residential GHG emissions.

It is important to note that these estimates represent the average treatment effect of net metering adoption on residential emissions across all states in our sample. This includes both states that experienced high uptake of net metering and those with lower participation rates. As such, this may mask potential heterogeneity in the impact of net metering policies across states with different levels of policy effectiveness or solar penetration. We explore this heterogeneity in subsequent analyses.

Table 3.4: Impact of Net Metering on Residential GHG Emissions.

	Log Residential GHG Emissions pc				
	(1)	(2)	(3)	(4)	(CS)
Net Metering Indicator	0.023 (0.029)	0.026 (0.032)	0.010 (0.030)	0.017 (0.029)	-0.823*** (0.214)
Observations	1,178	1,178	1,147	1,147	1,178
Year fixed effects	✓	✓	✓	✓	✓
State fixed effects	✓	✓			✓
Region fixed effects			✓	✓	
Controls		✓	✓	✓	
Linear time trends * Region				✓	

Notes: Columns (1) to (4) present different specifications for the TWFE estimator. Column (1) reports results without covariates. Column (2) includes the covariates mentioned in the data section. Column (3) adds region fixed-effects based on NOAA-defined climate regions. Column (4) incorporates region fixed-effects and linear time trends to allow for the possibility that states in different regions might be on different emissions trajectories. Column (5) displays the results for the CS estimator. Standard errors are clustered at the state level and are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Event Studies.

Figure 3.5 displays the event study results for the impact of net metering policies on residential GHG emissions. This plot reveals several key insights about the effects of net metering policy and the validity of our identification strategy. First, looking at the pre-treatment periods ($k < 0$), we see that the coefficients are close to zero and statistically indistinguishable from the baseline. This lack of a significant pre-trend in the differences supports our parallel trends assumption, giving us confidence that our estimates capture the causal effect of net metering policies. In the case of the post-treatment periods, we observe an interesting dynamic pattern. Immediately after policy adoption ($k = 0$), there is a small but statistically insignificant decrease in emissions. However, for both TWFE OLS and our preferred estimator, Callaway-Sant’Anna (CS), the effect becomes more pronounced and statistically significant by the second year post-adoption, with emissions decreasing by about 0.029% with the TWFE OLS estimator and by approximately 0.34% with the CS estimator.

The discrepancy between these two estimators is due to the fact that TWFE OLS is biased —as explained in the methodology section— towards zero in this case.

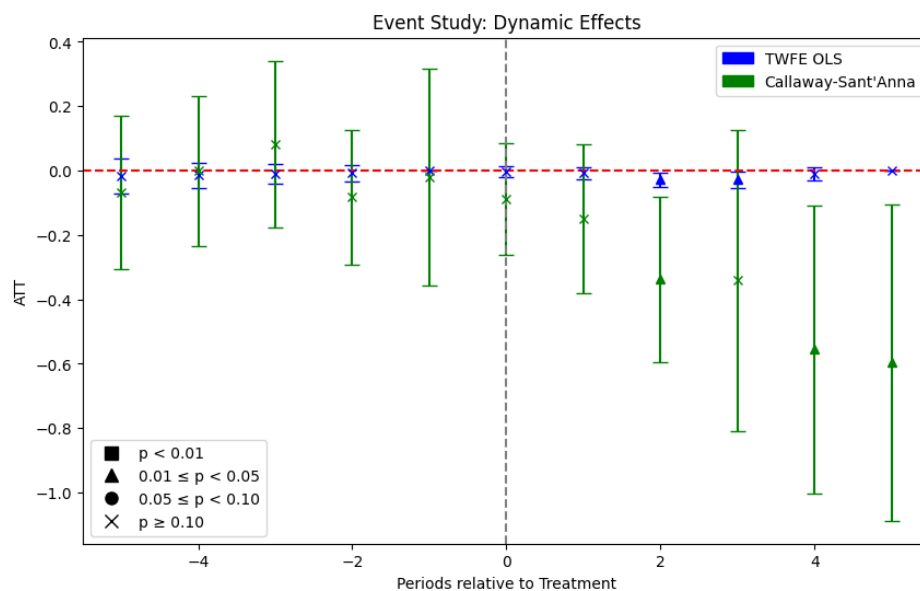


Figure 3.5: Impact of Net Metering Adoption on Residential GHG Emissions Relative to Adoption Year.

Notes: This figure presents event-study plots constructed using two different estimators: a dynamic version of the TWFE model (in blue with square markers) and Callaway & Sant'Anna (2021) (in green with triangle markers). The outcome variable is residential GHG emissions. The horizontal axis represents the periods relative to the net metering policy implementation, with $k = -1$ as the reference period excluded from the plot for the case of TWFE OLS. Each point denotes the average treatment effect (ATT) for the specified period, and the bars represent 95% confidence intervals with standard errors clustered at the state level.

This delayed impact makes sense given the nature of net metering policies. It likely takes time for households to learn about the new incentives, decide to install solar panels, and complete the installation process. The two-year lag we observe aligns with what we might expect given these practical constraints. Interestingly, the effect remains significant in year 5 with a reduction of about 0.6% in the case of the CS estimator. This sustained negative impact suggests that the net metering policy has a lasting effect on reducing emissions. The pattern indicates a robust and enduring response to the policy implementation, which warrants further investigation into its compound effect on residential emissions over time.

3.7.2 Examining Heterogeneity

Policy Favorability.

We start with estimating separate models for states with more favorable and less favorable net metering regulations.¹² As shown in Figure 3.6, the overall ATT for states with more favorable policies is positive at 2.49, but statistically insignificant ($p = 0.182$). In contrast, the overall ATT for states with less favorable policies is negative and significant at the 5% level, with a magnitude of -0.295 ($p = 0.039$).¹³

This stark difference suggests that the design and implementation of net metering policies play a crucial role in determining their effectiveness at reducing emissions. States with more favorable policies, such as higher capacity limits, more generous compensation rates, and fewer restrictions on eligible technologies, may create stronger incentives for households to adopt rooftop solar and other distributed renewable energy systems. However, the insignificant overall ATT for this group indicates that these incentives may not be sufficient to drive substantial reductions in residential emissions on average. On the other hand, the negative and significant ATT for states with less favorable policies is somewhat counterintuitive. One might expect that states with more restrictive net metering regulations would see smaller or no reductions in emissions. The scoring system used for determining favorability is likely the culprit, we provide a deeper explanation for this in the discussion section.

Examining the ATTs by adoption cohort provides insights into how the timing of the net metering policy adoption may influence its effectiveness. Looking at the ATTs by cohort (i.e., group of states that adopted the policy in the same year), we see that for states with more favorable policies, the treatment effects are positive and significant for the 1994 and 1997 cohorts, but negative and significant for the 2005 cohort. The large, positive, and highly significant ATT for the 1997 cohort

¹²Note that the dummy assignment was designed such that a score of one reflects a better outcome from the energy consumer standpoint. States with a score of at least 3 out of 5 (based on dummy variables for the five policy traits previously mentioned) are considered to have more favorable policies, while states with a score of 2 or less are considered to have less favorable policies. Never-treated states are included in all subsets.

¹³We discuss the intuition or counterintuition behind all the results in the discussion section.

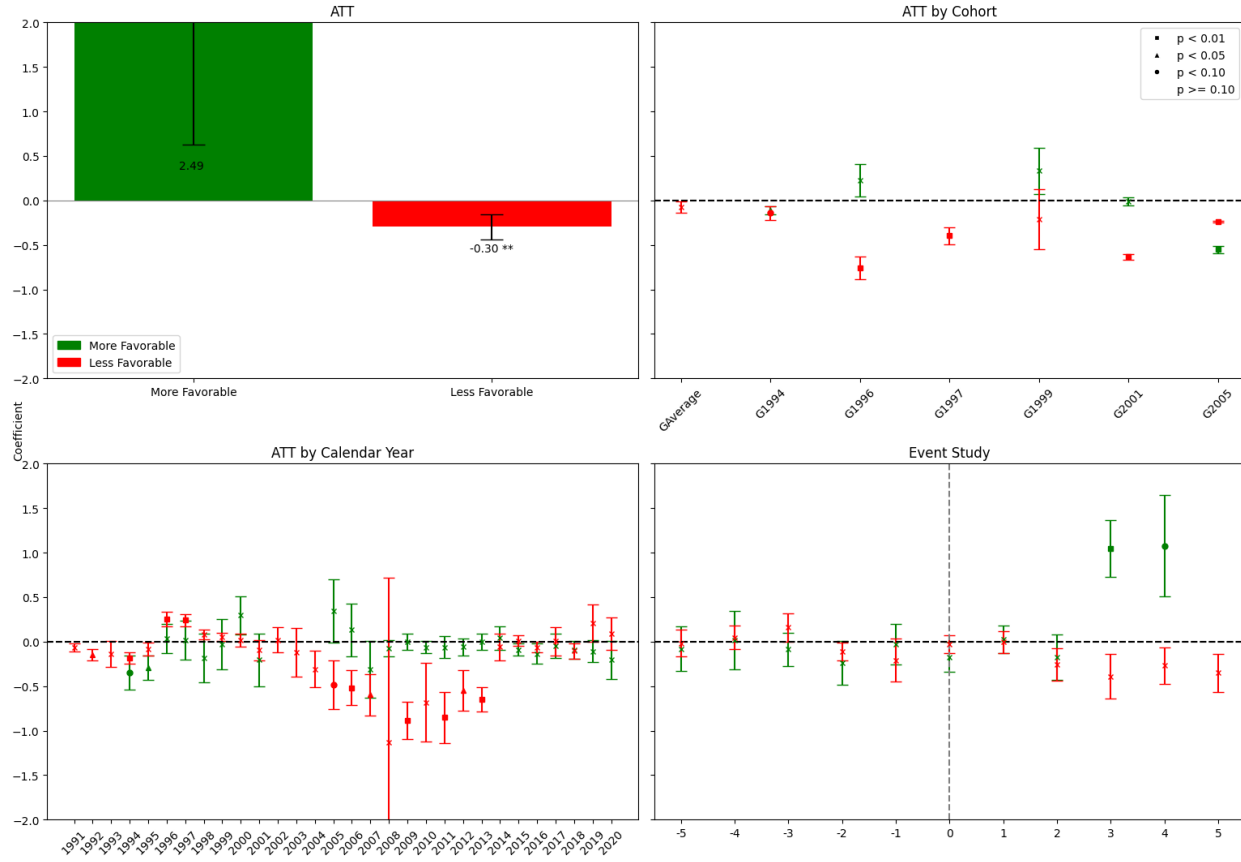


Figure 3.6: Heterogeneity between States of More versus Less Favorable Net Metering Policies.

Notes: The plots display the ATT for the different categories and time periods. The bar plots represent the overall ATT for each category. The cohort and calendar year plots show the ATT by cohort groups and calendar years, respectively. The event study plot illustrates the dynamic effects of treatment over time. The legends in the top quadrants apply to all the sub-quadrants. Significance levels are indicated by different markers: squares for $p < 0.01$, triangles for $p < 0.05$, circles for $p < 0.10$, and crosses for $p \geq 0.10$. Estimates greater than the absolute value of 2.0 are cropped out of the graph because they mask relevant patterns. These are, however, mentioned in the description of the plot if found statistically significant.

(20.68, $p < 0.001$) is particularly noteworthy and warrants further investigation. For states with less favorable policies, the ATTs are negative and significant across most cohorts.

Breaking down the ATTs by calendar year allows us to identify whether certain years were more conducive to the net metering policy success, regardless of when individual states adopted the policy. For states with more favorable policies, the treatment effects are negative and significant in the early years (1994-1995), but become positive and significant in later years (2002-2003), albeit

with large standard errors. For states with less favorable policies, the ATTs are consistently negative and often significant across most years, with some variation in magnitude over time.

Finally, in the case of the event study plots, for states with more favorable policies, the pre-treatment effects are generally insignificant, providing further support for the parallel trends assumption. The post-treatment effects are positive and significant in the third and fourth years after adoption which suggests a delayed impact on emissions. For states with less favorable policies, the pre-treatment effects are also insignificant, but the post-treatment effects are negative and become larger in magnitude over time, although not always statistically significant. These unexpected results here are likely due to the fact that, even when a group of states has the same favorability score, the composition of what policy traits drive their individual scores may be entirely different.

Policy Traits.

The heterogeneous effects of net metering policies on residential GHG emissions vary substantially across different policy design features. As shown in Figure 3.7, the overall ATTs are negative and significant for excess electricity compensation, system size caps, and REC ownership, but insignificant for the number of eligible technologies and program size caps. This suggests that more favorable compensation schemes, higher system size limits, and customer ownership of RECs are key drivers of the policy's effectiveness in reducing emissions.

The cohort-specific ATTs reveal some interesting patterns, with consistently negative and significant effects for excess electricity compensation, system size caps, and REC ownership, but more mixed results for the other policy categories. The calendar year ATTs show similar trends, with the most pronounced and consistent effects for excess electricity compensation, system size caps, and REC ownership.

The event study plots generally support the parallel trends assumption for these three policy features, with insignificant pre-treatment effects and negative, significant post-treatment effects that increase in magnitude over time. The number of eligible technologies and program size caps exhibit less clear patterns in the event study.

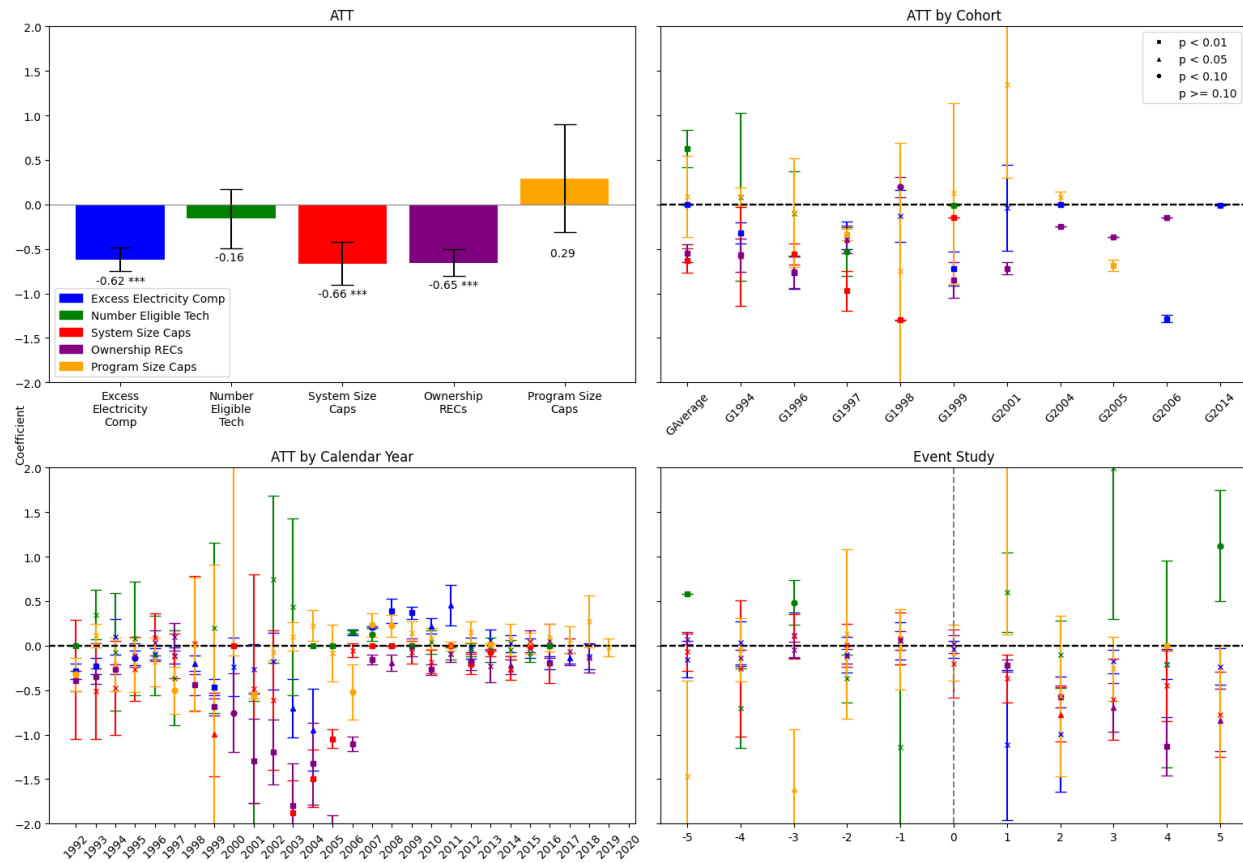


Figure 3.7: Heterogeneity by Policy Traits.

Notes: The plots display the ATT for the different categories and time periods. The bar plots represent the overall ATT for each category. The cohort and calendar year plots show the ATT by cohort groups and calendar years, respectively. The event study plot illustrates the dynamic effects of treatment over time. The legends in the top quadrants apply to all the sub-quadrants. Significance levels are indicated by different markers: squares for $p < 0.01$, triangles for $p < 0.05$, circles for $p < 0.10$, and crosses for $p \geq 0.10$. Estimates greater than the absolute value of 2.0 are cropped out of the graph because they mask relevant patterns. These are, however, mentioned in the description of the plot if found statistically significant.

Political Leanings.

The heterogeneous effects of net metering policies on residential GHG emissions differ markedly between Democratic and Republican states, as shown in Figure 3.8. As mentioned previously, this classification is based on the election results closest to but before the adoption date in each state. It is important to note that a state's political leaning can change over time, and the binary classification may not fully capture the nuances of a state's political environment. Moreover, the

political affiliation of a state's governor and legislature may not always align, leading to potential differences in policy implementation and effectiveness.¹⁴ The overall ATT for Democratic states is negative but not statistically significant, with a magnitude of -0.31. In contrast, the overall ATT for Republican states is about the same (-0.30) but highly significant ($p < 0.01$). This suggests that net metering policies have been more effective at reducing emissions in Republican-leaning states.

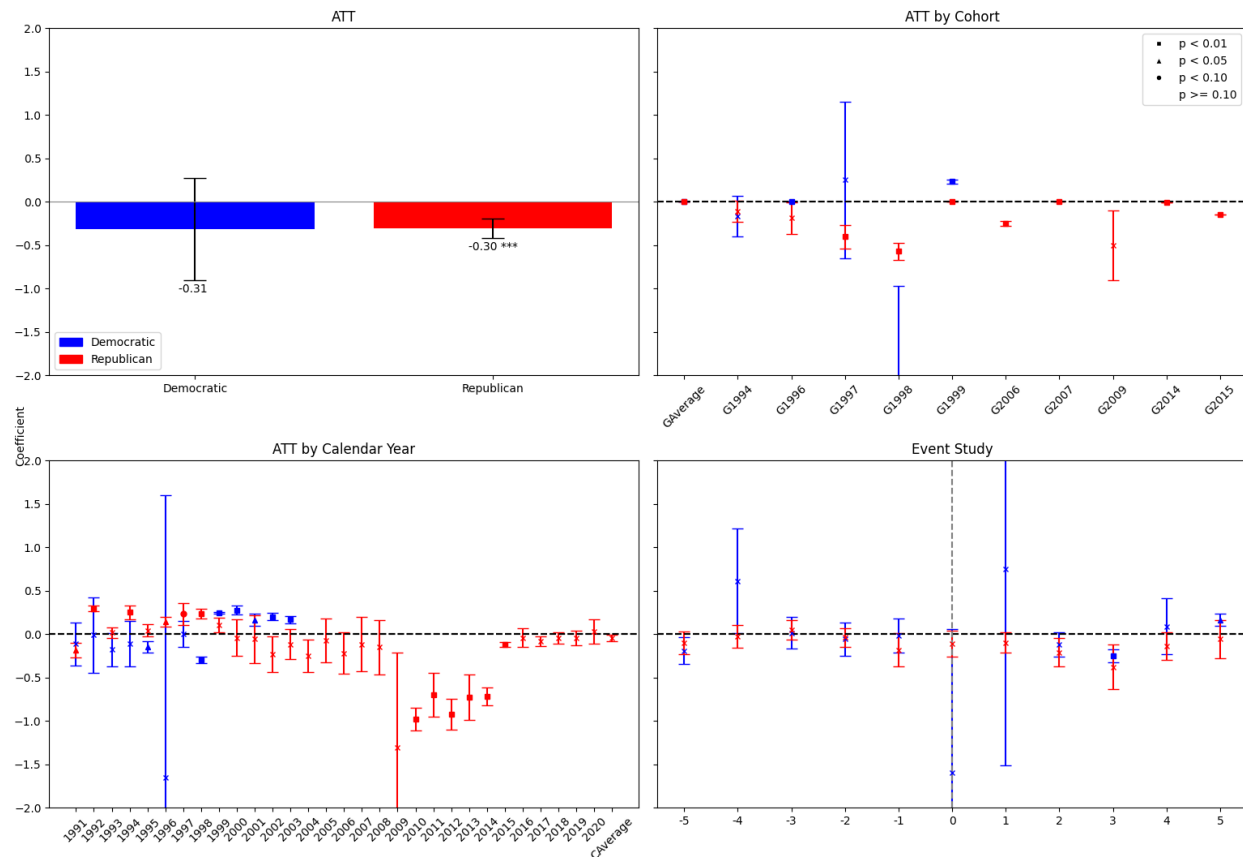


Figure 3.8: Heterogeneity between Democratic versus Republican States.

Notes: The plots display the ATT for the different categories and time periods. The bar plots represent the overall ATT for each category. The cohort and calendar year plots show the ATT by cohort groups and calendar years, respectively. The event study plot illustrates the dynamic effects of treatment over time. The legends in the top quadrants apply to all the sub-quadrants. Significance levels are indicated by different markers: squares for $p < 0.01$, triangles for $p < 0.05$, circles for $p < 0.10$, and crosses for $p \geq 0.10$. Estimates greater than the absolute value of 2.0 are cropped out of the graph because they mask relevant patterns. These are, however, mentioned in the description of the plot if found statistically significant.

¹⁴Future research could explore more granular measures of political context and how shifts in political power within states influence the impact of net metering policies.

Examining the group-specific ATTs reveals that Republican states show significant emissions reductions for the 1997, 1998, and 2006 adoption cohorts (ranging from -0.25 to -0.57). For Democratic states, the ATTs are more mixed, with only the 1999 cohort showing a significant effect (0.23), though in the opposite direction than expected. The calendar year ATTs show that Republican states experienced significant emissions reductions particularly during 2010-2014 (ranging from -0.72 to -0.98). In Democratic states, the effects vary over time, with some significant reductions in the late 1990s but positive and significant effects during 2001-2006.

In the case of the event study plots, the pre-treatment effects are insignificant for Republican states, supporting the parallel trends assumption. The post-treatment effects are negative and not statistically significant over the specified window. For Democratic states, the pre-treatment effects are also insignificant, but the post-treatment effects are more mixed, with significant negative impacts in the third year after adoption. These results suggest that the political context plays a crucial role in shaping the environmental effectiveness of net metering policies. Given the high level of statistical significance for Republican states, we argue that this is likely due to having more potential for GHG emissions reductions, given their typically higher baseline emissions levels.

Cumulative Effects.

While our primary analysis examines the impact of net metering policies at the state-year level, our dataset allows for a more granular investigation of cumulative effects over time. Specifically, we can exploit variation in the duration of exposure to net metering policies across states to understand how the impact of these policies evolves with longer implementation periods. To analyze these cumulative effects, we modify our baseline specification to incorporate a measure of policy exposure duration. We estimate the following equation:

$$\log(CO2_{st}) = \alpha_s + \gamma_t + \sum_{d=0}^{D=5} \beta_d \cdot \text{Duration}_{d(s,t)} + \mathbf{X}_{s,t} \cdot \delta + \epsilon_{s,t} \quad (3.8)$$

Here, everything else is as before. $\text{Duration}_{d(s,t)}$ is a set of indicator variables equal to 1 if state s at time t has had a net metering policy in place for d years. The duration of policy exposure

is calculated as $d = NM_{s,t} \cdot (t - \max\{T_s^{impl}, T_s^{start}\})$, where t represents the current year, T_s^{impl} is the year net metering was implemented in state s , T_s^{start} is the first year of our study period for state s , and $NM_{s,t}$ is an indicator for whether state s has a net metering policy in place by year t . We cap d at a maximum of 5 years to ensure sufficient sample size in each duration bin.¹⁵ This specification allows us to trace out the cumulative effects of net metering policies over time, providing insights into whether the impact of these policies strengthens, stabilizes, or potentially diminishes with longer exposure.¹⁶

As we can see in Figure 3.9, the net metering policy demonstrates a relatively stable pattern of small negative effects on residential GHG emissions per capita across all years of implementation. The effect is evident from the year of implementation, with a reduction of about 0.069% in emissions, though this initial effect is only marginally significant ($p < 0.10$). The impact becomes more pronounced and statistically significant in subsequent years, peaking at a 0.077% reduction two years after implementation.

While the effect strengthens and becomes highly significant ($p < 0.01$) in years 2 and 3, with reductions of 0.077 and 0.071% respectively, we observe a slight tapering and reduced significance ($p < 0.05$) in year 4, with a reduction of 0.052%. This pattern suggests that net metering policies lead to persistent, albeit small, reductions in residential GHG emissions, with the strongest and most significant effects observed in the middle years post-implementation.

The valleys in years 2-3, with reductions of 0.077 and 0.071% respectively (both significant at $p < 0.01$), represent the cumulative effect of early adopters fully integrating rooftop solar, combined with increasing participation from additional households over time. The slight decrease in the

¹⁵To illustrate the calculation of the *Duration* variable, consider the following example: Suppose State A implemented a net metering policy in 2010, and our study period begins in 2008. In 2012, the Duration for State A would be calculated as follows: $d = NM_{A,2012} \cdot (2012 - \max\{2010, 2008\}) = 1 \cdot (2012 - 2010) = 2$. This indicates that in 2012, State A had been exposed to the net metering policy for 2 years. In contrast, for a State B that implemented net metering in 2014, the Duration in 2012 would be: $d = NM_{B,2012} \cdot (2012 - \max\{2014, 2008\}) = 0 \cdot (2012 - 2014) = 0$. This reflects that State B had not yet implemented net metering by 2012.

¹⁶Other than for practical reasons, we cap d at 5 years because it is reasonable to assume that the most substantial uptake of net metering occurs within the first 3-5 years after policy implementation, as households respond to new financial incentives and installation costs decline.

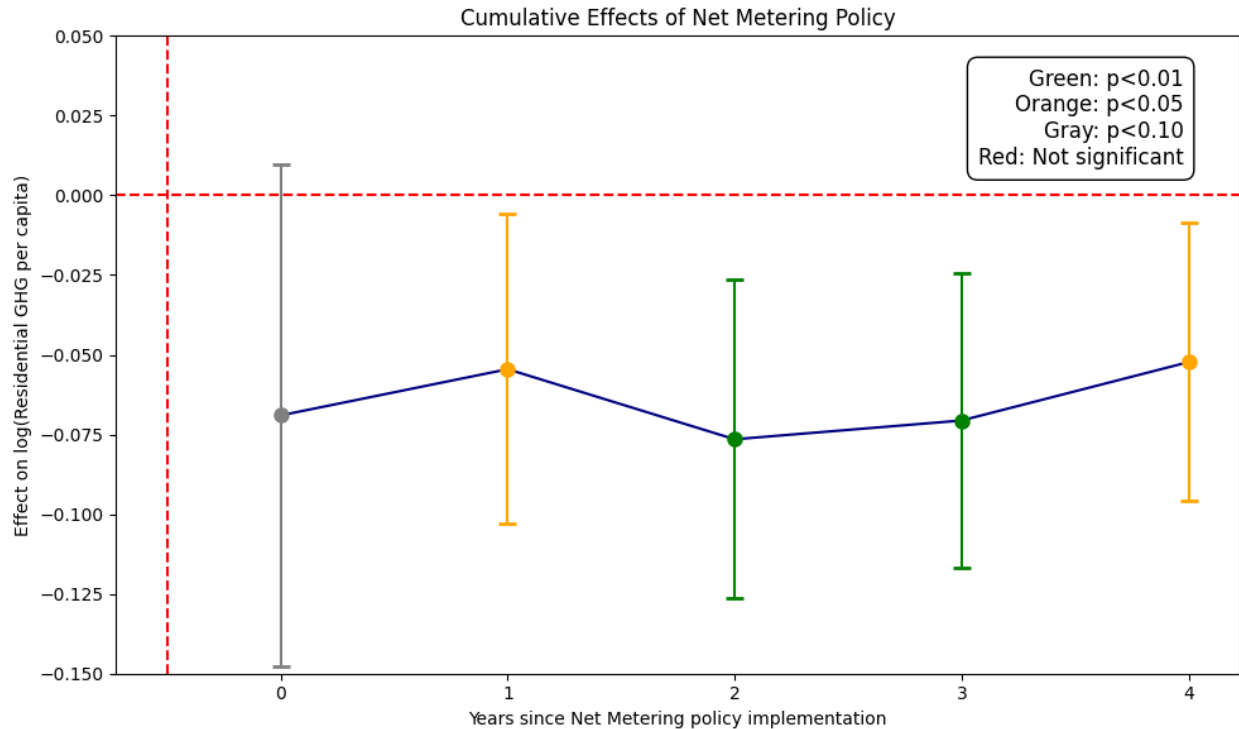


Figure 3.9: Cumulative Effects of the Net Metering Policy.

Notes: This figure displays the cumulative effects of the net metering policy on log(Residential GHG per capita) over time. The x-axis represents the number of years since the policy implementation, ranging from 0 (adoption year) to 4 years (the results did not converge for year 5). Each point shows the estimated effect for that year, with error bars indicating 95% confidence intervals. The colors of the points represent different levels of statistical significance: green for $p < 0.01$, yellow for $p < 0.05$, gray for $p < 0.10$, and red for $p \geq 0.10$. The dashed red horizontal line at $y = 0$ represents no effect, while the dashed red vertical line at $x = -0.5$ separates the pre-implementation period from the post-implementation period. The navy line connecting the points helps visualize the trend of the cumulative effects over time.

cumulative effect in year 4 (0.052%, $p < 0.05$) may not necessarily indicate a reduction in the policy's impact, but rather a potential slowdown in the rate of additional benefits. This could be due to a gradual approach towards market saturation, where the most enthusiastic adopters have already participated, or it might reflect a stabilization of the cumulative effects as the market matures. Importantly, the persistent negative coefficients across all years highlights the enduring cumulative climate benefits of the policy, even as the rate of additional impact may fluctuate over time.

3.7.3 Robustness Checks

Policy Diffusion/Spillover.

To account for potential policy spillovers across states, we extend our baseline model to include a measure of net metering policy adoption in neighboring states. This approach allows us to test whether a state's outcomes are influenced not only by its own policy but also by the policies of its neighbors. We modify our baseline difference-in-differences model as follows:

$$\log(CO2_{st}) = \alpha_s + \delta_t + \beta D_{st} + \gamma N_{st} + \mathbf{X}'_{st} \theta + \epsilon_{st} \quad (3.9)$$

where the new term, N_{st} , represents the count of neighboring states with active net metering policies. We construct N_{st} as follows:

$$N_{st} = \sum_{j \in J_s} \mathbb{1}(T_j \leq t) \quad (3.10)$$

where J_s is the set of neighboring (i.e., sharing a border) states s , T_j is the first year of treatment for state j , and $\mathbb{1}(\cdot)$ is the indicator function. This measure counts the number of neighboring states that have adopted net metering policies by year t .¹⁷ The coefficient γ captures the spillover effect from neighboring states' policies. A negative and significant coefficient would indicate positive policy spillovers, suggesting that a state's residential GHG emissions decrease when its neighbors adopt net metering policies. Conversely, a positive coefficient would suggest that neighboring policies have an adverse impact on a state's emissions.

The results from our policy diffusion analysis, presented in Table 3.5 and Figure 3.10, offer interesting insights when compared to our baseline findings. In the baseline fixed-effects model (Column 1 of Table 3.5), we observe that the coefficient on the net metering indicator (D_{st}) is positive

¹⁷For example, suppose neighboring state j adopts the net metering policy in 2003 ($T_j = 2003$). In years before adoption ($t = 2000, 2001, 2002$), the indicator function $\mathbb{1}(T_j \leq t)$ equals 0 because $T_j \leq t$ is false. Starting from the adoption year onward ($t = 2003, 2004, 2005$), the indicator equals 1 since $T_j \leq t$ holds true. This means N_{st} correctly counts the policy as active in neighboring state j from 2003 onward.

(0.033) but statistically insignificant. This is consistent with our previous baseline results, where we found small positive but insignificant effects of net metering policies on residential GHG emissions. However, the introduction of the neighboring policy variable reveals a negative and marginally significant effect (-0.043, $p < 0.10$), suggesting potential positive spillovers from neighboring states' policies.

Table 3.5: Overall ATTs and Signs of the Diffusion Variable across Specifications.

	Log Residential GHG Emissions	
	Baseline Model	CS Estimator
Net Metering Indicator (D_{st})	0.033 (0.036)	-0.153 (0.370)
Number of Neighboring Policies	-0.043* (0.025)	
Observations	1,178	956
State fixed-effects	✓	✓
Year fixed-effects	✓	✓

Notes: Column (1) presents results from the baseline fixed-effects model. Column (2) shows results from the CS estimator. The coefficients of covariates are not provided in the CS output which is why we have a blank space in the bottom right cell of the table. Standard errors are clustered at the state level and are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

The event study results in Figure 3.10 provide a more nuanced view of these effects over time. Comparing these to our previous event study results, we notice some key differences. The TWFE OLS estimates now show a slightly more pronounced negative trend in the post-treatment periods, with larger negative point estimates in periods 2 and 3 after treatment. This suggests that accounting for policy diffusion may reveal stronger emissions reduction effects of net metering policies over time.

Importantly, the pre-treatment trends in both estimators remain relatively flat and close to zero, supporting the parallel trends assumption and the validity of our difference-in-differences approach. These findings suggest that accounting for policy diffusion may be crucial in understanding the full impact of net metering policies on residential GHG emissions. The negative coefficient on the neighboring policy variable in the static model and the more pronounced negative trends in the

event study analyses indicate that the effects of these policies may extend beyond state borders. It is important to note, however, that while these results suggest potentially larger emission reduction effects when accounting for spillovers, the estimates from the CS estimator in particular show wide confidence intervals, indicating considerable uncertainty.

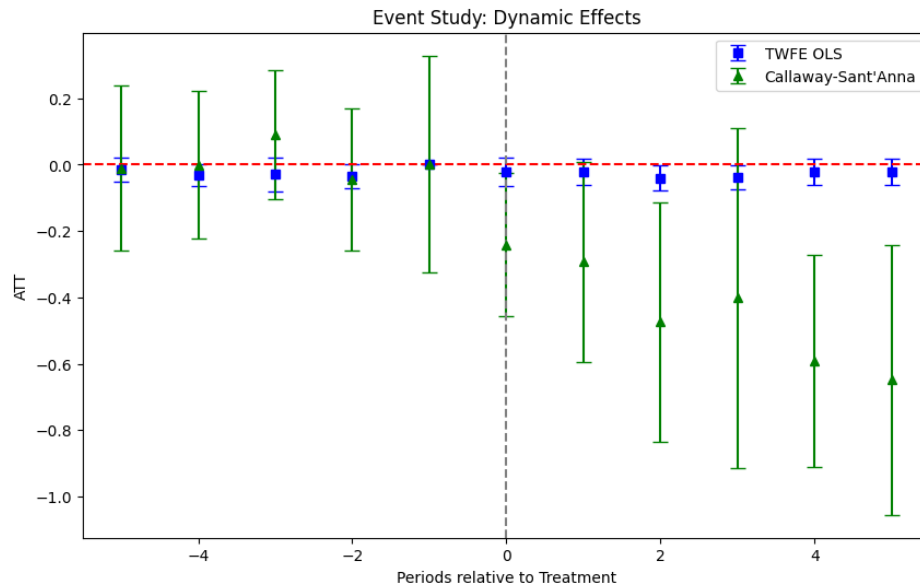


Figure 3.10: Event Study Results Accounting for Policy Diffusion.

Notes: This figure presents event-study plots constructed using two different estimators: a dynamic version of the TWFE model (in blue with square markers) and Callaway & Sant'Anna (2021) (in green with triangle markers). The outcome variable is residential GHG emissions. The horizontal axis represents the periods relative to the net metering policy implementation, with $k = -1$ as the reference period excluded from the plot for the case of TWFE OLS. Each point denotes the average treatment effect (ATT) for the specified period, and the bars represent 95% confidence intervals with standard errors clustered at the state level.

The spillover effect can be attributed to several mechanisms specific to the renewable energy sector and electricity markets, as evidenced by tangible examples across the United States. The adoption of net metering policies in neighboring states creates larger, more integrated markets for solar energy systems and services, as demonstrated in the Northeastern United States where companies like SolarCity and Sunrun expanded their operations across multiple states due to favorable policy environments. This expanded market attracts more firms to the region, particularly larger, more efficient operators. For instance, when Iowa enhanced its net metering policies, solar companies began offering services in neighboring Illinois and Missouri, which created

competitive pressure that improved efficiency and reduced prices. The resulting increase in market participants and competition leads to improved economies of scale, as seen in the Northeast where the concentration of solar companies has enabled bulk purchasing of equipment and shared infrastructure, driving down costs region-wide. Furthermore, successful implementation in adjacent states reduces perceived regulatory risks and thus lower financing costs for renewable energy projects. This is exemplified by New Jersey’s successful net metering implementation, which has increased investor confidence and led financial institutions to offer more favorable loan terms for renewable projects in neighboring states like Pennsylvania, despite their less progressive policies. The interconnected nature of electricity grids also facilitates efficient distribution of excess renewable energy, as evidenced by California’s exports of excess solar generation to Arizona and Nevada through the Western Interconnection grid. Additionally, regional initiatives like New England’s grid modernization efforts demonstrate how states can collaborate to handle increased renewable energy inputs more efficiently. These combined effects enhance the overall effectiveness of net metering policies in reducing residential GHG emissions across state boundaries by stimulating higher adoption rates of residential solar systems and improving regional renewable energy integration.

Placebo Tests.

Another critical technique employed by econometricians to validate the parallel trends assumption is the implementation of placebo tests. These tests serve to bolster the credibility of our DiD estimates by demonstrating that our results are not driven by spurious correlations or pre-existing trends. We focus on placebo period tests, as our use of not-yet-treated states as the control group rule out the possibility to use of placebo group tests.

The placebo period approach involves artificially shifting the treatment period to a time before the actual policy implementation.¹⁸ If our identification strategy is valid, we should observe no significant “treatment effect” during this placebo period. This absence of effect would support the notion that the changes we observe post-treatment are indeed causally related to the net metering

¹⁸Placebo group tests apply the treatment effect to a group that was not actually treated to see if any spurious effects emerge.

policies and not due to pre-existing differential trends between treated and control states. To implement this test, we modify our baseline equation as follows:

$$\log(CO2_{st}) = \alpha_s + \delta_t + \beta D_{st}^p + \mathbf{X}'_{st}\theta + \epsilon_{st} \quad (3.11)$$

where D_{st}^p is a placebo treatment indicator that equals 1 for treated states in each of the five years preceding their actual treatment year, and 0 otherwise.¹⁹ Formally:

$$D_{st}^p = \mathbb{1}(T_s - 5 \leq t < T_s) \quad (3.12)$$

where t is the current year, T_s is the actual year of net metering policy implementation in state s , and $\mathbb{1}(\cdot)$ is the indicator function.²⁰ We estimate this model using both our baseline fixed-effects approach and the CS estimator. If the parallel trends assumption holds, we expect the coefficient β to be statistically indistinguishable from zero.

These placebo tests serve multiple purposes. First, they provide a direct test of the parallel trends assumption in the pre-treatment period. Second, they allow us to assess whether our results are robust to potential anticipation effects or policy endogeneity. Finally, they help rule out the possibility that our estimated treatment effects are driven by other unobserved factors that might be correlated with the timing of net metering policy adoption.

Based on the placebo test results presented in Table 3.6 and Figure 3.11, we can draw the following conclusions. The placebo tests largely support the validity of our main findings. In Table 3.6, we can see that all the placebo treatment indicators are statistically insignificant.

Figure 3.11 provides a more detailed view of the placebo effects over time. Across all specifications - TWFE and CS estimators, with and without policy diffusion - we observe no clear pre-trends. The event study coefficients are predominantly insignificant across all periods and

¹⁹To avoid a constant five-year shift, we do so for a random subset of states that implemented the policy.

²⁰The condition $\mathbb{1}(T_s - 5 \leq t < T_s)$ ensures that the indicator captures the pre-treatment period up to the actual treatment year, without including any post-treatment effects. For example, if $T_s = 2010$, the indicator will be 1 for the years 2005 to 2009, and 0 for all other years.

Table 3.6: Placebo Tests: Overall ATTs across Specifications.

	Log Residential GHG Emissions			
	Without Policy Diffusion		With Policy Diffusion	
	Baseline Model	CS Estimator	Baseline Model	CS Estimator
Placebo Treatment Indicator (D_{st}^p)	0.011 (0.021)	0.040 (0.184)	-0.005 (0.018)	-0.084 (0.174)
Number of Neighboring Policies			-.047* (0.026)	
Observations	1,054	818	1,054	801
State fixed-effects	✓	✓	✓	✓
Year fixed-effects	✓	✓	✓	✓

Notes: Column (1) and (2) present results from the baseline fixed-effects model and the CS estimator, respectively, without accounting for policy diffusion. Columns (3) and (4) present results with accounting for policy diffusion. The coefficients of covariates are not provided in the CS output which is why we have a blank space in the bottom right cell of the table. After applying the time shifter, States with resulting treatment years outside of the study range are dropped from the analysis. Standard errors are clustered at the state level and are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

specifications. This lack of significant effects in the pre-treatment periods reinforces the parallel trends assumption underlying our difference-in-differences approach. The consistency of these insignificant results across different model specifications and estimators adds robustness to our main findings. It suggests that the significant effects we observe in our primary analysis are likely due to the actual impact of net metering policies rather than spurious correlations or pre-existing trends.

3.7.4 Potential Mechanisms

To investigate the mechanisms through which net metering policies influence residential CO_2 emissions, we analyze detailed state-level net metering statistics obtained from the EIA for the years 2011-2020.²¹ It is important to note that these data, not available for years prior to 2011, unfortunately overlap with only a few states that adopted the policy after 2011, preventing us from including more states in the DiD analysis. Our analysis focuses on South Carolina as the treated

²¹The data can be downloaded from the EIA website here: <https://www.eia.gov/electricity/data/eia861m/>

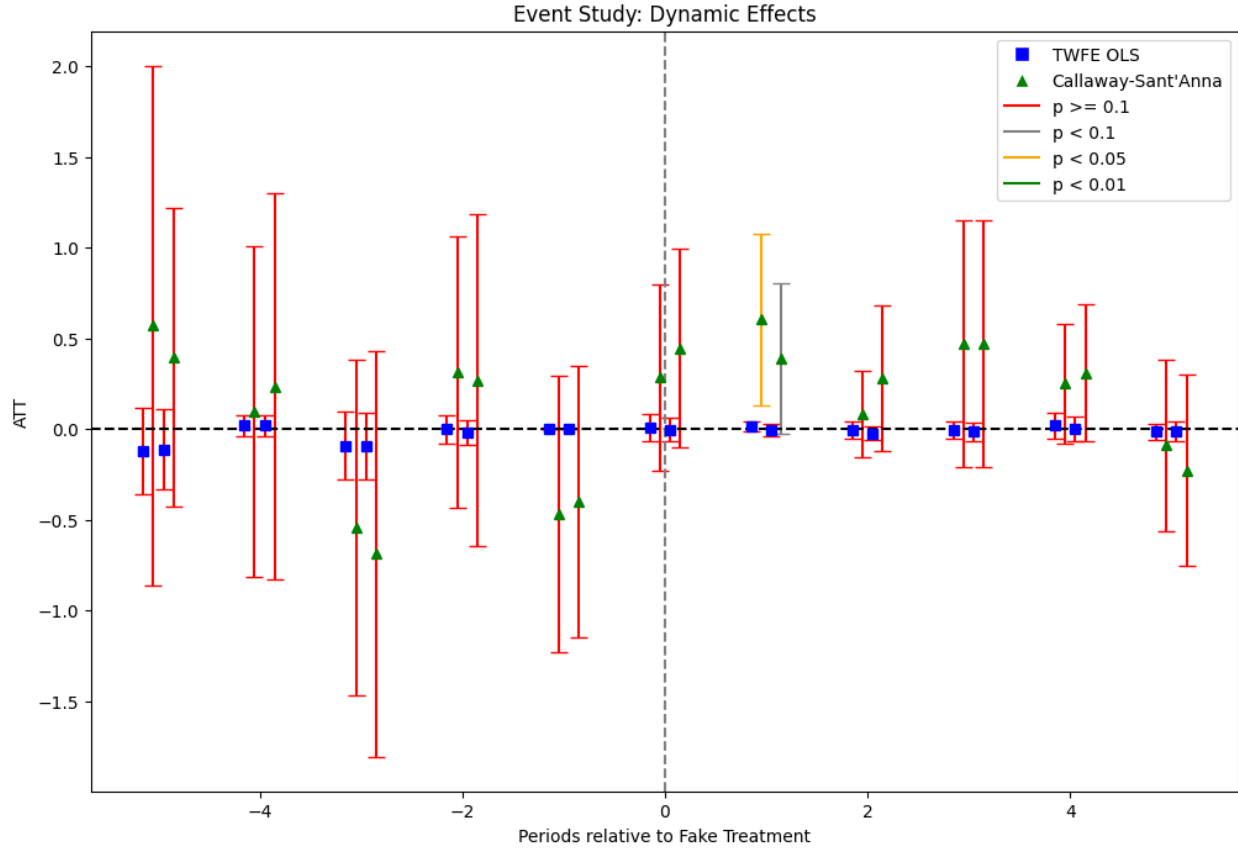


Figure 3.11: Placebo Tests: Event Studies across Specifications.

Notes: This figure presents event-study plots constructed using two different estimators: a dynamic version of the TWFE model (in blue with square markers) and Callaway & Sant'Anna (2021) (in green with triangle markers). The outcome variable is residential GHG emissions. The horizontal axis represents the periods relative to the net metering policy implementation, with $k = -1$ as the reference period excluded from the plot for the case of TWFE OLS. Each point denotes the average treatment effect (ATT) for the specified period, and the bars represent 95% confidence intervals with standard errors clustered at the state level. The negatively jittered markers show cases without policy diffusion, while the positively jittered markers show cases with policy diffusion.

state, which implemented net metering policies during this period, and uses Alabama, South Dakota, and Tennessee as the control group, representing states that did not adopt such policies.²²

We estimate the causal effect of net metering policies on several key intermediate outcomes related to residential renewable energy adoption and utilization. Our specification remains similar to the main analysis, with state and year fixed effects, a treatment indicator, and time-varying state

²²There are barely data available for Mississippi, which adopted the policy in 2015.

characteristics as controls. However, instead of residential GHG emissions, we now use various measures of renewable energy adoption and utilization as outcome variables.

The outcome variables we consider include residential photovoltaic net metering customer count, installed capacity (MW), and energy sold back to the grid (MWh). We also examine similar metrics for wind energy, as well as aggregate measures across all renewable technologies. These outcomes allow us to assess whether net metering policies effectively stimulate the adoption of distributed renewable energy resources and increase grid interaction, which are crucial mechanisms for reducing residential CO_2 emissions. Note that, since we only have one treated state, our control group becomes the *never treated* group.

Given the structure of our data, with South Carolina as the sole treated state, our analysis reduces to a classical DiD framework. This setup allows us to estimate the causal effect of net metering policies on various intermediate outcomes related to renewable energy adoption and utilization. Our empirical specification takes the following form:

$$Y_{st} = \alpha_s + \delta_t + \beta_1 \text{Treat}_s + \beta_2 \text{After}_t + \beta_3 (\text{Treat}_s \times \text{After}_t) + \mathbf{X}'_{st} \theta + \epsilon_{st} \quad (3.13)$$

where Y_{st} represents the log per capita value of our outcome variables for state s in year t , α_s are state fixed effects, δ_t are year fixed effects, Treat_s is an indicator for South Carolina, After_t is an indicator for the post-treatment period, $\mathbf{X}_{st}$ is a vector of time-varying covariates, and ϵ_{st} is the error term. The coefficient of interest, β_3 , captures the average treatment effect of net metering policy implementation on the intermediate outcomes.

The results in Table 3.7 illuminate the key mechanisms driving the reduction in residential CO_2 emissions observed in our main analysis under both unconditional (excluding the covariates) and conditional cases. The intermediate outcomes, particularly those related to solar PV adoption (first three rows), serve as the primary channels through which net metering policies influence residential emissions. It's noteworthy that the results for all technologies closely mirror those of solar PV, which is unsurprising given that solar accounts for over 95% of residential renewable

energy systems across all categories. This dominance of solar PV underscores its pivotal role in shaping the overall impact of net metering policies on residential energy landscapes. The significant increases in solar capacity and energy sold back to the grid indicate that net metering effectively incentivizes residential solar installations and promotes greater grid interaction. These changes in energy production and consumption patterns directly contribute to the reduction in residential CO_2 emissions by displacing fossil fuel-based electricity. The slight discrepancy between unconditional and conditional DiD estimates suggests that while net metering policies are influential, in some cases, the effects become insignificant after adding covariates, which indicates that other factors beyond the policy itself may also play a crucial role in driving adoption and emissions reductions.

The stark contrast between solar and wind results, particularly the large increase in wind energy sold back without corresponding increases in capacity or customers, likely reflects the concentrated nature of wind energy adoption. Unlike solar installations, which are more accessible to average homeowners, wind turbines require significant upfront investment and suitable land, which limits the customer base. However, existing wind energy customers appear to be expanding their operations within their current capacity. This could explain why we observe increased energy sold back to the grid without seeing proportional increases in new capacity installations or customer counts. The pattern suggests that wind energy growth is driven by existing wind energy companies producing more rather than new wind energy firms entering the market.

3.8 Discussions

Our analysis reveals that net metering policies lead to modest but statistically significant reductions in residential GHG emissions. The cumulative effect reaches approximately 0.6% five years after policy adoption, which translates to an average reduction of about 44,640 metric tons of residential CO_2 emissions per state per year by the fifth year.²³ To put this into perspective, this reduction is roughly 1/15th of the energy-related residential emissions for the District of Columbia in 2021,

²³This calculation is based on the means provided in Table 3.2.

Table 3.7: Understanding Mechanisms: DiD Estimates.

Dependent Variable	Unconditional DiD	Conditional DiD
ln (Solar Energy Sold Back) pc	1.957** (0.512) [33]	0.486 (0.646) [33]
ln (Solar Installed Capacity) pc	2.280** (0.629) [34]	1.409* (0.447) [34]
ln (Solar Customers) pc	1.991* (0.643) [34]	1.041 (0.419) [34]
ln (Wind Energy Sold Back) pc	5.199*** (0.030) [22]	4.521* (0.907) [22]
ln (Wind Installed Capacity) pc	0.185 (0.199) [23]	0.454 (0.366) [23]
ln (Wind Customers) pc	0.248 (0.308) [23]	0.486 (0.375) [23]
ln (Renewable Energy Sold Back) pc	2.622** (0.483) [33]	1.584 (0.955) [33]
ln (Renewable Installed Capacity) pc	2.732** (0.489) [34]	2.395* (0.790) [34]
ln (Renewable Customers) pc	2.560** (0.437) [34]	2.136* (0.759) [34]

Notes: This table presents the DiD estimates for various models. All variables represent the residential sector. The estimates represent the coefficient for the interaction term (after_treat). The standard errors are reported in parentheses below the estimates, with the number of observations (N) reported in square brackets. *** p < 0.01, ** p < 0.05, * p < 0.10.

which totaled 0.67 million metric tons.²⁴ This is also equivalent to removing around 9,700 cars off the road in each state, every single year.²⁵ While these reductions may appear small in percentage terms, their aggregate impact across multiple states over time represents a substantial contribution to climate mitigation efforts.

²⁴According to EIA data which can be obtained from: <https://www.eia.gov/environment/emissions/state/>

²⁵According to the EIA, an average passenger car produces approximately 4.6 metric tons of CO₂ annually.

The effectiveness of net metering policies varies considerably across states, highlighting the importance of policy design and local context. States with more favorable policy features based on our scoring system did not necessarily experience larger emissions reductions. This is likely due to the fact that each policy feature has a different weight on potential residential emissions reductions. For instance, a state implementing net metering policies with the two most impactful traits might achieve greater emissions reductions than a state adopting three less influential features. Specific policy elements, including favorable excess electricity compensation, higher system size caps, and customer ownership of renewable energy credits, emerged as primary drivers of these reductions. Surprisingly, we found that Republican-leaning states experienced about the same but statistically significant emissions reductions compared to Democratic-leaning states. This seemingly counterintuitive result likely stems from the higher baseline adoption of renewable energy for electricity generation in Democratic states prior to net metering implementation. With a larger share of their energy mix already coming from clean sources, Democratic states experienced minimal additional reductions in GHG emissions from net metering policies. In contrast, Republican-leaning states, starting from a lower baseline of renewable energy adoption, were able to reduce their GHG emissions at a slightly faster rate through net metering policies, as they had more opportunities for immediate emissions reductions that Democratic states had already captured through earlier policies. Moreover, our evidence of positive spillover effects, where emissions decreased in states neighboring those that adopted net metering policies, underscores the potential for regional policy coordination to amplify benefits. This finding implies that the impact of net metering extends beyond state borders, possibly through mechanisms such as expanded markets for solar energy systems or knowledge sharing across jurisdictions.

Despite these insights, our study has limitations that warrant discussion. Particularly, our reliance on state-level aggregate data may obscure important local or household-level variations. Nonetheless, we believe our results provide valuable insights for understanding the role of net metering policies in reducing residential GHG emissions. The core mechanisms through which

these policies affect emissions – primarily by incentivizing the adoption of residential solar PV systems – are likely to remain relevant in the foreseeable future. Furthermore, technological advancements in solar PV systems and smart grid technologies over the past two decades may have enhanced rather than diminished the impact of net metering policies. Improvements in solar panel efficiency, reductions in installation costs, and the integration of energy storage solutions could potentially make net metering even more effective in reducing residential GHG emissions.

3.9 Conclusions

Our study breaks new ground in the literature on renewable energy policies as the first empirical evaluation of net metering’s impact on residential greenhouse gas emissions in the World. We hope this pioneering work will pave the way for further research into the effectiveness of such policies in combating climate change. Our findings demonstrate that these policies have played a modest but meaningful role in reducing emissions, with effects intensifying over time. The heterogeneity in policy effectiveness across states emphasizes the importance of tailored policy design. Policymakers may focus on enhancing specific features that have demonstrated efficacy, such as generous compensation rates and higher system size caps, while considering the unique contexts of their states, including existing electricity market structures and renewable energy initiatives, when crafting or refining net metering policies. Our analysis also highlights the potential for regional cooperation to amplify the benefits of net metering policies. The observed spillover effects suggest that coordinated approaches across neighboring states could lead to more substantial emissions reductions.

While our study focuses on residential GHG emissions, the implications extend to broader discussions on sustainable energy transitions and climate policy. The evidence that net metering policies facilitate increased adoption of distributed renewable energy resources supports their role in national and state-level strategies to reduce carbon emissions. As states and nations strive to meet ambitious climate goals, understanding the effectiveness of policies like net metering

becomes increasingly critical. Future research could expand on this work by exploring longer-term effects, examining impacts on other sectors, or assessing the dynamics between net metering and other renewable energy incentives. In summary, net metering policies represent a viable tool for reducing residential GHG emissions, but their effectiveness is contingent upon thoughtful design and consideration of local contexts. As the energy landscape continues to evolve, policymakers and researchers must remain attentive to emerging technologies and changing market dynamics to ensure that such policies continue to contribute meaningfully to climate mitigation efforts.

CHAPTER 4

LEVERAGING REMOTE SENSING AND MACHINE LEARNING TO PREDICT U.S. COUNTY-LEVEL RICE YIELD: THE ROLE OF GHG EMISSIONS¹

¹J. Augustin, G. Munisamy, B. Karali & Y. Rao. To be submitted to IEEE.

4.1 Abstract

This study leverages remote sensing and machine learning techniques to predict U.S. county-level rice yields from 2008 to 2022 across 67 counties in six major rice-producing states. The comprehensive dataset integrates satellite-derived vegetation indices and climate variables from NOAA, soil properties from Google Earth Engine, methane emissions data from the Environmental Protection Agency, and rice production statistics from USDA. We leverage various machine learning models, including Explainable Boosting Machines, XGBoost, Ridge Linear Regression, Decision Trees, LASSO, Random Forest, Support Vector Machines, and Convolutional Neural Networks for predictions. XGBoost and EBM emerge as top performers, accurately predicting yields without overfitting. A key finding reveals that while models struggle with out-of-time predictions, they excel at out-of-season forecasts, accurately predicting yields as early as April-June of the growing season. Feature importance analysis highlights soil properties, particularly pH and texture at various depths, as critical predictors for both yield and emissions. This study also uniquely explores yield-emissions trade-offs using the Non-dominated Sorting Genetic Algorithm II (NSGA-II), revealing an unexpected positive correlation between yield improvement and methane emissions reduction.

4.2 Introduction

Rice is arguably the most important staple food crop globally, providing essential calories and nutrition for more than 3 billion people, equivalent to over half of the world's population (Hossain & Fischer, 1995; Fageria, 2007). In Asia alone, rice comprises over 70% of total calorie intake (Bishwajit et al., 2013), highlighting deep connections between rice yields and broader food security challenges, particularly in developing nations (Bandumula, 2018; Maraseni et al., 2018; Bishwajit et al., 2013). For example, Asia accounts for over 87% of total global rice production and 35% of rice exports, dominated by major producers like China, India, Thailand, and Vietnam (Bandumula, 2018). However, efforts to expand rice cultivation are constrained by a combination of threats, including land degradation, urban expansion on agricultural areas, and the mounting pressures of global climate change manifested through shifts in temperature, rainfall, and extreme weather events (Maraseni et al., 2018). Sustainably predicting and maintaining stable rice yields in the face of these challenges remains an urgent priority with implications spanning from household food security to global commodity markets.

At the same time, flooded rice paddies are also a major source of greenhouse gas (GHG) emissions, accounting for upwards of 11% of total emissions from the agricultural sector (Maraseni et al., 2018). This predominantly reflects substantial methane generation under the anaerobic soil conditions associated with continuous flooding (Qian et al., 2023). For example, Chinese rice paddies alone are estimated to contribute nearly 40% of the nation's total agricultural methane emissions (Li et al., 2006). While offering certain agronomic advantages, paddy flooding therefore represents a tradeoff between production levels and environmental impact. As such, optimized water management practices like mid-season drainage or intermittent flooding have potential to reduce methane emissions from rice cultivation by over 50% (Qian et al., 2023; Li et al., 2006). However, realizing these kinds of mitigation opportunities requires developing accurate models, calibrated to

regional or county-level scales, that can predict fluctuations in both rice yield and associated GHG dynamics under different climate and management scenarios moving forward.

The U.S. rice industry is concentrated primarily across six rice-producing states—Arkansas, California, Louisiana, Mississippi, Missouri, and Texas—which collectively harvested nearly 10.73 million tons of rice in 2021 (USDA Economic Research Service, 2024). This was cultivated over 2.8 million acres, mostly on small family farms that each contribute around \$1 million to local economies annually (USARice, 2024). In total, the U.S. rice industry generates over \$34 billion for the national economy and directly supports over 125,000 jobs across farming, processing, and distribution activities (USARice, 2024). The United States is uniquely positioned as the world's fifth largest rice exporter, shipping 5% of global trade, despite accounting for less than 2% of total production (USDA Economic Research Service, 2024). Domestically, Arkansas leads production of long grain varieties while California specializes in medium and short grain rice. Moreover, the United States achieves some of the highest rice yields globally through intensive irrigation and advanced genetics (USDA Economic Research Service, 2024).

The importance of accurate rice yield forecasting extends beyond simple production estimates. It plays a crucial role in informing agricultural policies, guiding resource allocation, and ensuring food security. However, current forecasting methods often struggle to capture the complex interaction of factors affecting rice yields, particularly in the face of changing climate conditions and the need for sustainable farming practices. This study proposes a novel approach to address these challenges by combining three key elements: comprehensive analysis of environmental data over space and time, interpretable machine learning techniques, and methods to quantify prediction uncertainty. By integrating these components, we aim to enhance our ability to forecast rice yields accurately while also considering the environmental impact of rice production, particularly methane emissions.

This study focuses on the U.S. rice industry for several key reasons. First, the concentrated nature of production across six states provides a manageable yet diverse study area which allows for the analysis of varying climatic and soil conditions. Second, the U.S. rice industry's high

yields and advanced production techniques offer an opportunity to study best practices that could potentially be applied globally. Third, as the fifth major rice-exporting country in the world as of 2024 (Statista, 2024), understanding U.S. rice production dynamics has implications for global food security and trade. Lastly, the availability of detailed, county-level data on rice production, climate, and soil characteristics in the U.S. enables a robust analysis of the factors influencing rice yields and associated greenhouse gas emissions.

According to the Environmental Protection Agency, the agriculture sector accounted for 10% of total U.S. GHG emissions in 2021 (Environmental Protection Agency, 2023). Major sources from agriculture include nitrous oxide emissions from soil management activities like fertilizer application, methane from enteric fermentation in livestock, and methane and nitrous oxide emissions from manure management. While not broken out separately, rice cultivation contributes to agriculture emissions through methane released from flooded paddies. Overall, emissions from agriculture increased 7% from 1990 to 2021, highlighting the need for optimized management practices to enhance productivity while mitigating climate impacts.

Current approaches to crop yield prediction often rely on traditional statistical methods or complex machine learning models. Statistical methods, such as multiple linear regression or time series analysis, offer simplicity and interpretability but may fail to capture complex, non-linear relationships in agricultural systems. On the other hand, advanced machine learning techniques like neural networks or random forests can model intricate patterns in data, potentially improving prediction accuracy. However, these “black-box” models often lack transparency, making it difficult for agricultural practitioners to understand and trust the underlying decision-making process (Hu et al., 2023). The limitations of these existing methods have led to a growing interest in more interpretable machine learning approaches.

In this context, the so-called “glass-box” machine learning methods, such as explainable boosting machines (EBMs), have potential to elucidate the hidden relationships learned between various environmental predictors and target crop yield outcomes (Celik et al., 2023). Rather than

functioning as black boxes, techniques such as EBM can identify and highlight the relative influence of different variables, from rainfall to soil moisture, in determining yield estimates. Such model interpretability and transparency is critical for generating actionable insights to guide breeding efforts and optimized management strategies related to both rice production levels and associated GHG emissions mitigation. However, transparent glass box methods have not yet been extensively applied or validated in agricultural modeling contexts. Developing reliable approaches to predict county-level rice yields, while accounting for spatiotemporal fluctuations in climate and advocating for model explicability, is therefore an essential innovation to support regional food security planning and greenhouse gas regulation.

While remote sensing and machine learning methods show promise for rice yield modeling, major knowledge gaps persist in applying these techniques across different rice growing regions of the United States. For example, the majority of previous efforts developing advanced yield prediction models have focused on major cereal crops like maize and soybeans grown across the Midwest Corn Belt (Huntington et al., 2020). Comparatively fewer studies have explored integration of climate, soil, and remote sensing data specifically for rice yield forecasting concentrated in production areas like California, Arkansas, Louisiana, and Mississippi (Espe et al., 2016).

Additionally, while mid-season drainage presents a promising avenue for mitigating methane emissions from flooded paddies, broad scale implementation constraints exist due to limited water control infrastructure, variability in farmer access to actionable information, and lack of technical guidance on optimized drainage timing and duration (Qian et al., 2023). Most prior assessments of drainage practices for methane regulation have occurred at relatively small experimental scales. Therefore generating generalized insights into balancing county-level rice yields and methane mitigation requires spatially extensive modeling, validated across heterogeneous cultivated landscapes (Li et al., 2006). Furthermore, drainage and other water management decisions cannot occur in isolation—rather they intersect closely with cultivar selection, planting schedules, fertilizer application, and overall intensification strategies which collectively determine yields

(Fageria, 2007; Qian et al., 2023). Disentangling these complex agro-environmental relationships places a premium on interpretable machine learning which illuminates the relative influence of different variables.

Rice agriculture in the United States is deeply intertwined with economic policies, market dynamics, and the livelihoods of farming communities concentrated across Southern and Central California, Arkansas, Louisiana, Mississippi, and Missouri (Espe et al., 2016). While biophysical crop-climate modeling helps characterize yield potentials and constraints, translating these insights into tangible production outcomes requires integrating social dimensions like agricultural policies, commodity prices, consumer demand, and farmer decision-making (Baker, 2004). This encompasses factors ranging from crop insurance subsidies, to efficiency of extension services, to strength of market incentives rewarding sustainable intensification over extensification (Espe et al., 2016).

This research aims to advance the field of data-driven rice yield modeling in the United States, with a particular focus on integrating environmental factors and emissions. The primary goals are: (1) to develop and validate machine learning models for predicting county-level rice yields across major U.S. rice-growing regions using a combination of remote sensing, climate, and soil data; (2) to assess the spatial and temporal robustness of these predictive models; (3) to identify and quantify the key environmental drivers influencing rice yields through interpretable machine learning techniques; and (4) to explore the potential trade-offs between maximizing rice yields and minimizing rice-derived methane emissions.

This study makes multiple key contributions to the literature. First, to the best of our knowledge, it is the first application focusing on U.S. county-level rice yield prediction by harnessing remote sensing and machine learning methods. While extensive research has explored these techniques for major cereal crops across the Midwest, rice has received comparatively little attention despite strong linkages to domestic agricultural policies and global food security through export markets. Second, the methodology blends state-of-the-art spatiotemporal analytics with transparent, interpretable machine learning techniques to not only forecast yields but also elucidate

the environmental drivers and quantify predictive uncertainties. The integration of satellite mapping of crop health proxies with auxiliary climate and soil datasets is demonstrated across the main rice-growing states. Finally, the analysis uniquely assesses tradeoffs around maximizing production versus minimizing greenhouse gas emissions, especially methane, which comprises a disproportionate share of the rice agricultural footprint. While studies have appraised aspects like water management practices for emissions control in Asian contexts, this study is the first targeting a major exporter in the Western hemisphere.

Our findings reveal several key insights. Explainable Boosting Machine emerged as a powerful glass-box model for rice yield prediction. The models demonstrate strong spatial robustness and accurately capture yield patterns across different counties. Temporal analysis indicates that while the models perform well within growing seasons, their ability to generalize to future years remains limited, which emphasizes the need for regular model updates. Feature importance analysis highlights soil properties, particularly pH and texture at various depths, as critical factors that influence rice yields. Water-related variables, vegetation indices, and climate factors also play significant roles. Importantly, our models can provide accurate yield predictions as early as the beginning of the growing season (April-June), which offers valuable insights for farmers' decision-making. Finally, we uncovered a positive correlation between yield improvement and emissions reduction, which suggests that practices that enhance productivity also contribute to lower methane emissions.

The rest of the paper is structured as follows: Section 3 provides a comprehensive literature review, Section 4 details the data used in the study, Section 5 outlines the methods employed, Section 6 presents the results in depth, Section 7 discusses the implications of our findings, and Section 8 concludes with policy implications.

4.3 Literature Review

A rapidly accelerating body of literature has explored integration of satellite remote sensing and machine learning techniques for crop yield modeling and forecasting. This review synthesizes recent research in this interdisciplinary domain, with particular emphasis on efforts targeting rice yield prediction across different geographical contexts spanning major production regions from Asia to the United States. A seminal study by Sun et al. (2019) proposed a combined convolutional and recurrent neural network architecture for soybean yield forecasting at the U.S. county level. Their model takes as input time series of gridded climate data, Moderate Resolution Imaging Spectroradiometer (MODIS) land surface temperature, and MODIS surface reflectance converted to histogram-based tensors. These remote sensing and meteorological datasets are synthesized using the cloud-based Google Earth Engine platform to demonstrate scalable analytics. When evaluated against standalone Long-Short-Term Memory (LSTM) and Convolutional Neural Networks (CNN) benchmarks, the CNN-LSTM hybrid demonstrated superior accuracy for both end-of-season and crucially within-season predictions. The authors note the potential to extend their deep learning fusion approach to other major crops like maize and wheat as well as finer spatial scales given sufficient training data. One limitation is the reliance on black-box neural networks which fail to provide intrinsic explanations.

In a related analysis focused on winter wheat in the United States, Feng et al. (2021) employed a spatiotemporally weighted neural network (GTWNN) to explicitly account for non-stationarity in yield response across both space and time. Their GTWNN integrates satellite indices and meteorological data similar to Sun et al. (2019) but uses an explicit geotemporal weighting kernel in place of the generic deep network architecture. Side-by-side comparisons against support vector regression, time-series neural networks, and geotemporal regression verified the advantages of the proposed special-purpose architecture for this crop. However, a lack of model transparency persists

as a limitation, motivating future work to open the black box through methods like SHAP (SHapley Additive exPlanations) value analysis or representative feature extraction from inner layers.

Beyond the U.S. context, multiple recent studies have tailored machine learning models specifically for rice yield prediction spanning top production zones in Asia. For instance, Cao et al. (2021) developed county-level random forest, LSTM, and LASSO regression models to forecast yields across all rice growing regions of China using vegetation indices, climate, and soil datasets synthesized within the Google Earth Engine cloud platform. The LSTM convincingly outperformed the alternate machine learning approaches, demonstrating potential for national scale yield monitoring leveraging only publicly available remote sensing resources. However, the role of notorious methane emissions stemming from Chinese rice paddies was not directly assessed, representing an area for further integration with climate and agricultural emissions modeling.

In analogous efforts demonstrating the portability of integrated yield forecasting pipelines, Park et al. (2018) combined satellite-derived NDVI measures from the Moderate Resolution Imaging Spectroradiometer (MODIS) with an artificial neural network model incorporating spatial interpolation of climate inputs to predict rice yields and climate impacts in South Korea at high resolution. Meanwhile Ji et al. (2007) demonstrated the superiority of neural networks over linear regression alternatives for forecasting yields specifically in mountainous terrain across numerous sites in China's Fujian province. However, both studies were limited to demonstration over regional analysis domains.

Beyond Asia, remote sensing and crop modeling have also been combined for yield prediction and benchmarking in major rice exporters like the United States. As an example, Espe et al. (2016) employed the ORYZA model to estimate rice yield potential across environments encompassing the majority of total U.S. rice area. The significance of elevated nighttime temperatures and cold shock events for simulating yields in California rice systems emerged as a key insight, underscoring the value of process-based approaches effective in new geographical areas. However, such crop

models require extensive parameterization and calibration. Meanwhile statistical methods offer greater flexibility but faced limitations extrapolating locally derived relationships.

In terms of underlying methodology for yield prediction, while deep neural networks and tree-ensembles have shown immense promise, they are often criticized as opaque “black box” systems. Very recently, glass box methods like EBM have been proposed as an interpretable alternative (Celik et al., 2023). Transparent and trustworthy modeling approaches like EBM, which have shown promise in cotton yield prediction using multisource data, are prime candidates for application in rice forecasting and other agricultural contexts, offering the potential to unpack complex relationships and provide interpretable insights.

Finally, zooming to the global scale, multiple analyses point to rice agriculture as a pivotal nexus balancing productivity gains, emissions mitigation imperatives, and food security for hundreds of millions of vulnerable smallholder producers and consumers across monsoon Asia (Maraseni et al., 2018; Qian et al., 2023).² For instance, Qian et al. (2023) estimate that optimized water management via intermittent flooding could reduce methane emissions from Chinese rice paddies by over 50% while largely maintaining yields and output. However, uncertainties and assumptions remain regarding both in situ emissions baselines as well as ideal optimized implementation schemes. How findings would translate to alternate nations in the region further compounds the challenges. Therefore improved quantification using data resources like satellite mapping could prove invaluable.

While advancements have been made in forecasting accuracy and spatiotemporal modeling, challenges remain in model interpretability, emissions considerations, and cross-regional applicability. We address these gaps by combining interpretable machine learning techniques with comprehensive spatiotemporal data to predict both rice yields and associated methane emissions in the U.S. context. Our approach aims to balance productivity, environmental sustainability, and food security concerns

²Monsoon Asia encompasses Asian countries that experience substantial seasonal rainfall and dry periods due to the monsoon winds. This climatic pattern is essential for agricultural practices, especially rice farming, and significantly influences the socioeconomic conditions of the region.

and thus contributes to the development of more comprehensive and actionable rice yield prediction models.

4.4 Data

4.4.1 Study Area

In the United States, rice cultivation is concentrated in several key regions, each with unique agro-ecological conditions and management practices that influence productivity and sustainability outcomes. The study focuses on six main rice-growing states: Arkansas (AR), Louisiana (LA), Missouri (MO), Mississippi (MS), Texas (TX), and California (CA). As shown in Figure 4.1, these states encompass the four primary rice-producing regions in the country: the Arkansas Grand Prairie, the Mississippi Delta (parts of Arkansas, Mississippi, Missouri, and Louisiana), the Gulf Coast (Texas and Southwest Louisiana), and the Sacramento Valley of California.

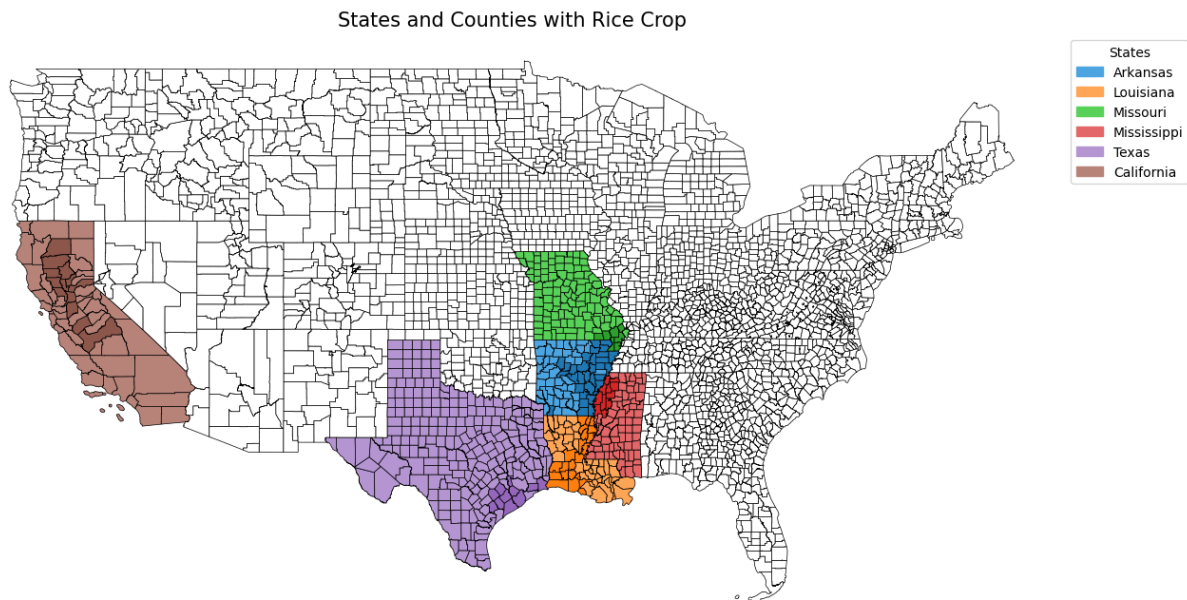


Figure 4.1: States and Counties (Darker Areas) where Rice is Grown.

Each of these regions specializes in specific rice variety, generally categorized by grain length (long, medium, or short).³ Long-grain rice accounts for approximately 75% of U.S. rice production and is primarily grown in the southern states. Medium-grain rice makes up about 24% of the total production and is mainly cultivated in California and Arkansas. Short-grain rice, almost exclusively grown in California, represents the remaining 1% of U.S. rice production (USDA Economic Research Service, 2024).

In 2021, the United States produced 10.73 million tons of rough rice, slightly above the 2019 crop but down 16% from 2020. The 2022 rice crop further declined to 8.98 million tons, largely due to drought conditions in California (USDA Economic Research Service, 2024). These fluctuations in production highlight the importance of understanding the spatiotemporal drivers of rice yields and developing robust prediction models to support sustainable production efforts. The darker areas in Figure 4.1 represent the specific counties involved in rice cultivation within each state. Initially, 94 counties were considered for the study; however, 27 were dropped due to lack of data across variables, resulting in a final set of 67 counties.

4.4.2 Data Collection and Processing

The dataset spans from 2008 to 2022 and includes climate variables, soil properties, rice production data, satellite-derived vegetation indices, and rice-derived methane emissions. We collect data from various sources and process them to align with the county-level spatial resolution and monthly temporal resolution. Table 4.1 provides an overview of the data sources and variables used in this study.

The climate variables are obtained from the National Oceanic and Atmospheric Administration (NOAA).⁴ These include average temperature (tavg), maximum temperature (tmax), minimum temperature (tmin), precipitation (pcp), cooling degree days (cdd), heating degree days (hdd),

³In our models, we do not include rice type because most of the yield data reports published by USDA do not specify the type of rice.

⁴NOAA climate data link: <https://www.ncei.noaa.gov/access/monitoring/climate-at-a-glance/county/time-series>

Table 4.1: Summary of Data Sources and Variables.

Variable	Source	Content
Climate Variables	NOAA	average temperature, maximum temperature, minimum temperature, precipitation, cooling degree days, heating degree days, palmer drought severity index, palmer hydrological drought index, palmer modified drought index, palmer z-index
Soil Properties	GEE	carbon content, water content, pH, texture
Rice Data	USDA	yield, production, area planted, area harvested
Satellite Data	NOAA	normalized difference vegetation index, leaf area index, fraction of absorbed photosynthetically active radiation
Emissions Data	(Maasakkers et al., 2023)	CH_4

Notes: NOAA: National Oceanic and Atmospheric Administration; GEE: Google Earth Engine; USDA: United States Department of Agriculture.

Palmer Drought Severity Index (pdsi), Palmer Hydrological Drought Index (phdi), Palmer Modified Drought Index (pmdi), and Palmer Z-Index (zndx). The data are programmatically downloaded at the county and monthly level through the NOAA Application Programming Interface (API).

Soil properties, including carbon content,⁵ water content, pH, and texture, are obtained from the Google Earth Engine (GEE) platform⁶ using the GEE API. The soil pH data (Hengl, 2018a), soil water content data (Hengl & Gupta, 2019), and soil texture classes data (Hengl, 2018b) are collected at 250 m resolution for six standard depths (0, 10, 30, 60, 100, and 200 cm). As these soil properties do not change significantly in the short term, the data is reported at the yearly level.

Rice production data, including yield, production, area planted, and area harvested, are acquired from the United States Department of Agriculture (USDA) through their API.⁷ These data are available at the yearly level. Satellite-derived vegetation indices, namely NDVI (Vermote,

⁵This variable is later dropped due to data sparsity.

⁶Google Earth Engine data link: <https://developers.google.com/earth-engine/datasets/catalog>

⁷USDA QuickStats data link: <https://quickstats.nass.usda.gov/>

2019b), Leaf Area Index (LAI), and Fraction of Absorbed Photosynthetically Active Radiation (FAPAR) (Vermote, 2019a), are obtained from the NOAA Terrestrial Climate Data Records.⁸ The data are available at the daily level, but we calculate and use the mean at the monthly level given that there are many days throughout a particular month with poor data quality due to adverse weather conditions. The satellite data are stored on Amazon S3 and was accessed using the S3 API.

Methane emissions data, specifically rice-related methane emissions, are obtained from the gridded EPA U.S. methane greenhouse gas inventory (Maasakkers et al., 2023).⁹ The dataset, available from 2012 to 2018, includes annual methane emission maps with $0.1^\circ \times 0.1^\circ$ spatial resolution (approximately 10 x 10 km) and monthly temporal resolution for the contiguous United States (CONUS). To calculate the monthly emissions for rice cultivation, we utilize the conversion factor provided by Maasakkers et al. (2023).¹⁰ The conversion factor transforms the monthly scale factors into total emissions, taking into account the spatial resolution, number of days in each month, and the molar mass of methane.

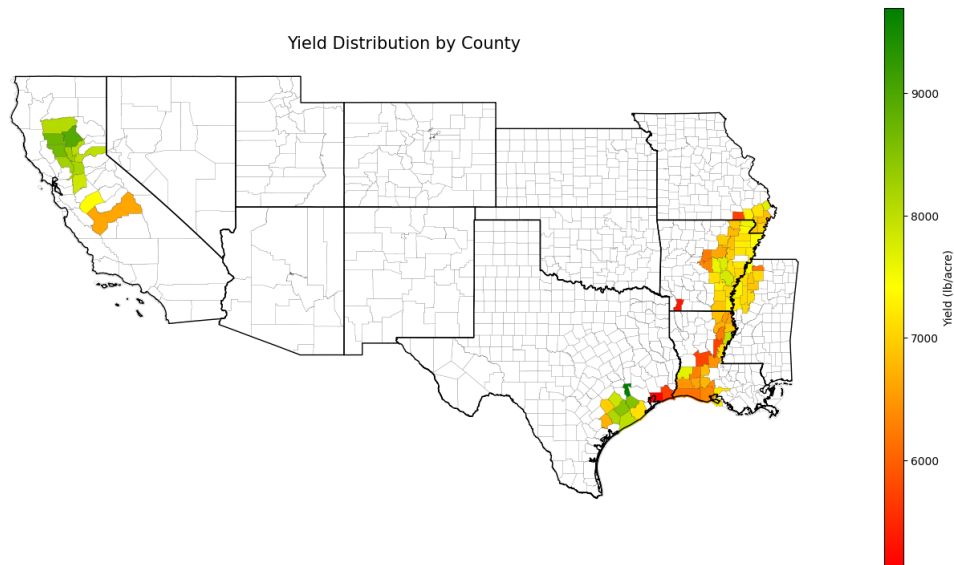
To obtain county-level rice emissions, we first scale the methane emissions data to match the spatial resolution of the rice grid cells derived from the USDA Cropland Data Layers (USDA National Agricultural Statistics Service, 2024). We then perform a spatial intersection between the scaled methane emissions data and the rice grid cells to identify the emissions specifically associated with rice cultivation areas. Finally, we aggregate the emissions of rice fields within each county to obtain total county-level rice emissions. Figure 4.2 shows the distribution of average yield and rice-derived methane emissions by county. Counties in California seem to have the lowest rice-related emissions despite having the highest rice yields. This decoupling of high productivity from high emissions can be attributed to several factors, including strict environmental regulations, adoption

⁸NOAA terrestrial climate data records: <https://www.ncei.noaa.gov/products/climate-data-records/terrestrial>

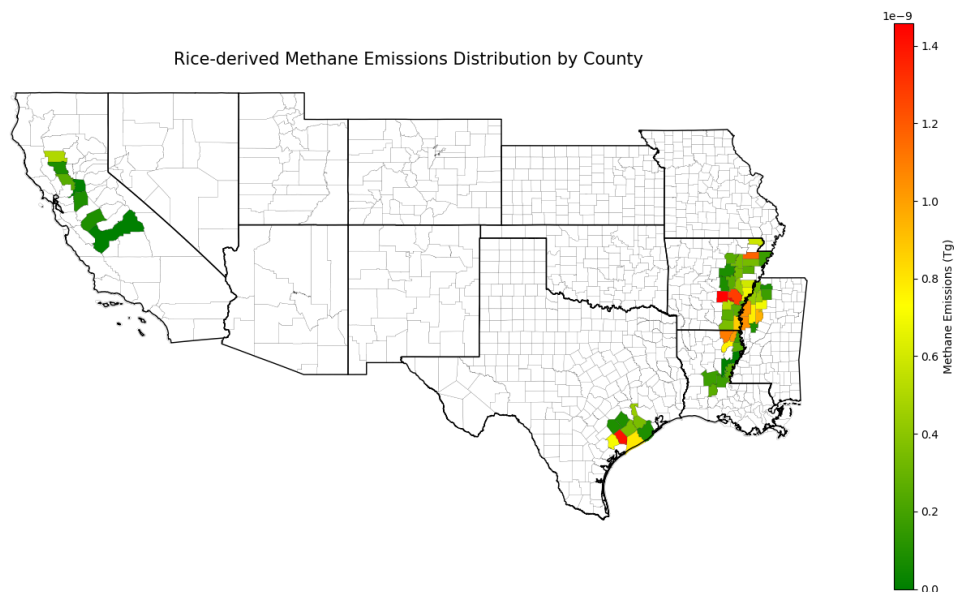
⁹Gridded EPA U.S. Methane Greenhouse Gas Inventory: <https://zenodo.org/records/8367082>

¹⁰The conversion factor was calculated using the following constants: Avogadro's number (6.022×10^{23} molecules per mole), molar mass of methane (16.04×10^{-12} Tg per mole), and the number of seconds in a day (86400). The grid cell area was determined based on the latitude and the $0.1^\circ \times 0.1^\circ$ spatial resolution. For each grid cell and time point, the number of days in the month was considered to account for leap years. The resulting emissions in Tg were obtained by multiplying the monthly scale factors by the calculated conversion factor and the grid cell area.

of water-efficient farming techniques such as alternate wetting and drying, and implementation of methane-reducing practices like dry seeding and residue (i.e., rice straw) management. While this pattern isn't universal across all counties, it provides evidence that improving yields while reducing environmental impact is achievable with appropriate agricultural practices and policy frameworks.



(a) Average Rice Yield by County.



(b) Average Methane Emissions by County.

Figure 4.2: Comparison of Average Yield and Rice-based Methane Emissions by County.

To ensure spatial consistency and accuracy, the satellite and soil property data are also processed using the USDA Cropland Data Layers shapefiles to identify the specific pixel-level grids where rice is grown in each county. The average values of these grids are then calculated to represent the county-level averages. Additionally, the methane emissions data are available for a shorter period (2012-2018) compared to the other variables (2008-2022). Consequently, models that include emissions data have eight fewer years of data available for analysis. This results in 3,555 observations for the longer dataset and 1,710 observations for the shorter dataset, both across 67 counties.

This sample size and time span are comparable to those used in similar studies predicting crop yields in the U.S. context. For instance, Feng et al. (2021) employed data from 2008 to 2019 for winter wheat yield prediction at the county level, while Celik et al. (2023) utilized data from 2017 to 2021 for cotton yield estimation across 214 counties. Furthermore, it is common in crop yield prediction studies to use variables at different temporal frequencies. For example, our study incorporates daily, monthly, and yearly data, similar to the approach taken by Sun et al. (2019), Celik et al. (2023), and Feng et al. (2021), who also integrated data at various temporal resolutions.¹¹

Table 4.2 presents the summary statistics for the variables used in this study, along with their respective units and spatial resolutions. The satellite-derived vegetation indices, NDVI, LAI, and FAPAR, have mean values of 0.38, 1.23, and 0.43, respectively, and are measured at a 0.05° spatial resolution. Climate variables such as average temperature, maximum temperature, and minimum temperature are measured in degrees Fahrenheit (°F), with mean values of 74.99, 85.77, and 64.18, respectively. Precipitation has a mean value of 4.13 inches, while cooling degree days and heating degree days have mean values of 332.41 and 25.79 degree days, respectively. The drought indices have mean values close to zero, indicating relatively normal conditions on average.

¹¹For the variables with daily and monthly frequencies, we only include data for the months of the growing season (from planting until harvest) for a particular state. This information can be found here: <https://www.ers.usda.gov/topics/crops/rice/rice-sector-at-a-glance/>

Rice production variables, including area harvested, area planted, production, and yield, exhibit substantial variability across the counties. The mean area harvested and area planted are 38,243.84 and 40,984.75 acres, respectively. The mean production is 158,683.784 tons, while the mean yield is 7,228.99 pounds per acre (lb/acre). We report summary statistics of soil properties are reported only at 30 cm depth for the sake of space. The spatial resolution for the soil properties is 250 meters. The table shows considerable variation for the levels of pH and water content across the study area. The mean soil pH in water is 60.40, and the mean soil water content is 24.23%.

Table 4.2: Summary Statistics of Variables with Units and Resolution.

Variable	Mean	Std	Min	Max	Units
NDVI	0.38	0.13	0.13	0.82	index: 0.05°
Average temperature	74.99	7.27	49.90	88.10	°F
Maximum temperature	85.77	7.15	62.00	99.90	°F
Minimum temperature	64.18	8.05	37.90	78.70	°F
Precipitation	4.13	2.87	0.00	17.29	inches
Cooling degree days	332.41	179.59	0.00	715.00	degree days
Heating degree days	25.79	54.68	0.00	467.00	degree days
Palmer drought severity index	0.07	2.43	-8.97	6.04	index
Palmer hydrological drought index	0.24	2.53	-8.97	6.12	index
Z-index	0.20	2.06	-6.16	8.82	index
Area harvested	38243.84	33881.12	232.00	163000.00	acres
Area planted	40984.75	34357.68	600.00	164000.00	acres
Production	158683.78	148193.64	663.82	778400	tons
Yield	7228.99	884.89	3620.00	10600.00	lb/acre
Soil pH in water (30cm)	60.40	6.65	50.00	78.40	pHw : 250 meters
Soil water Content (30cm)	24.23	1.92	19.33	32.53	%: 250 meters
LAI	1.23	0.60	0.00	3.55	index: 0.05°
FAPAR	0.43	0.14	0.00	0.75	index: 0.05°

Notes: We use crop layers to compute averages at the county for the climate, soil, and satellite variables. Rice-related variables are at the county level.

4.5 Methods

Prior to conducting the primary analysis, we perform a series of diagnostic tests on the dataset to assess data quality, identify potential anomalies, and make necessary adjustments. In the following

subsection, we delineate the diagnostic tests performed. Readers may refer to the appendix for the complete interpretation of the results of these diagnostic tests. We leverage the PiML Python library (Sudjianto et al., 2023) to conduct some of these analyses.

4.5.1 Diagnostic Tests

For each feature, we visualize the distribution using appropriate graph-based techniques such as histograms for numerical variables and frequency-based bar plots for categorical variables. For the bivariate analysis, we compute the Pearson’s correlation coefficient to quantify pairwise linear relationships.

Outliers can have a substantial impact on the analysis and performance of ML models. To identify potential outliers in the numeric features, we employ a boxplot-based approach. Each boxplot provides a visual representation of the median, interquartile range (IQR), and whiskers extending to the most extreme data points within 1.5 times the IQR from the box edges. Data points falling outside the whiskers are considered potential outliers.

Data drift refers to the phenomenon where the statistical properties of the data change over time. Assessing data drift is crucial to ensure that the distributions of the features in the testing dataset are not fundamentally different than those of the training dataset. We focus on marginal distribution drift, which assesses the distance between the marginal distributions of each feature in the training and testing sets. The Population Stability Index (PSI) is used to quantify the difference between two probability distributions. A feature with a significantly higher PSI value compared to other features indicates a greater degree of data drift for that specific feature. This implies that the marginal distribution of the feature has changed substantially between the training and testing sets, potentially leading to reduced model performance and generalization ability.

To perform the data drift test, we conduct 100 random 5-fold splits of the dataset using the K-Fold cross-validation technique. For each split, we calculate the PSI between the training and testing sets for each feature. The PSI between two distributions P and Q is defined as:

$$\text{PSI}(P, Q) = \sum_{i=1}^B (p_i - q_i) \ln \frac{p_i}{q_i} \quad (4.1)$$

where B represents the number of bins, and p_i and q_i denote the proportions of the two samples in each bin. After completing all the splits and repeats, we calculate the average PSI for each feature across all instances.

4.5.2 Spatiotemporal Analysis

The spatiotemporal analysis aims to disentangle complex relationships between rice yields and various environmental factors, including climate variables, soil properties, and satellite-derived indices over space and time. To investigate the spatial correlations between rice yields and environmental factors, we calculate Pearson's correlation coefficient for each county. Let Y_c denote the rice yield for county c , and $X_{c,v}$ represent the v -th environmental factor for county c . The correlation coefficient for county c 's rice yield and environmental factor v , $\rho_{c,v}$, is calculated as:

$$\rho_{c,v} = \frac{\sum_{t=1}^T (Y_{c,t} - \bar{Y}_c)(X_{c,t,v} - \bar{X}_{c,v})}{\sqrt{\sum_{t=1}^T (Y_{c,t} - \bar{Y}_c)^2} \sqrt{\sum_{t=1}^T (X_{c,t,v} - \bar{X}_{c,v})^2}} \quad (4.2)$$

where $Y_{c,t}$ and $X_{c,t,v}$ are the rice yield and the v -th environmental factor for county c at year t , and $\bar{Y}_c$ and $\bar{X}_{c,v}$ are their mean values over the study period T , respectively.

To investigate the temporal trends of the interaction of each feature with rice yields, we calculate Pearson's correlation coefficient between rice yields and environmental factors for each county and each month. We then average these across all counties to obtain the overall correlation between rice yields and each environmental factor for each month of the growing season. This allows us to capture the temporal patterns of correlation. The correlation coefficient for rice yields in county c , environmental factor v for month m is denoted as $\rho_{c,v,m}$ and is calculated as:

$$\rho_{c,v,m} = \frac{\sum_{t=1}^T (Y_{c,t} - \bar{Y}_c)(X_{c,t,v,m} - \bar{X}_{c,v,m})}{\sqrt{\sum_{t=1}^T (Y_{c,t} - \bar{Y}_c)^2} \sqrt{\sum_{t=1}^T (X_{c,t,v,m} - \bar{X}_{c,v,m})^2}} \quad (4.3)$$

where $Y_{c,t}$ and $X_{c,t,v,m}$ are the rice yields and the v -th environmental factor for county c in month m at year t , and $\bar{Y}_c$ and $\bar{X}_{c,v,m}$ are the mean values of rice yields and the v -th environmental factor for county c in month m over the study period T , respectively.

The spatiotemporal correlations between rice yields and environmental factors are visualized using maps for spatial correlations and time series boxplots for temporal correlations. The spatial correlation maps show the average correlation for each county, while the temporal boxplots display the distribution of correlations across different months to highlight the variability and trends over time.

4.5.3 Yield Analysis

The yield analysis focuses on developing and comparing various machine learning models for predicting rice yields at the county level. Let $\mathbf{X}_c = [X_{c,1}, X_{c,2}, \dots, X_{c,p}]$ denote the input features for county c , where p is the number of features, and let Y_c represent the corresponding rice yield. The goal of the yield analysis is to learn a function f that maps the input features $\mathbf{X}_c$ to the yield Y_c , i.e., $f(\mathbf{X}_c) = Y_c$. As presented in the data section, our dataset has a temporal mismatch: Y_c is at the annual level, while $\mathbf{X}_c$ contains monthly and yearly data.¹² In the context of agricultural yield prediction studies, it is common to use higher-frequency input data to predict lower-frequency outcomes, as machine learning models can handle high-dimensional data and identify complex relationships. For instance, Sun et al. (2019) and Feng et al. (2021) also utilized monthly environmental data to predict annual crop yields in their studies on soybean and winter wheat, respectively. One of the drawbacks of this is the introduction of noise which can lead to overfitting. To assess the performance of the

¹²We aggregate the three variables that are at the daily level (NDVI, LAI, and FAPAR) to the monthly level to avoid data redundancy since the majority of the features are at the monthly level.

models used and ensure the robustness of our results, we employ a 5-fold cross-validation approach as mentioned earlier.¹³

Feature Engineering

To capture the complex relationships between environmental factors and rice yields, we perform feature engineering by creating interaction terms. We include interactions between the variable month and the climate variables, as well as the satellite and soil variables. Additionally, we create interaction terms between the maximum temperature and all other features except for the county code.¹⁴ These interaction terms allow the models to capture the time-varying (seasonality) and temperature-dependent effects of these factors on rice yields.

From an agronomic perspective, including these interaction terms is crucial for capturing the stage-specific effects of environmental factors on rice growth and yield formation. Rice plants have specific growth stages, each with different requirements for water, temperature, and solar radiation. The impact of these factors on rice yields can vary significantly depending on the growth stage at which they occur. For example, during the vegetative phase, rice plants require adequate water and nutrients for proper growth and development, while during the reproductive phase, they are more sensitive to temperature extremes. By including interactions with month and maximum temperature, the models can capture these stage-specific effects and provide a more accurate representation of the complex relationships between environmental conditions, crop growth, and yield formation throughout the growing season.

To handle the wide ranges of feature values, we also apply min-max scaling to all numeric variables. Min-max scaling transforms the features to a common scale between 0 and 1, which can improve the convergence and performance of many machine learning algorithms. The min-max scaling formula is given by:

¹³We report cross-validated metrics throughout the results to ensure that our results are not due to random luck.

¹⁴We create a unique identifier for each county by concatenating the two-digit state FIPS code to the respective three-letter county FIPS code. This is important to ensure that counties in different states with the same name are not treated as one.

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (4.4)$$

where X is the original feature value, X_{min} and X_{max} are the minimum and maximum values of the feature, respectively, and X_{scaled} is the scaled feature value.

Model Development and Evaluation

For glass-box models, we employ EBM, Ridge Linear Regression, Decision Trees, and LASSO. The EBM model (Nori et al., 2019), an additive model that learns feature contributions while maintaining interpretability, is compared to other glass-box models such as Ridge Linear Regression, which mitigates overfitting through L2 regularization, Decision Trees that recursively partition the feature space based on the most informative features, and LASSO, which performs feature selection via L1 regularization.

Among the black-box models, CNN are employed to capture spatial dependencies in the input features, while ensemble methods like Random Forest and Extreme Gradient Boosting (XGBoost) are used to combine multiple decision trees and sequentially add weak learners to minimize residual errors, respectively. Support Vector Machines (SVM), a kernel-based method that finds the optimal hyperplane to separate the feature space into different classes, is also included in the comparison.

Hyperparameter tuning is performed using GridSearch to find the best combination of hyperparameters for each model. The models are evaluated using various metrics, including mean absolute error (MAE), root mean squared error (RMSE), and R-squared (R^2), which are calculated using cross-validation. These metrics are defined as follows:

$$MAE = \frac{1}{n} \sum_{c=1}^n |Y_c - \hat{Y}_c| \quad (4.5)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{c=1}^n (Y_c - \hat{Y}_c)^2} \quad (4.6)$$

$$R^2 = 1 - \frac{\sum_{c=1}^n (Y_c - \hat{Y}_c)^2}{\sum_{c=1}^n (Y_c - \bar{Y})^2} \quad (4.7)$$

where $\hat{Y}_c$ is the predicted yield for county c , and $\bar{Y}$ is the mean yield of the training set.

Feature importance is assessed using SHAP (SHapley Additive exPlanations) values (Lundberg & Lee, 2017), which provide a unified framework for interpreting the output of any machine learning model by computing the contribution of each feature to the final prediction. The SHAP value for feature v and county (i.e., observation) c is given by:

$$\phi_{c,v} = \sum_{S \subseteq F \setminus \{v\}} \frac{|S|!(p - |S| - 1)!}{p!} [f_S(\mathbf{X}_{c,S \cup \{v\}}) - f_S(\mathbf{X}_{c,S})] \quad (4.8)$$

where F is the set of all features, S is a subset of features, and f_S is the model trained on the subset S .¹⁵

Robustness Analyses

Spatial robustness is evaluated using maps that compare actual and predicted rice yields across different counties or regions. Any substantial discrepancies between the predicted and actual yield values would be evident in these maps. We expect the models to perform well for a majority of the spatial units, with similar patterns emerging between the actual and predicted yield maps. However, some variability is anticipated, particularly in regions with unique micro-climates, soil conditions, or management practices that may not be fully captured by the input features.

¹⁵For ML algorithms such as EBM that do not support SHAP, we use their respective built-in functions to obtain feature importance.

Temporal robustness is evaluated using an out-of-time validation approach, where the models are trained using data from earlier years (e.g., 2008-2018) and tested on later years (e.g., 2019-2022). The performance metrics are computed for each year to assess the model’s ability to generalize over time. Finally, the optimal timing for yield prediction is determined by comparing the model performance when using data from the beginning of the season (e.g., April-June) versus the late season (e.g., July-September). The earliest point in the growing season when yield predictions can be made with acceptable accuracy is identified.

4.5.4 Yield-Emissions Trade-off Analysis

While maximizing rice yields is a primary goal for rice farmers, it is increasingly important to consider the environmental impact of rice cultivation practices, particularly methane emissions. We explore the trade-offs between rice yields and rice-derived methane emissions using multi-objective optimization techniques. By understanding these trade-offs, we aim to identify strategies that enable rice farmers to maintain high yields while minimizing the methane footprint of their rice production systems. We use the EBM model developed in the previous section for yield and emissions prediction. These models serve as the foundation for our multi-objective optimization approach. As mentioned before, we have fewer years for this analysis. As a result, we retrain the models using the same set of features for predicting yield and emissions.¹⁶ The multi-objective optimization problem is formulated as follows:

$$\max_x [f_y(x), -f_e(x)] \quad (4.9)$$

where $f_y(x)$ represents the predicted yield $\hat{y}$ and $f_e(x)$ represents the predicted emissions $\hat{e}$, both obtained from the EBM models. The goal is to maximize yield while minimizing emissions, subject to constraints on the input variables x , defined by their feasible ranges in the dataset.

¹⁶Since we need data alignment for the multi-objective function, the yield model needs to be retrained given that we have emissions data for a shorter timeframe. We report the performance metrics for both models in the results section, along with the sorted important features.

We utilize the Non-dominated Sorting Genetic Algorithm II (NSGA-II) of Deb et al. (2002) to solve this multi-objective optimization problem. NSGA-II is an evolutionary algorithm that efficiently generates a set of Pareto-optimal solutions, representing the trade-offs between objectives. The algorithm operates on a population of candidate solutions, evolving them over generations using genetic operators such as selection, crossover, and mutation.¹⁷

In our implementation, we configure NSGA-II with the following settings: a population size of 100 solutions, 100 new solutions created in each generation, random sampling for the initial population, simulated binary crossover with a probability of 0.9 and a distribution index of 15, polynomial mutation with a distribution index of 20, and duplicate elimination enabled to ensure diversity in the population.¹⁸ This configuration allows NSGA-II to explore a wide range of potential solutions while maintaining diversity and efficiently converging towards the Pareto-optimal front. To explore a wide range of trade-off scenarios, we vary the weights assigned to the yield and emissions objectives using a linear space from 0 to 1 with 100 evenly spaced points. For each weight combination, the objectives are combined into a weighted sum as follows:

$$f_{weighted}(x) = w_y \cdot (-f_y(x)) + w_e \cdot f_e(x) \quad (4.10)$$

where w_y represents the weight for yield and w_e represents the weight for emissions, with $w_y + w_e = 1$.

The Pareto-optimal solutions obtained from the multi-objective optimization are visualized using a comprehensive trade-off plot. We then look at the trade-off plot to identify the “sweet spot” that balances the objectives of maximizing rice yield and minimizing methane emissions. Once the optimal trade-off point is determined, we extract the corresponding values for each input variable

¹⁷This algorithm was initially developed in the biology field but has been widely used in other fields such as engineering, economics, and computer science, given its ability to efficiently solve complex multi-objective optimization problems and find a set of Pareto-optimal solutions that represent the best trade-offs between conflicting objectives.

¹⁸These NSGA-II settings balance exploration (searching widely in the solution space) and exploitation (refining promising solutions): (1) Population size of 100 balances diversity and computational cost. (2) 100 new solutions per generation ensures full population replacement. (3) Random initial sampling promotes diversity. (4) Simulated binary crossover (probability 0.9, distribution index 15) encourages exploration while focusing on local exploitation. (5) Polynomial mutation (distribution index 20) allows for fine-tuning (i.e., small adjustments to gene values). (6) Duplicate elimination maintains population diversity.

from the associated Pareto-optimal solution. This optimal trade-off point represents a specific combination of yield and emissions weights that provides the most desirable outcome in terms of both agricultural productivity and environmental sustainability.

4.6 Results

4.6.1 Diagnostic Tests

The extended results of the diagnostic tests are presented in the appendix. Climate variables show relatively symmetric distributions, with temperature-related variables approximating normal distributions. Soil pH variables display positively skewed distributions, indicating a prevalence of lower pH values across the study area. Rice production and area-related variables exhibit positive skewness, suggesting the presence of a few large-scale production sites among predominantly smaller cultivation areas. Strong positive correlations exist among temperature-related variables and cooling degree days, while precipitation shows moderate negative correlations with these variables. Soil texture strongly correlates with water content and moderately with soil pH, reflecting the influence of soil composition on water retention and chemical properties. Satellite-derived vegetation indices show positive correlations with each other and moderate positive correlations with temperature-related variables. Outlier detection reveals some extreme values, particularly in the heating degree days variable, which we subsequently remove from the analysis. Data drift tests identify significant changes in the distributions of average temperature, precipitation, and production variables over time, while soil texture features remain relatively stable. These diagnostic results inform subsequent data preprocessing steps and model selection.

4.6.2 Spatiotemporal Analysis

Spatial Analysis.

For brevity, we focus on a subset of key variables representing different categories: NDVI, maximum temperature, precipitation, palmer drought severity index, production, soil pH, soil water content and soil texture class at 60 cm depth. Figure 4.3 shows the spatiotemporal subplot for each of the aforementioned variables. At the spatial level, NDVI shows a fairly consistent positive association with yields across most counties, indicating that healthier vegetation tends to support higher productivity. However, the strength of this relationship does vary, with the strongest correlations found in certain counties of California, Texas, and Arkansas, while a few counties in Arkansas and Mississippi show little to no correlation.

Temperature effects on yields are less spatially consistent but are negative overall, suggesting that higher maximum temperatures tend to slightly reduce rice yields in most counties. The strongest negative effects are seen in parts of Louisiana and Texas, while some counties in Mississippi and California show neutral to slightly positive temperature impacts. Precipitation and yield correlations vary a lot geographically, spanning from moderately negative to moderately positive across the study region. The most beneficial precipitation effects are found in some counties in Texas, northern Louisiana and southeast Arkansas, while most counties in California and Arkansas see the most negative impacts of increasing rainfall.

The palmer drought severity index shows generally positive associations with yields, confirming the importance of moisture availability. Correlations are strongest in counties across Louisiana, Arkansas, and Mississippi. It is interesting to see that drought has a particularly damaging effect on rice yield across all the counties in California. The relationship between production and yield varies dramatically between counties, from very strongly negative to very strongly positive. Most counties in Texas and Louisiana exhibit the closest coupling of yields and production, while most counties across Arkansas and Mississippi show strongly opposing trends.

Soil characteristics also show highly variable effects on yields. Soil pH correlations span from strongly negative to strongly positive, with the most positive associations in counties in California. The strongest negative pH-yield correlations are found in parts of Arkansas and Mississippi. Soil

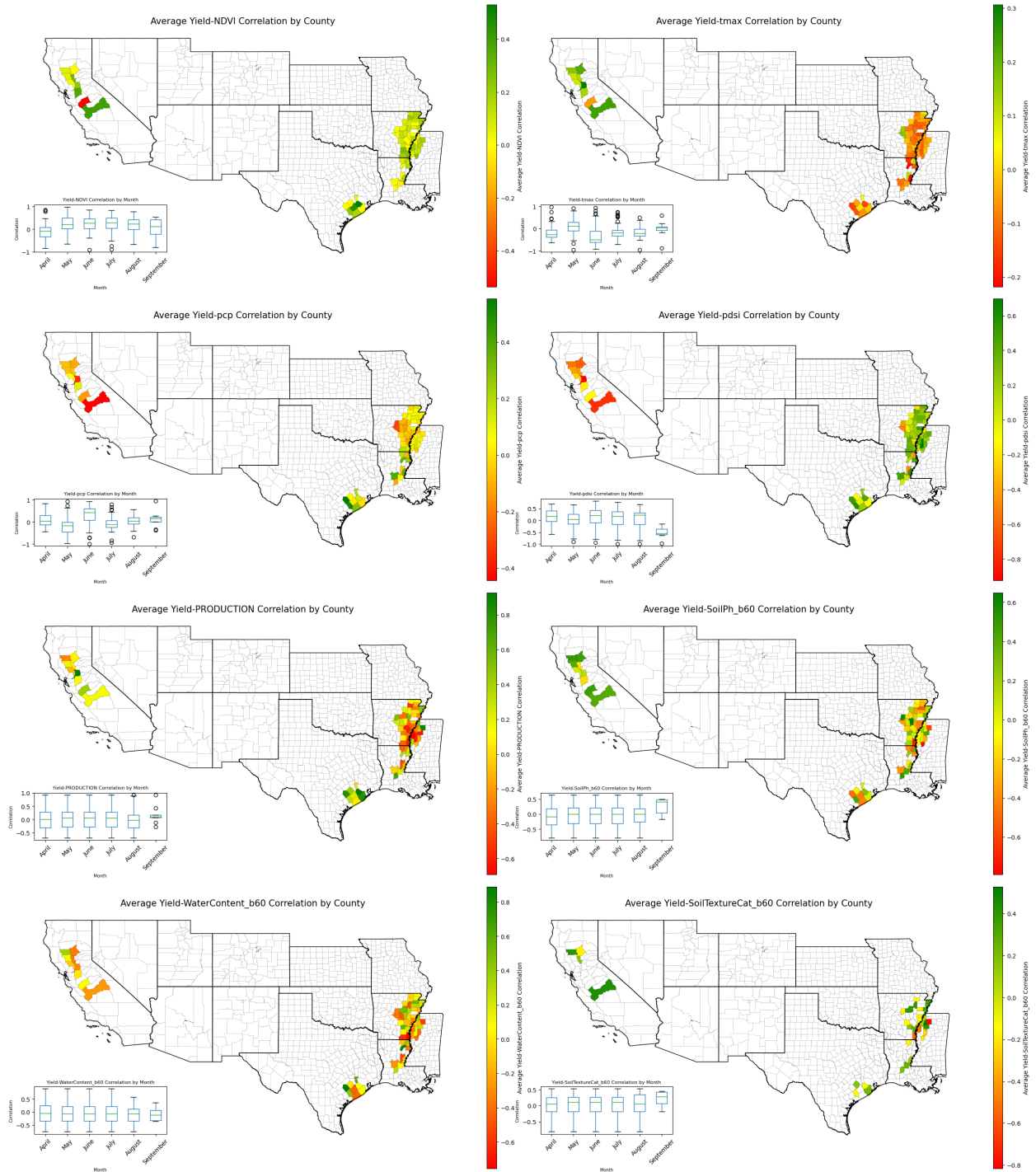


Figure 4.3: Panel of Spatiotemporal Plots.

water content also shows a wide range of yield relationships, from moderately negative to moderately positive. The most favorable soil moisture conditions for yields are found in a handful of counties in Arkansas and Texas, while the strongest negative associations occur in some counties across Arkansas, Louisiana, and Mississippi.

The soil texture variable exhibits generally negative correlations with yield. The correlation coefficients for this variable are undefined in counties where there is limited variation in either soil texture class or yield values over the study period. This occurs when a county has the same soil texture class across all years, resulting in a standard deviation of zero for the texture variable. When converting this feature into numbers, higher numbers are assigned to better soils.¹⁹ The best soil types for rice are generally those that can retain water well, such as loam, clay loam, and silty clay loam. A negative correlation between yield and soil texture class at 60 cm depth in this context indicates that rice yields are lower in counties with better soil textures for rice cultivation. This could be due to factors such as improper water management, soil nutrient imbalances, or other environmental constraints that prevent the realization of potential yield benefits of these soils for rice.

Temporal Analysis.

At the temporal level, the correlations between rice yields and environmental factors show distinctive patterns and transitions from the early to late growing season (April-August). For NDVI, most counties exhibit a trend of increasing positive correlations with yields as the season progresses from April to July/August. For example, in Arkansas, correlations in counties switch from negative in April to moderately positive by July. Similar patterns are observed in Louisiana and Mississippi. This suggests that higher vegetation vigor becomes increasingly crucial for determining yields in the later stages of the season. Maximum temperature correlations show a contrasting temporal pattern in many counties. In Arkansas, Louisiana, and Mississippi, correlations are often more negative in

¹⁹The mapping used for soil texture classes is as follows: Sandy Clay (SaCl) is assigned the lowest value, followed by Silty Clay (SiCl), Silty Loam (SiLo), Sandy Clay Loam (SaClLo), Silty Clay Loam (SiClLo), Clay Loam (ClLo), and Loam (Lo) being assigned the highest value.

the early months (April-May) but become less negative or even slightly positive in the later months (June-August). This could indicate that higher temperatures are more detrimental to yields in the sensitive early growth stages, while the crop may develop some tolerance later in the season.

Precipitation-yield correlations also display a temporal shift in some counties. In Arkansas and Louisiana, correlations are positive in April but turn negative in the later months, suggesting that while early season rainfall is beneficial, excessive precipitation during the later stages could adversely affect yields. However, this pattern is not consistent across all counties. PDSI-yield correlations generally remain positive throughout the season in most counties, indicating the persistent importance of adequate moisture supply. However, the strength of the relationship varies, with some counties showing higher correlations in the middle of the season in Louisiana and Mississippi.

Production-yield correlations appear to be relatively stable throughout the season for most counties. This suggests that the overall seasonal production is more strongly related to yields than monthly variations. Soil pH and yield correlations do not show a clear temporal pattern across counties. The influence of pH on yields likely depends more on local soil conditions and management practices. Soil water content and yield correlations are generally consistent in sign throughout the season for most counties, reflecting the importance of soil moisture. However, some counties in Arkansas show more negative correlations later in the season, possibly due to the adverse effects of water saturation. Soil texture and yield correlations also do not display a strong temporal trend. The impact of soil texture on yields is likely more spatially dependent and related to the inherent properties of the soil.

This temporal analysis reveals that the influence of environmental factors on rice yields can vary throughout the growing season. Factors like NDVI and temperature show the most pronounced temporal patterns, with the direction and strength of correlations shifting from the early to late season months in many counties. Other variables, such as PDSI, production, and soil characteristics, exhibit more temporally consistent relationships with yields.

4.6.3 Yield Analysis

Model Results

In this subsection, we compare the performance of various machine learning models for predicting rice yields at the county level under different interaction settings.²⁰ Table 4.3 presents the five-fold cross-validated scores across the glass-box and black-box models for both training and testing sets. This way, we can detect potential overfitting or underfitting in the models. A significant discrepancy between the training and testing performance is a sign of overfitting, where the model fits the training data too closely and fails to generalize well to unseen data. On the other hand, if both the training and testing performances are poor, it indicates underfitting, where the model is unable to capture the underlying patterns in the data.

Overall, the black-box models outperform the glass-box models in terms of both training and testing metrics, regardless of the interaction settings. Among the black-box models, XGBoost and Random Forest consistently achieve the highest performance, with XGBoost slightly outperforming Random Forest in most cases. The CNN and SVM, despite being powerful black-box models, do not perform as well as XGBoost and Random Forest in this study. Interestingly, the inclusion of interaction terms does not lead to a consistent improvement in model performance across all models. In fact, for many models, including XGBoost and EBM, the best performance is achieved without any interaction terms. This suggests that the base features alone are sufficient to capture the underlying patterns in the data, and the additional complexity introduced by the interaction terms may not be necessary for these models. Among the glass-box models, EBM exhibits the best performance, particularly when no interaction terms are included. The EBM model strikes a balance between interpretability and predictive power, making it a valuable tool for understanding the factors influencing rice yields. Notably, LASSO performs quite poorly in contrast with the other

²⁰To handle the computational complexity of training the ML models across all the different settings and GridSearch parameter space, we leverage Tensor Processing Units (TPUs) available on Google Colab. This significantly reduces the time for model processing.

Table 4.3: Comparison of Model Performance Across Different Settings.

Model	Interactions	MAE		RMSE		R ²	
		Training	Testing	Training	Testing	Training	Testing
Panel A: <i>glass-box models</i>							
EBM	No	0.037	0.050	0.049	0.066	0.885	0.789
Ridge Linear Regression	No	0.079	0.080	0.105	0.106	0.463	0.451
Decision Tree	No	0.035	0.051	0.053	0.083	0.862	0.662
LASSO	No	0.097	0.097	0.128	0.128	0.211	0.211
EBM	Just tmax	0.036	0.057	0.047	0.075	0.893	0.724
Ridge Linear Regression	Just tmax	0.077	0.079	0.103	0.105	0.483	0.463
Decision Tree	Just tmax	0.034	0.053	0.053	0.086	0.865	0.645
LASSO	Just tmax	0.097	0.097	0.128	0.128	0.211	0.211
EBM	Just month	0.038	0.057	0.049	0.075	0.881	0.723
Ridge Linear Regression	Just month	0.078	0.079	0.104	0.106	0.481	0.459
Decision Tree	Just month	0.002	0.041	0.008	0.084	0.995	0.657
LASSO	Just month	0.097	0.097	0.128	0.128	0.211	0.211
EBM	Both	0.030	0.054	0.039	0.072	0.925	0.748
Ridge Linear Regression	Both	0.059	0.061	0.083	0.086	0.663	0.641
Decision Tree	Both	0.036	0.058	0.053	0.089	0.862	0.616
LASSO	Both	0.097	0.097	0.128	0.128	0.211	0.211
Panel B: <i>black-box models</i>							
Random Forest	No	0.013	0.034	0.019	0.051	0.982	0.874
Extreme Gradient Boosting	No	0.001	0.029	0.002	0.044	1.000	0.906
Support Vector Machines	No	0.046	0.060	0.071	0.085	0.755	0.647
Convolutional Neural Network	No	0.070	0.073	0.093	0.098	0.582	0.535
Random Forest	Just tmax	0.015	0.039	0.022	0.058	0.977	0.834
Extreme Gradient Boosting	Just tmax	0.011	0.035	0.015	0.049	0.989	0.883
Support Vector Machines	Just tmax	0.046	0.061	0.072	0.087	0.752	0.632
Convolutional Neural Network	Just tmax	0.072	0.075	0.096	0.101	0.554	0.509
Random Forest	Just month	0.014	0.039	0.021	0.057	0.978	0.840
Extreme Gradient Boosting	Just month	0.012	0.034	0.016	0.048	0.988	0.887
Support Vector Machines	Just month	0.065	0.071	0.083	0.093	0.670	0.579
Convolutional Neural Network	Just month	0.073	0.077	0.097	0.103	0.543	0.489
Random Forest	Both	0.015	0.041	0.022	0.059	0.976	0.829
Extreme Gradient Boosting	Both	0.008	0.034	0.011	0.048	0.994	0.887
Support Vector Machines	Both	0.018	0.039	0.036	0.062	0.938	0.815
Convolutional Neural Network	Both	0.061	0.066	0.083	0.090	0.664	0.603

Notes: Five-fold cross-validated scores are reported for all metrics. Panel A contains glass-box models, while Panel B contains black-box models. The GridSearch function was used for hyperparameter tuning, and the models were trained on the best subset of hyperparameters.

glass-box models, indicating that the L1 regularization may be too aggressive in this case, leading to underfitting.

We select two candidate models for subsequent analysis based on their strong performance and consistency across training and testing sets, as well as their interpretability and predictive power. We consider one glass-box model and one black-box model to provide a balance between interpretability and accuracy. For the glass-box model, we select EBM without any interaction terms. This model achieves good performance on both the training set (MAE = 0.037, RMSE = 0.049, $R^2 = 0.885$) and the testing set (MAE = 0.050, RMSE = 0.066, $R^2 = 0.789$). The relatively small difference between the training and testing scores indicates that the model generalizes well to unseen data without severe overfitting. Moreover, the EBM model offers the advantage of interpretability, as it provides insights into the individual feature contributions and their relationships with the target variable.

For the black-box model, we choose XGBoost without any interaction terms, which consistently outperforms other models across different interaction settings, demonstrating its robustness and predictive power. Without interactions, XGBoost achieves impressive performance on both the training set (MAE = 0.001, RMSE = 0.002, $R^2 = 1.000$) and the testing set (MAE = 0.029, RMSE = 0.044, $R^2 = 0.906$). The relatively small difference between the training and testing scores indicates that the model effectively captures the underlying patterns in the data without overfitting. Although XGBoost is a black-box model and lacks the interpretability of glass-box models, it offers superior predictive performance, making it a valuable tool for accurate yield predictions.

Feature Importance

Figure 4.4 presents the feature importance for both the EBM and XGBoost models. This analysis provides valuable insights into the key factors influencing rice yields and offers guidance for optimizing agricultural practices.

Soil properties emerge as the most critical factors for rice yield prediction in both models. Soil pH at various depths (0-200cm) consistently ranks among the top features, indicating the crucial role of soil acidity in rice production. This suggests that maintaining optimal soil pH levels throughout the soil profile through appropriate liming or acidification practices could significantly impact yield outcomes. Additionally, soil texture categories show high importance, particularly at deeper layers (200cm), highlighting the relevance of subsoil characteristics in determining rice yields, possibly due to their influence on water retention and nutrient availability. Farmers and agronomists should consider deep soil properties when making management decisions or selecting suitable rice varieties.

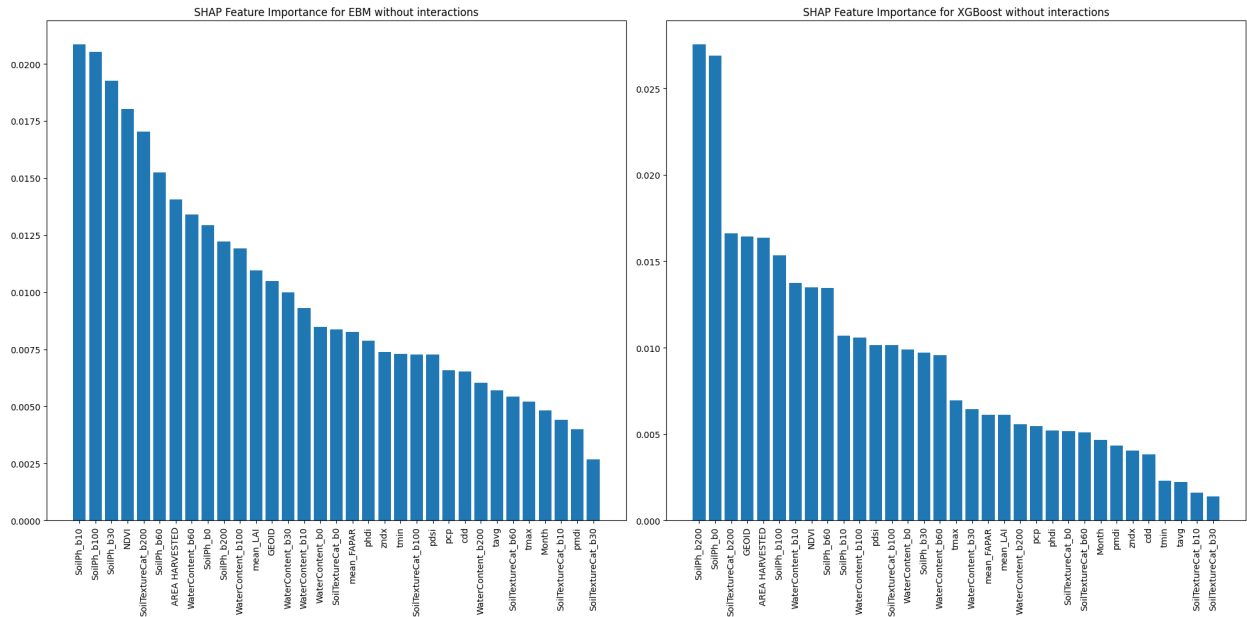


Figure 4.4: Feature Importance for the EBM and XGBoost Models.

Notes: tavg: average temperature; tmax: maximum temperature; tmin: minimum temperature; pcp: precipitation; cdd: cooling degree days; hdd: heating degree days; pdsi: Palmer Drought Severity Index; phdi: Palmer Hydrological Drought Index; pmdi: Palmer Modified Drought Index; zndx: Z-index.

Water-related factors, including soil water content at different depths and precipitation (pcp), are also prominent in both models. This highlights the importance of water management in rice cultivation, emphasizing the need for efficient irrigation strategies and water conservation practices to ensure optimal moisture availability for crop growth. Vegetation indices derived from satellite

data, such as NDVI, LAI, and FAPAR, play significant roles in yield prediction. These indices reflect crop health and photosynthetic activity, suggesting that monitoring and maintaining optimal crop growth throughout the season is crucial for maximizing yields.

Climate variables, including temperature metrics (t_{max} , t_{min} , t_{avg}) and drought indices ($pdsi$, $phdi$, $zndx$), show moderate to low importance in both models. While these factors influence rice yields, their impact may be less direct or more complex than soil and water-related variables. Nonetheless, considering these factors in crop management and variety selection decisions remains important for optimizing yields. Interestingly, the county identifier and *area harvested* variables also appear as important features, particularly in the XGBoost model. This suggests that location-specific factors and the scale of cultivation play roles in determining yields, highlighting the potential benefits of tailoring management practices to specific geographic contexts and considering economies of scale in rice production. Lastly, the *month* variable shows relatively low importance in both models, indicating that the timing within the growing season may be less critical than other factors in determining overall yield outcomes.

Robustness Checks

Spatial Robustness

To assess the spatial robustness of the EBM and XGBoost models, we compare the predicted yields with the actual yields at the county level. Figure 4.5 presents a visual comparison of the actual yields and the predicted yields from both models across the study area. The maps reveal a high degree of spatial consistency between the actual and predicted yields for both models. The EBM and XGBoost predictions closely match the spatial patterns observed in the actual yield map, with higher yields predominantly concentrated in the northern counties of California and Arkansas, as well as the eastern counties of Texas. Lower yields are generally predicted and observed in the southern counties of Louisiana and Mississippi.

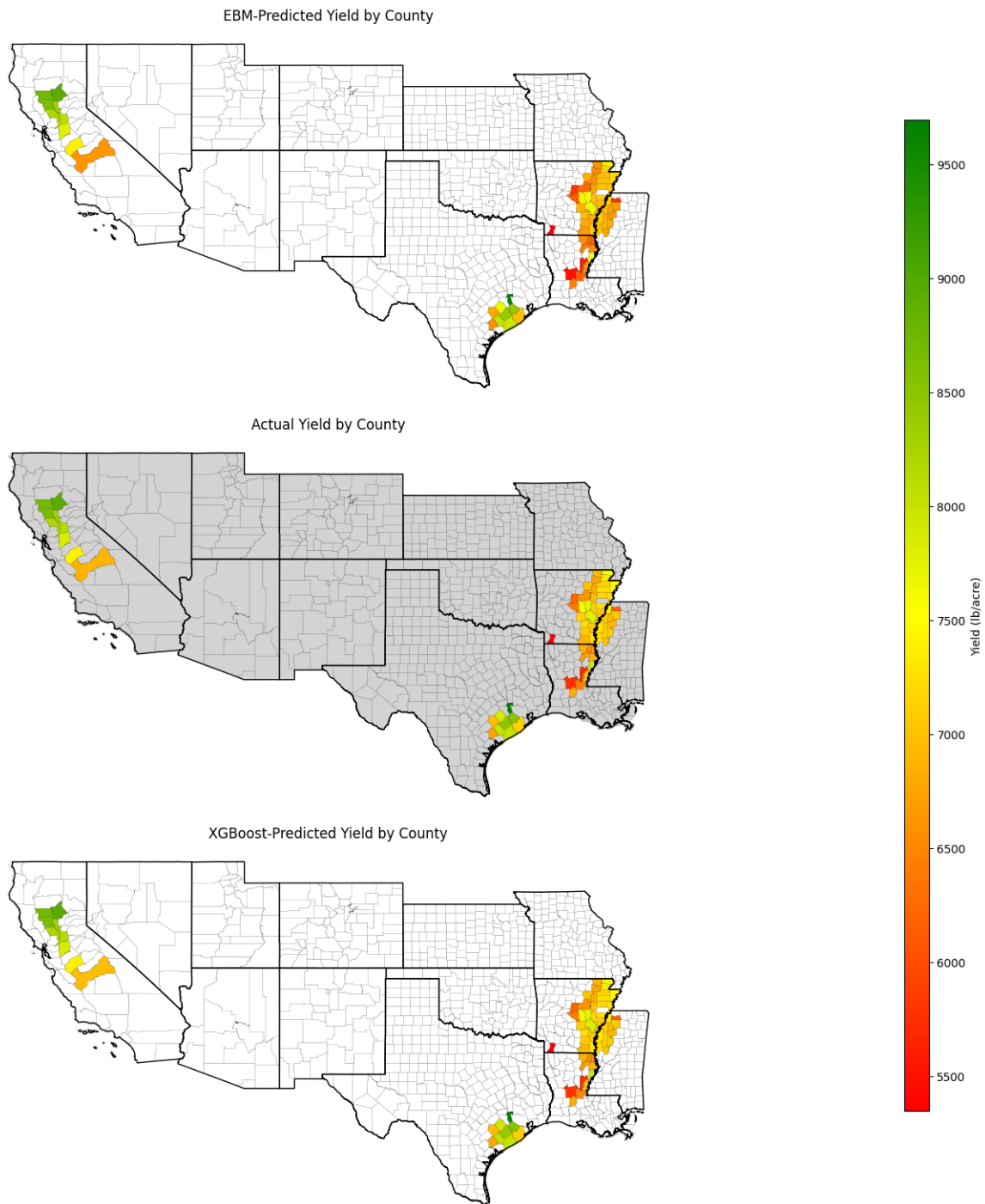


Figure 4.5: Comparison of Actual vs Predicted Yield over Space by Model.

However, some discrepancies between the actual and predicted yields can be observed in certain counties. For example, the EBM model appears to slightly underestimate yields in a few Arkansas counties compared to the actual values. These differences may be attributed to local variations in management practices, soil conditions, or microclimatic factors that are not fully captured by the input features. Conversely, there is no noticeable differences in the performance of the XGBoost across space compared to the actual yield values. Despite minor discrepancies in the case of the EMB model, both models demonstrate a strong ability to capture the overall spatial variability in rice yields across the study area. The high spatial consistency between the actual and predicted yields suggests that the models are robust to spatial variations and can effectively predict yields at the county level.

Temporal Robustness

To assess the temporal robustness of the EBM and XGBoost models, we employ an out-of-time validation approach. The models are trained using data from earlier years and tested on later years, with the training and testing periods progressively shifted forward in time. Table 4.4 presents the performance metrics (RMSE, MAE, and R^2) for both models under different training and testing splits. To ensure the best possible performance, GridSearch was again used for hyperparameter tuning of both models in all cases. The results demonstrate that both models exhibit strong performance on the training data across all time splits, with high R^2 values and low RMSE and MAE scores. This indicates that the models effectively capture the relationships between the input features and rice yields within the training periods.

However, when the models are tested on future years, a decrease in performance is observed. The R^2 values drop, while the RMSE and MAE scores increase, suggesting that the models' ability to generalize to new, unseen data is limited. This is a common challenge in machine learning, known as concept drift, where the relationships between the input features and the target variable change over time.

Table 4.4: Model Performance with Out-of-Time Validation.

Model	Years	RMSE		MAE		R ²	
		Train	Test	Train	Test	Train	Test
EBM	2008-2017 / 2018-2022	0.019	0.116	0.014	0.090	0.983	0.170
	2008-2018 / 2019-2022	0.019	0.118	0.015	0.089	0.982	0.149
	2008-2019 / 2020-2022	0.022	0.115	0.017	0.083	0.977	0.193
	2008-2020 / 2021-2022	0.024	0.110	0.018	0.082	0.972	0.193
	2008-2021 / 2022	0.026	0.099	0.019	0.076	0.969	0.189
XGBoost	2008-2017 / 2018-2022	0.006	0.111	0.005	0.085	0.998	0.244
	2008-2018 / 2019-2022	0.007	0.102	0.005	0.078	0.998	0.366
	2008-2019 / 2020-2022	0.022	0.098	0.017	0.073	0.977	0.414
	2008-2020 / 2021-2022	0.008	0.095	0.006	0.072	0.997	0.399
	2008-2021 / 2022	0.003	0.087	0.002	0.063	1.000	0.386

Despite the performance drop, it is important to note that the XGBoost model generally outperforms the EBM model in terms of temporal robustness. The XGBoost model maintains higher R^2 values and lower RMSE and MAE scores on the testing data compared to the EBM model across all time splits. This suggests that the XGBoost model is better able to capture the temporal patterns and adapt to the changing relationships between the input features and rice yields over time. Furthermore, it is observed that the performance of both models tends to improve as more recent years are included in the training data. For instance, in the last three time splits, the R^2 values are higher, and the RMSE and MAE scores are lower compared to the earlier time splits. This indicates that incorporating more recent data in the training process helps the models to better capture the current relationships and improve their temporal robustness. These results highlight the importance of regularly updating and retraining machine learning models for rice yield prediction. As new data becomes available, it is crucial to incorporate it into the training process to capture the evolving relationships between the input features and rice yields. This allows the models to adapt to changing environmental conditions, management practices, and technological advancements, thereby improving their temporal robustness and predictive accuracy.

To further investigate the temporal robustness of our models, we implement a rolling window approach for out-of-time prediction. This method involves training the models on a fixed-width window of recent years and testing on the subsequent years, progressively moving the window forward. We hypothesize that this approach will provide insights into how well the models adapt to more recent changes in agricultural conditions and environmental factors and whether they maintain predictive power when trained on a smaller, more temporally focused dataset. Table 4.5 presents the results of this analysis for both the EBM and XGBoost models.

Table 4.5: Out-of-Time Prediction with Rolling Window Approach.

Model	Years	RMSE		MAE		R ²	
		Train	Test	Train	Test	Train	Test
EBM	2013-2017 / 2018-2022	0.018	0.099	0.014	0.075	0.980	0.401
	2014-2018 / 2019-2022	0.016	0.106	0.012	0.080	0.986	0.312
	2015-2019 / 2020-2022	0.015	0.112	0.012	0.085	0.988	0.238
	2016-2020 / 2021-2022	0.017	0.112	0.013	0.078	0.985	0.162
	2017-2021 / 2022	0.014	0.113	0.011	0.072	0.989	-0.056
XGBoost	2013-2017 / 2018-2022	0.002	0.091	0.001	0.067	0.999	0.486
	2014-2018 / 2019-2022	0.001	0.091	0.001	0.066	0.999	0.495
	2015-2019 / 2020-2022	0.001	0.096	0.001	0.071	0.999	0.443
	2016-2020 / 2021-2022	0.002	0.104	0.001	0.069	0.999	0.284
	2017-2021 / 2022	0.001	0.107	0.001	0.059	0.999	0.065

The rolling window approach reveals an interesting pattern. For both EBM and XGBoost models, we observe an improvement in test performance for the first three prediction windows compared to the cumulative approach in Table 4.4. For instance, the EBM model's test R² values increase from 0.170 to 0.401 for the 2018-2022 prediction period, while XGBoost's R² improves from 0.244 to 0.486. This suggests that training on the most recent five years allows the models to capture more relevant and up-to-date relationships between the predictors and rice yields. However, this improvement doesn't hold for the last two prediction windows. We observe a decline in performance for both models, with R² values dropping significantly. The EBM model's performance deteriorates more drastically, reaching a negative R² value for the 2022 prediction.

This indicates that for very recent years, even the rolling window approach struggles to maintain predictive power.

Out-of-Season Prediction

To determine the optimal timing for yield prediction, we compare the model performance when using data from the beginning of the season (April-June) versus the late season (July-September). Additionally, we investigate the model performance when using data from April-July for training and August-September for testing. Table 4.6 presents the performance metrics for both the EBM and XGBoost models under these different scenarios.

Table 4.6: Model Performance for Early vs Late Season.

Model	Period	RMSE		MAE		R ²	
		Train	Test	Train	Test	Train	Test
EBM	April-June / July-September	0.040	0.055	0.030	0.041	0.921	0.852
	April-July / August-September	0.038	0.053	0.028	0.039	0.930	0.861
XGBoost	April-June / July-September	0.007	0.030	0.005	0.018	0.998	0.955
	April-July / August-September	0.007	0.029	0.005	0.018	0.998	0.958

The results show that both models perform well when using data from the beginning of the season (April-June) for training and the late season (July-September) for testing. The XGBoost model achieves an impressive R² value of 0.955 on the test set, with low RMSE and MAE scores of 0.030 and 0.018, respectively. The EBM model also performs well, with an R² value of 0.852 and relatively low RMSE and MAE scores of 0.055 and 0.041, respectively.

When the training period is extended to include data from April-July and the testing period is limited to August-September, both models show a slight improvement in performance. The XGBoost model achieves an R² value of 0.958 on the test set, with RMSE and MAE scores of 0.029 and 0.018, respectively. Similarly, the EBM model's R² value increases to 0.861, and the RMSE and MAE scores decrease to 0.053 and 0.039, respectively.

These results indicate that accurate yield predictions can be made as early as the beginning of the growing season (April-June) using the XGBoost and EBM models. This is a significant

finding for rice farmers, as it allows them to make informed decisions regarding crop management, resource allocation, and marketing strategies well in advance of the harvest. By having reliable yield predictions early in the season, farmers can optimize their farming practices, such as irrigation, fertilization, and pest control, to maximize their yields and profitability. Furthermore, the ability to make accurate yield predictions early in the season can help farmers to better plan their labor requirements, storage facilities, and transportation arrangements. This can lead to more efficient use of resources and reduce the risk of post-harvest losses.

It is important to note that the strong performance of the models in the early vs late season analysis contrasts with the relatively poor results obtained from the out-of-time forecasting exercise. This suggests that while the models are able to capture the relationships between the input features and rice yields within a given growing season, their ability to generalize to future years is limited. This again emphasizes the importance of regularly updating and retraining the models as new data becomes available to maintain their predictive power over time.

4.6.4 Yield-Emissions Trade-off

The results of the EBM models developed for yield and emissions prediction are shown in Table 4.7. The models were able to accurately predict both yield and emissions, with high R^2 values and low RMSE and MAE scores on the test set. The strong performance of the EBM models in capturing the complex relationship between input features and the target variables (yield and emissions) lays a solid foundation for the subsequent trade-off analysis.

Table 4.7: Model Performance for Yield and Emission Models.

Model	RMSE		MAE		R^2	
	Train	Test	Train	Test	Train	Test
Yield Model	0.024	0.044	0.018	0.032	0.965	0.893
Emission Model	0.030	0.060	0.023	0.044	0.981	0.932

The feature importance rankings derived from the EBM models, presented in Table 4.8, provide valuable insights into the key factors influencing yield and emissions. For the yield prediction model, the top-ranking features are consistent with those identified in the previous yield analysis section, with harvested area, soil texture, water content, and soil pH at various depths being the most influential variables.

Table 4.8: EBM-based Feature Importance: Yield vs. Emissions Models.

Yield Feature	Importance	Emissions Feature	Importance
AREA HARVESTED	0.016	GEOID	0.043
SoilTextureCat_b200	0.014	WaterContent_b200	0.027
WaterContent_b10	0.013	SoilTextureCat_b30	0.024
GEOID	0.011	SoilTextureCat_b100	0.022
SoilPh_b60	0.011	SoilPh_b0	0.021
SoilPh_b0	0.010	WaterContent_b0	0.019
SoilPh_b30	0.010	SoilPh_b10	0.019
SoilPh_b200	0.009	SoilTextureCat_b10	0.018
WaterContent_b30	0.009	AREA HARVESTED	0.018
WaterContent_b0	0.009	SoilPh_b200	0.018
SoilPh_b10	0.008	WaterContent_b10	0.017
SoilPh_b100	0.008	SoilPh_b100	0.017
SoilTextureCat_b60	0.008	SoilPh_b60	0.016
SoilTextureCat_b100	0.007	WaterContent_b30	0.016
WaterContent_b200	0.007	SoilTextureCat_b200	0.016
WaterContent_b60	0.007	SoilTextureCat_b0	0.014
WaterContent_b100	0.007	WaterContent_b100	0.014
SoilTextureCat_b10	0.006	SoilPh_b30	0.013
mean_FAPAR	0.006	WaterContent_b60	0.012
zndx	0.005	SoilTextureCat_b60	0.010
pcp	0.005	pdsi	0.009
SoilTextureCat_b0	0.005	pcp	0.009
SoilTextureCat_b30	0.005	pmdi	0.009
tmin	0.005	Month	0.009
cdd	0.004	mean_FAPAR	0.008
NDVI	0.004	tmax	0.008
mean_LAI	0.004	NDVI	0.008
phdi	0.004	tmin	0.008
pmdi	0.004	tavg	0.007
Month	0.004	phdi	0.007
tavg	0.004	zndx	0.006
pdsi	0.003	cdd	0.005
tmax	0.003	mean_LAI	0.005

In the case of the emissions prediction model, we observe that the most important features differ from those driving yield variability. The county identifier variable (GEOID) emerges as the top-ranking feature, suggesting that spatial variability and regional characteristics play a significant role in determining methane emissions from rice cultivation. This finding highlights the need for region-specific mitigation strategies and the importance of considering local conditions when developing emissions reduction policies.

Other key features influencing emissions include water content and soil texture at different depths, particularly at the 200 cm and 30 cm layers. The prominence of these variables indicates that water management and soil properties are crucial factors in regulating methane emissions from rice fields. This aligns with the understanding that methane is produced under anaerobic conditions, which are typically associated with flooded rice paddies. The results suggest that optimizing water management practices, such as alternate wetting and drying or midseason drainage, could be effective strategies for mitigating methane emissions.

Interestingly, soil pH also appears among the top features for emissions prediction, albeit at different depths compared to the yield model. This finding suggests that soil pH influences not only rice growth and productivity but also the microbial processes responsible for methane production. Managing soil pH through liming or other amendments may offer opportunities to reduce methane emissions while maintaining optimal conditions for rice cultivation.

The presence of harvested area as an important feature in both the yield and emissions models highlights the trade-offs inherent in rice production. Increasing the area under rice cultivation may boost total production and very likely yield due to economies of scale, but it also has the potential to increase overall methane emissions. This emphasizes the need for a balanced approach that considers both productivity and environmental sustainability.

The trade-off plot in Figure 4.6 reveals a surprising and encouraging relationship between rice yields and methane emissions. Instead of the traditional trade-off where higher yields come at the cost of higher emissions, the graph shows a positive correlation between yield improvement and

emissions reduction. As we move along the Pareto front towards higher yield weights, we observe an increase in predicted yields (blue line) alongside a decrease in predicted emissions (red line), particularly when yield reaches at least 10,000 kg/ha. This unexpected pattern suggests that the farming practices and technologies used to boost rice yields are also contributing to lower methane emissions. Inefficient agricultural practices, such as excess fertilizer use, poor water management in rice fields, and inefficient livestock management, are known to increase greenhouse gas emissions, particularly methane. However, the trade-off plot indicates that the methods employed to enhance rice productivity are simultaneously mitigating these emissions.

This pattern can be explained by examining the feature importance rankings from our EBM models, as shown in Table 4.8. The county identifier variable is a top feature for both yield and emissions, which indicates that local conditions and region-specific practices significantly influence both productivity and environmental impact. Soil characteristics, including texture and pH at various depths, are important for both models. Optimal soil conditions can enhance nutrient retention and water management, potentially increasing yields while reducing methane production. Water content at various soil depths is also crucial, which suggests that efficient water management practices contribute to both higher yields and lower emissions. The area harvested is an important feature for both models. This indicates that larger, more efficiently managed farms may be better equipped to implement advanced technologies that simultaneously boost yields and mitigate emissions.

To better understand the relationship between key features and the model outputs, we analyze partial dependence plots (PDPs) for the common features among the top 10 most important predictors in both the yield and emissions models. These common features are: *area harvested*, *soilph_b0*, *watercontent_b0*, and *soilph_b200*. We exclude *geoid* from this analysis as it's a categorical, non-ordinal variable unsuitable for this method. PDPs help visualize the marginal effect of a feature on the predicted outcome (Friedman, 2001).

For a fitted model $\hat{f}(\mathbf{X})$, where $\mathbf{X}$ represents the complete set of features, let $\mathbf{X}_1 = \{x_1, x_2, \dots, x_k\}$ represent a subset of features of interest and $\mathbf{X}_2 = \{x_{k+1}, x_{k+2}, \dots, x_p\}$ represent

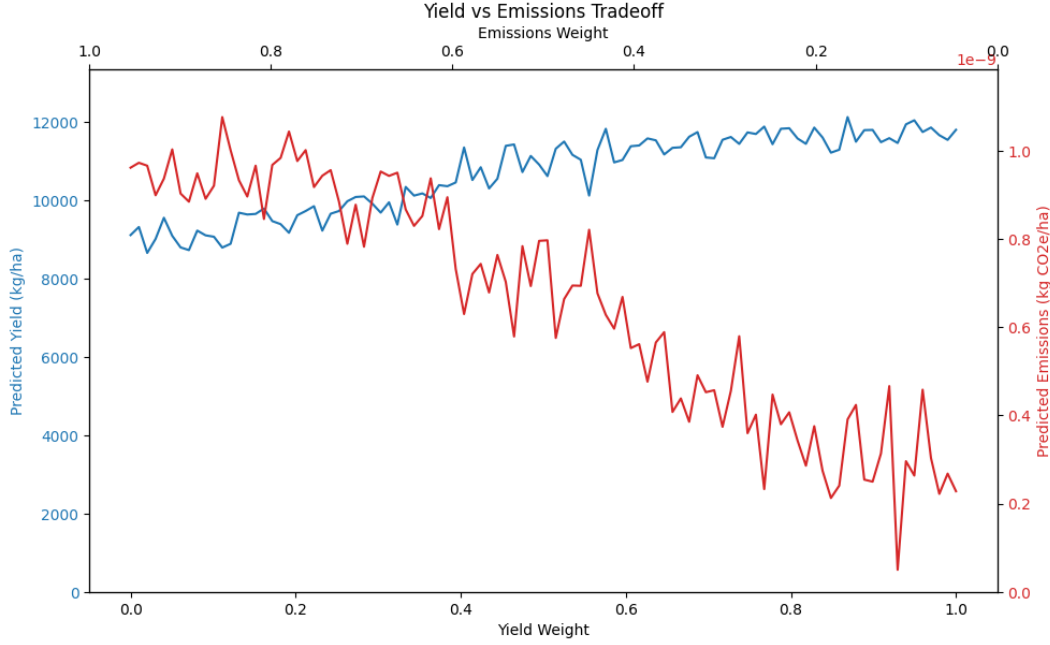


Figure 4.6: Yield-Emissions Trade-off Plot.

Notes: The upper x-axis shows the weight assigned to emissions, while the lower x-axis shows the weight assigned to rice yield. For instance, the farthest left combination represents the case where the full weight is placed on reducing methane emissions, and no weight is placed on yield maximization (the (1,0) pair). In other words, as we move from left to right, we care more about maximizing yield and less about polluting the environment. The next combination represents the (0.99, 0.1) pair, and this pattern continues, changing by 0.1 increments, until we reach the (0,1) combination. We impose a constraint of achieving at least the median value for yield prediction. The blue line represents predicted yield, while the red line represents predicted rice-derived methane emissions.

the remaining features (i.e., the complement of $\mathbf{X}_1$). The partial dependence function for $\mathbf{X}_1$ is defined as:

$$\text{PD}_1(\mathbf{X}_1) = \mathbb{E}[\hat{f}(\mathbf{X}_1, \mathbf{X}_2)] = \int \hat{f}(\mathbf{X}_1, \mathbf{X}_2) p(\mathbf{X}_2) d\mathbf{X}_2$$

where $p(\mathbf{X}_2)$ is the marginal probability density of $\mathbf{X}_2$. The integral is taken over the possible values of $\mathbf{X}_2$, averaging out its influence on the model while keeping $\mathbf{X}_1$ fixed. Given the difficulty to obtain the exact computation for this integral, it is typically approximated using the training data in the following manner:

$$\text{PD}_1(\mathbf{X}_1) \approx \frac{1}{n} \sum_{i=1}^n \hat{f}(\mathbf{X}_1, \mathbf{X}_2^{(i)})$$

where $\mathbf{X}_2^{(i)}$ represents the values of the complement features $\mathbf{X}_2$ for the i -th training instance. In our context, when plotting the partial dependence of *soilph_b0*, $\mathbf{X}_1$ represents the feature *soilph_b0* itself, while $\mathbf{X}_2$ comprises all the other features in the model—including *area harvested*, *watercontent_b0*, *soilph_b200*, and any additional predictors. This means that we are examining how changes in *soilph_b0* affect the model's predictions while averaging out the influence of all other variables.

Figure 4.7 displays the PDPs for our selected features. For *area harvested*, we observe a general upward trend in yield predictions as the harvested area increases, particularly pronounced in the lower range. This suggests economies of scale in rice production. Interestingly, emissions also tend to increase with area, but the relationship is non-linear and shows a sharp decline at higher values. This could indicate that larger operations may implement more efficient practices that reduce the intensity of emissions.

When looking at surface soil pH, yield shows a sharp increase around the mid-range of pH values, suggesting an optimal pH zone for rice productivity. Emissions, conversely, exhibit a slight downward trend as pH increases, with more variability at lower pH levels. This inverse relationship suggests that managing soil pH could potentially offer a win-win scenario for increasing yield while reducing emissions. The PDP for surface soil water content reveals that yield appears relatively stable across different water content levels, with slight increases at certain points. Emissions show more variability, with a notable peak in the mid-range. This complex relationship highlights the delicate balance in water management for rice cultivation, where optimal water levels for yield may not align with minimal emission levels. For subsoil pH at 200 cm depth, the yield curve is flatter compared to surface pH, indicating a more stable but still positive relationship. Emissions show more variability with subsoil pH, which suggests that deeper soil characteristics play a complex role in methane production and release.

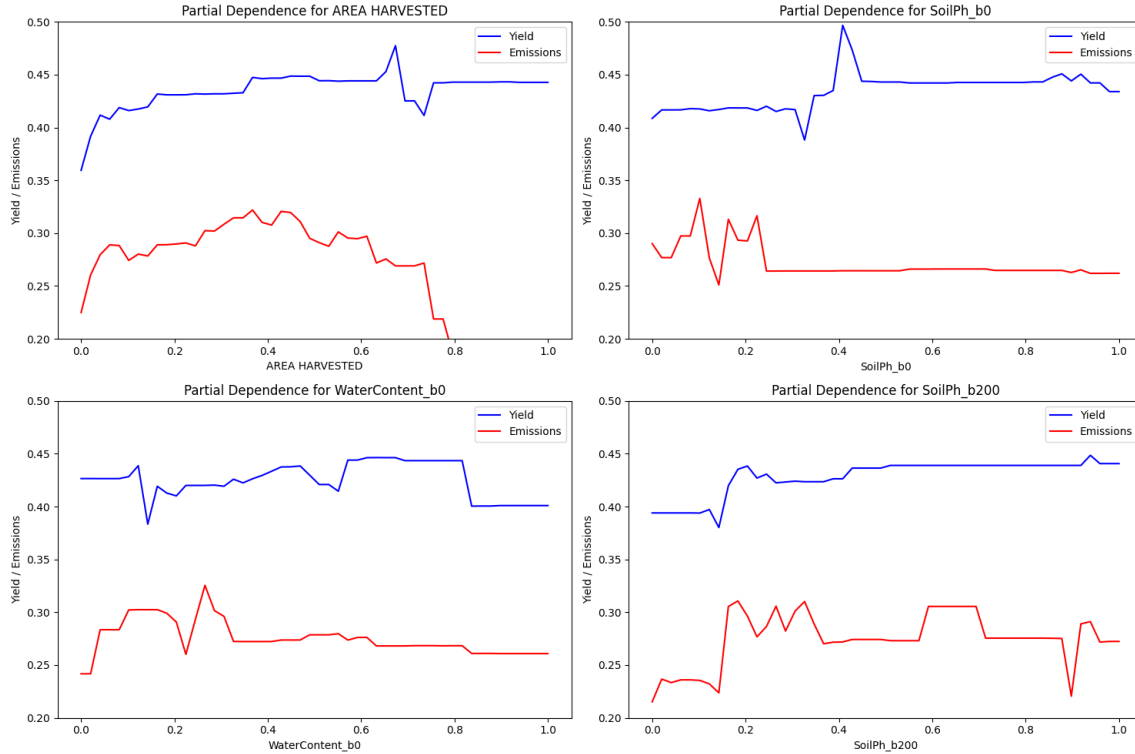


Figure 4.7: PDPs for Common Important Features across Yield and Emission Models.

Notes: The values of both the features reported and the response variable are standardized to be between zero and one. The y-axis range is set between 0.2 and 0.5 to facilitate visual comparison within subplots and between subplots.

The underlying models suggest that practices targeting optimal soil pH levels, especially at the surface, could be particularly effective in simultaneously boosting yields and reducing emissions. The varying impacts of surface versus subsoil pH further emphasize the importance of considering soil properties at different depths. Water management emerges as a critical factor, with its non-linear effects on both yield and emissions highlighting the need for precision in irrigation practices. The relationship between harvested area and outcomes points to potential benefits of scale, possibly due to the adoption of more advanced management techniques in larger operations.

Several factors could explain the positive correlation between yield improvement and emissions reduction. Precision agriculture techniques optimize resource use, ensuring that inputs

like water and fertilizers are applied efficiently, minimizing waste and reducing the potential for emissions. Better soil management practices can enhance carbon sequestration, while optimized fertilizer use reduces excess nitrogen that leads to emissions. Interestingly, this synergy between yield enhancement and emissions mitigation aligns with the observation in Figure 4.2 where counties with higher rice yields, particularly those in California, appeared to have lower rice-related methane emissions.

The implications of this finding are significant. It suggests that investing in efficient, sustainable agricultural practices can lead to a win-win scenario, where improvements in rice productivity go hand in hand with reduced environmental impact. By adopting precision agriculture techniques and optimizing resource use, rice farmers can maximize yields while minimizing methane emissions.

However, it is important to note that the relationship between yield and emissions along the Pareto front is not perfectly linear, as evident from the variability in the trade-off plot. This indicates that not all yield-improving methods may have an equal impact on emissions reduction. Therefore, careful evaluation and selection of sustainable practices are necessary to achieve the desired balance between productivity and environmental sustainability.

4.7 Discussions

Through the use of advanced machine learning techniques and a comprehensive dataset including remote sensing, climate variables, and soil properties, we identify key determinants of rice productivity and environmental impact. Our analysis reveals that soil properties, especially soil pH and texture across various depths, play a pivotal role in both rice yields and methane emissions. This finding corroborates previous research on the importance of soil acidity for water availability and nutrient uptake (Espe et al., 2016), and extends the understanding to greenhouse gas emissions. The dual impact of soil characteristics on productivity (i.e., rice yield) and environmental outcomes (i.e., methane emissions) suggests that targeted soil management practices could simultaneously

enhance yields and reduce methane emissions. This aligns with the observations of Qian et al. (2023), who identified diminishing returns to yield as cultivated area expands, which emphasizes the need for improved soil quality and management practices alongside area expansion to prevent productivity losses.

The pronounced spatiotemporal variability in the relationships between environmental factors and rice yields emphasizes the complexity of agricultural systems. Our findings support the observations of Park et al. (2018), who noted that rice yield is highly sensitive to temperature variability, particularly during the reproductive phase of the crop. The correlation between precipitation and rice yield also exhibits high spatial variability, with some regions benefiting from increased rainfall while others suffer from waterlogging and associated yield declines. This variability necessitates models that incorporate spatial heterogeneity and temporal dynamics to enhance the precision of yield forecasts. Our results demonstrate the effectiveness of satellite-derived vegetation indices, including NDVI, LAI, and FAPAR, as predictors of rice yields. This aligns with the findings of Cao et al. (2021), who suggested that crop yields can be predicted 1-2 months before maturity with satellite vegetation indices, which has important implications for establishing food security early warning systems. Notably, the predictive power of these indices increases as the growing season advances, which suggests that remote sensing data are particularly valuable during the later stages of crop development.

The application of advanced machine learning models, especially ensemble methods such as XGBoost and EBM, proves highly effective in the capture of complex, non-linear relationships between environmental variables and rice yields. This corroborates the findings of You et al. (2023), who demonstrated that deep learning models, particularly CNNs, excel at capturing complex, non-linear relationships in high-dimensional, multi-source datasets that include satellite imagery and ground-truth data like soil and weather variables. The observed data drift over time highlights the dynamic nature of agricultural systems and the challenges associated with model accuracy maintenance amidst changing environmental conditions. This aligns with the concerns raised by

Cao et al. (2021), whose data drift analysis showed that the marginal distribution of climate variables changes significantly over time, necessitating regular model updates to maintain predictive accuracy. The shifting distributions of climate variables indicate that static models may quickly become outdated, which compromises their predictive capabilities. The unexpected positive correlation between yield improvement and emissions reduction challenges the conventional notion of an inherent trade-off between agricultural productivity and environmental sustainability. Our findings suggest that the adoption of efficiency-enhancing practices, including precision agriculture and optimized resource management, can lead to simultaneous gains in yield and reductions in methane emissions. This synergy supports the paradigm of sustainable intensification, where the goal is to increase agricultural output without exacerbation of environmental pressures.

4.8 Conclusions and Policy Implications

This study provides a comprehensive analysis of rice yield prediction and methane emissions in U.S. rice production through the leverage of advanced machine learning techniques and integration of diverse data sources. The findings have significant implications for agricultural economics, policy, and practice in the context of climate change and food security. The strong predictive capabilities of machine learning models, especially when enriched with high-resolution satellite data and detailed soil information, highlight the significant potential of precision agriculture approaches. Policymakers and agricultural stakeholders should consider investment in and promotion of technologies that facilitate data-driven decision-making at the farm level. Implementation of practices such as variable-rate fertilizer application and optimized irrigation scheduling based on real-time monitoring can improve resource use efficiency, increase yields, and minimize environmental impacts.

The significant influence of climate variables on rice yields indicates the urgent need for climate-smart agricultural policies. Strategies such as promotion of drought-resistant rice varieties, investment in advanced water management infrastructure, and development of early warning systems for extreme weather events are essential to enhance the resilience of rice production systems. Support

for research into adaptive farming practices that mitigate the adverse effects of climate change will further strengthen the sector's capacity to cope with environmental uncertainties. Recognition of the pivotal role of soil properties in both yields and emissions should lead to policies that emphasize soil conservation and enhancement practices. Provision of incentives for methods such as cover cropping, reduced tillage, and other soil health-promotion techniques can significantly improve soil quality. Moreover, implementation of targeted soil amendment programs informed by spatially explicit soil data can optimize soil pH and texture for rice cultivation while concurrently reduce methane emissions. Such practices are fundamental to the achievement of sustainable soil management, which underpins long-term agricultural productivity and environmental stewardship.

The observed synergy between yield improvement and emissions reduction indicates that policies promoting sustainable yield intensification could lead to mutually beneficial outcomes. Policymakers might consider the design of incentive programs or carbon credit schemes that reward farmers for implementation of practices like alternate wetting and drying irrigation, which enhances yields while reducing methane emissions. Such initiatives can accelerate the adoption of sustainable practices and facilitate the transition towards more environmentally friendly agricultural systems. The importance of diverse, high-quality data sources in this study points to the necessity of robust agricultural data infrastructures. Support for the creation of standardized data collection protocols, promotion of open data initiatives, and establishment of platforms for data sharing among researchers, farmers, and policymakers can significantly enhance the accuracy and timeliness of yield forecasts. Improved data accessibility and integration will facilitate better-informed decision-making and contribute to more effective food security planning at both regional and national scales.

The data drift and temporal shifts in yield-environment relationships observed in this study highlight the need for adaptive and flexible policymaking. Agricultural policies should accommodate evolving environmental conditions and integrate emerging scientific knowledge. Regular policy evaluations informed by the latest modeling and empirical data can ensure that interventions remain effective under changing climatic scenarios, thus protect agricultural productivity and promote

sustainability. The significant spatial heterogeneity in yield-environment relationships indicates that uniform, blanket policies may not be effective across diverse agricultural landscapes. Policymakers should consider the adoption of integrated landscape management approaches that are sensitive to local environmental conditions, farming practices, and socio-economic contexts. Crafting of region-specific best practice guidelines and tailoring of support programs to meet local needs can improve the efficacy of interventions and promote equitable agricultural development across different regions. Overall, this study demonstrates the power of advanced analytics combined with diverse data sources to gain actionable insights for sustainable rice production. The embrace of data-driven, climate-smart, and locally adapted approaches by policymakers and agricultural stakeholders can enhance the resilience and sustainability of U.S. rice production in the face of environmental challenges. This not only contributes to national food security but also supports global efforts in climate change mitigation and sustainable development.

CHAPTER 5

CONCLUSIONS

This dissertation presents a rigorous, multifaceted analysis of the complex relationships between greenhouse gas emissions, energy systems, economic activity, and climate change impacts in the United States. By employing advanced econometric techniques, quasi-experimental methods, and machine learning algorithms, we uncover crucial insights into the sectoral and regional heterogeneity of emissions drivers, the effectiveness of subnational renewable energy policies, and the dynamics between agricultural productivity and methane emissions. The findings highlight the urgent need for nuanced, context-specific climate policies that account for the diverse economic, technological, and ecological factors shaping emissions patterns across the U.S.

In the first chapter, we explored the causal relationships between economic growth, sectoral energy consumption, and CO_2 emissions. One notable takeaway is the outsized influence of the transportation and industrial sectors on overall emissions. This suggests a clear need for focused interventions—whether through stricter fuel efficiency standards, low-carbon fuel mandates, or carbon pricing—to accelerate decarbonization within these high-emission sectors. At the same time, the role of renewable energy in mitigating emissions highlights the importance of continued investment in clean energy initiatives. Expanding renewable portfolio standards, streamlining grid interconnection, and strengthening transmission infrastructure are crucial to advancing the shift toward a low-carbon economy.

In Chapter Two, the causal analysis of state net metering programs revealed a gradual but significant reduction in residential emissions. This highlights the potential for reforming electricity markets to maximize the benefits of distributed solar energy. Redesigning retail rates to better reflect the true value of distributed generation, for example, could help. Additionally, policies like minimum bill requirements would ensure fair distribution of grid maintenance costs. More

transformative changes, such as adopting uniform interconnection standards or removing capacity restrictions, could further democratize access to clean energy. Combined with federal incentives, these reforms could significantly boost the deployment of distributed renewables.

The third chapter used explainable machine learning to predict rice yields and methane emissions, offering insights for climate-smart agriculture. The strong performance of the predictive models demonstrates the potential for advanced algorithms and remote sensing data in sustainable land management. Notably, this research revealed surprising synergies between yield optimization and methane mitigation. This suggests that practices like precision nutrient application or conservation tillage could simultaneously increase food security and reduce emissions. Scaling up these techniques through extension services and policy support could provide farmers with the tools they need to adopt more climate-resilient practices.

Together, these three studies emphasize the need for context-specific nuanced climate mitigation strategies. The stark variations in emissions profiles, renewable energy potential, and policy contexts across the U.S. mean that a one-size-fits-all approach won't work. Federal policies—such as carbon pricing or clean energy standards—must be designed with enough flexibility to support regional diversity. By setting ambitious national goals while providing targeted resources and incentives, the federal government can drive bottom-up action and ensure an equitable transition to a low-carbon future.

More broadly, this dissertation highlights the value of data-driven decision-making in climate policy. With increasing access to geospatial data, sensor networks, and powerful computational tools, policymakers can create more precise, adaptable, and impactful strategies. Interdisciplinary collaboration, drawing insights from fields as varied as atmospheric science and political economy, will deepen our understanding of the forces driving emissions. In turn, this will enable more holistic and effective responses. Partnerships between researchers, policymakers, and local communities can further help translate scientific findings into actionable, on-the-ground solutions that reflect local needs and priorities.

As the U.S. faces the mounting threats posed by climate change—ranging from extreme weather events to worsening public health impacts—the urgency for ambitious emissions reductions is undeniable. Recent legislative milestones, such as the Inflation Reduction Act and the Infrastructure Investment and Jobs Act, represent a significant step forward. However, to limit global warming to 1.5°C, much more needs to be done. The recommendations presented in this dissertation can guide the development of effective, evidence-based climate policies that are both equitable and transformative.

Ultimately, achieving a net-zero future will involve difficult trade-offs, unexpected consequences, and complex distributional impacts. Navigating these challenges requires an ongoing commitment to learning, adaptation, and inclusive dialogue. Climate policies must be grounded in rigorous data and analysis, but they must also center the voices of vulnerable populations. If the United States can harness its innovation and collective will, it has the potential to lead the global effort in forging a future that is not only sustainable but also equitable and resilient. While the climate crisis poses immense challenges, it also offers a unique opportunity to rebuild our infrastructure, revitalize industries, and reimagine how we live and work, all while respecting the planet's natural limits. This research is just one small contribution to that greater, intergenerational effort.

REFERENCES

- Acaravci, A., & Ozturk, I. (2010). On the relationship between energy consumption, co2 emissions and economic growth in europe. *Energy*, 35, 5412-5420. doi: 10.1016/j.energy.2010.07.009
- Acheampong, A. O. (2018, 8). Economic growth, co2 emissions and energy consumption: What causes what and where? *Energy Economics*, 74, 677-692. doi: 10.1016/j.eneco.2018.07.022
- Apergis, N., & Payne, J. E. (2010, 1). The emissions, energy consumption, and growth nexus: Evidence from the commonwealth of independent states. *Energy Policy*, 38, 650-655. doi: 10.1016/j.enpol.2009.08.029
- Baker, J. T. (2004, 4). Yield responses of southern us rice cultivars to co2 and temperature. *Agricultural and Forest Meteorology*, 122, 129-137. doi: 10.1016/j.agrformet.2003.09.012
- Bandumula, N. (2018, 12). Rice production in asia: Key to global food security. *Proceedings of the National Academy of Sciences India Section B - Biological Sciences*, 88, 1323-1328. doi: 10.1007/s40011-017-0867-7
- Begum, R. A., Sohag, K., Abdullah, S. M. S., & Jaafar, M. (2015). *Co2 emissions, energy consumption, economic and population growth in malaysia* (Vol. 41). Elsevier Ltd. doi: 10.1016/j.rser.2014.07.205
- Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly journal of economics*, 119(1), 249–275.
- Bielen, S. (2024). Prosecutors and crime deterrence: Evidence from a difference-in-differences analysis with staggered treatment. *Journal of Criminal Justice*, 90, 102147.

- Bishnoi, S., Datt, R., Arya, S., Gupta, S., Gupta, R., Sharma, S. N., ... Gupta, V. (2022). Engineered cathode buffer layers for highly efficient organic solar cells: a review. *Advanced Materials Interfaces*, 9. doi: 10.1002/admi.202101693
- Bishwajit, G., Sarker, S., Kpoghomou, M. A., Gao, H., Jun, L., Yin, D., & Ghosh, S. (2013, 7). *Self-sufficiency in rice and food security: A south asian perspective* (Vol. 2). BioMed Central Ltd. doi: 10.1186/2048-7010-2-10
- Breitung, J. (2001). The local power of some unit root tests for panel data. In *Nonstationary panels, panel cointegration, and dynamic panels* (pp. 161–177). Emerald Group Publishing Limited.
- Brown, D. P., & Sappington, D. E. (2017, 5). Designing compensation for distributed solar generation: Is net metering ever optimal? *Energy Journal*, 38, 1-32. doi: 10.5547/01956574.38.3.dbro
- Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of econometrics*, 225(2), 200–230.
- Cao, J., Zhang, Z., Tao, F., Zhang, L., Luo, Y., Zhang, J., ... Xie, J. (2021, 2). Integrating multi-source data for rice yield prediction across china using machine learning and deep learning approaches. *Agricultural and Forest Meteorology*, 297. doi: 10.1016/j.agrformet.2020.108275
- Celik, M. F., Isik, M. S., Taskin, G., Erten, E., & Camps-Valls, G. (2023). Explainable artificial intelligence for cotton yield prediction with multisource data. *IEEE Geoscience and Remote Sensing Letters*, 20. doi: 10.1109/LGRS.2023.3303643
- Chang, S., Cho, J., Heo, J., Kang, J., & Kobashi, T. (2022). Energy infrastructure transitions with pv and ev combined systems using techno-economic analyses for decarbonization in cities. *Applied Energy*, 319, 119254.

- Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE transactions on evolutionary computation*, 6(2), 182–197.
- Dehdar, F., Fuinhas, J. A., Alavijeh, N. K., Nazeer, N., & Zangoei, S. (2023, 2). Investigating the determinants of carbon emissions in the usa: a state-level analysis. *Environmental Science and Pollution Research*, 30, 23023-23034. doi: 10.1007/s11356-022-23831-x
- Delgado, B. M., Kotireddy, R., Cao, S., Hasan, A., Hoes, P. J., Hensen, J. L., & Sirén, K. (2018, 3). Lifecycle cost and co2 emissions of residential heat and electricity prosumers in finland and the netherlands. *Energy Conversion and Management*, 160, 495-508. doi: 10.1016/j.enconman.2018.01.069
- Environmental Protection Agency. (2023). *Sources of greenhouse gas emissions*. <https://www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions>.
- Espe, M. B., Yang, H., Cassman, K. G., Guilpart, N., Sharifi, H., & Linquist, B. A. (2016, 1). Estimating yield potential in temperate high-yielding, direct-seeded us rice production systems. *Field Crops Research*, 193, 123-132. doi: 10.1016/j.fcr.2016.04.003
- Fageria, N. K. (2007, 6). *Yield physiology of rice* (Vol. 30). doi: 10.1080/15226510701374831
- Feldman, D., Barbose, G., Margolis, R., Bolinger, M., Chung, D., Fu, R., . . . Wiser, R. (2015). *Photovoltaic system pricing trends. historical, recent, and near-term projections* (Tech. Rep.). National Renewable Energy Lab.(NREL), Golden, CO (United States).
- Feng, L., Wang, Y., Zhang, Z., & Du, Q. (2021, 9). Geographically and temporally weighted neural network for winter wheat yield prediction. *Remote Sensing of Environment*, 262. doi: 10.1016/j.rse.2021.112514
- Forsyth, T. L., Pedden, M., & Gagliano, T. (2002). *The effects of net metering on the use of small-scale wind systems in the united states*. Retrieved from <http://www.doe.gov/bridge>

- Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189–1232.
- Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of econometrics*, 225(2), 254–277.
- Hadri, K. (2000). Testing for stationarity in heterogeneous panel data. *The Econometrics Journal*, 3(2), 148–161.
- Hamit-Haggar, M. (2012). Greenhouse gas emissions, energy consumption and economic growth: A panel cointegration analysis from canadian industrial sector perspective. *Energy Economics*, 34, 358–364. doi: 10.1016/j.eneco.2011.06.005
- Hashemi, M., Jenkins, G., & Milne, F. (2023, 12). Rooftop solar with net metering: An integrated investment appraisal. *Renewable and Sustainable Energy Reviews*, 188. doi: 10.1016/j.rser.2023.113803
- Hengl, T. (2018a). *Soil ph in h2o at 6 standard depths (0, 10, 30, 60, 100 and 200 cm) at 250 m resolution (version v02) [data set]*. Retrieved from <https://doi.org/10.5281/zenodo.1475459> (Accessed: 2024-05-28)
- Hengl, T. (2018b). *Soil texture classes (usda system) for 6 soil depths (0, 10, 30, 60, 100 and 200 cm) at 250 m (version v02) [data set]*. Retrieved from <https://doi.org/10.5281/zenodo.1475451> (Accessed: 2024-05-28)
- Hengl, T., & Gupta, S. (2019). *Soil water content (volumetric percent) for 33kpa and 1500kpa suctions predicted at 6 standard depths (0, 10, 30, 60, 100 and 200 cm) at 250 m resolution (version v01) [data set]*. Retrieved from <https://doi.org/10.5281/zenodo.2629589> (Accessed: 2024-05-28)
- Hossain, M., & Fischer, . (1995). *Rice research for food security and sustainable agricultural development in asia: Achievements and future challenges* (Vol. 35).

- Hu, T., Zhang, X., Bohrer, G., Liu, Y., Zhou, Y., Martin, J., . . . Zhao, K. (2023). Crop yield prediction via explainable ai and interpretable machine learning: Dangers of black box models for evaluating climate change impacts on crop yield. *Agricultural and Forest Meteorology*, 336, 109458.
- Huntington, T., Cui, X., Mishra, U., & Scown, C. D. (2020, 5). Machine learning to predict biomass sorghum yields under future climate scenarios. *Biofuels, Bioproducts and Biorefining*, 14, 566-577. doi: 10.1002/bbb.2087
- Iliopoulos, T. G., Fermeiglia, M., & Vanheusden, B. (2020, 7). The eu's 2030 climate and energy policy framework: How net metering slips through its net. *Review of European, Comparative and International Environmental Law*, 29, 245-256. doi: 10.1111/reel.12339
- Im, K. S., Pesaran, M. H., & Shin, Y. (2003). Testing for unit roots in heterogeneous panels. *Journal of econometrics*, 115(1), 53–74.
- Ji, B., Sun, Y., Yang, S., & Wan, J. (2007, 6). Artificial neural networks for rice yield prediction in mountainous regions. In (Vol. 145, p. 249-261). doi: 10.1017/S0021859606006691
- Jian, J., Fan, X., He, P., Xiong, H., & Shen, H. (2019, 9). The effects of energy consumption, economic growth and financial development on co2 emissions in china: A vecm approach. *Sustainability (Switzerland)*, 11. doi: 10.3390/su11184850
- Johansen, S., & Juselius, K. (1990). Maximum likelihood estimation and inference on cointegration—with applications to the demand for money. *Oxford Bulletin of Economics and statistics*, 52(2), 169–210.
- Kao, C. (1999). Spurious regression and residual-based tests for cointegration in panel data. *Journal of econometrics*, 90(1), 1–44.
- Kramarz, T., Park, S., & Johnson, C. (2021). Governing the dark side of renewable energy: A typology of global displacements. *Energy Research & Social Science*, 74, 101902.

- Kwiatkowski, D., Phillips, P. C., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of econometrics*, 54(1-3), 159–178.
- Legislative Analyst Office. (2022). *Climate change impacts across california*. Retrieved from <https://lao.ca.gov/Publications/Report/4575>
- Levin, A., Lin, C.-F., & Chu, C.-S. J. (2002). Unit root tests in panel data: asymptotic and finite-sample properties. *Journal of econometrics*, 108(1), 1–24.
- Li, C., Salas, W., DeAngelo, B., & Rose, S. (2006, 7). Assessing alternatives for mitigating net greenhouse gas emissions and increasing yields from rice production in china over the next twenty years. *Journal of Environmental Quality*, 35, 1554-1565. doi: 10.2134/jeq2005.0208
- Liu, X., & Yuan, X. (2023). Novel research methods for energy use, carbon emissions, and economic growth: evidence from the usa. *Economic Research-Ekonomska Istrazivanja*, 36, 1735-1750. doi: 10.1080/1331677X.2022.2092763
- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30.
- Maasakkers, J. D., McDuffie, E. E., Sulprizio, M. P., Chen, C., Schultz, M., Brunelle, L., . . . others (2023). A gridded inventory of annual 2012–2018 us anthropogenic methane emissions. *Environmental science & technology*, 57(43), 16276–16288.
- Maddala, G. S., & Wu, S. (1999). A comparative study of unit root tests with panel data and a new simple test. *Oxford Bulletin of Economics and statistics*, 61(S1), 631–652.
- Maraseni, T. N., Deo, R. C., Qu, J., Gentle, P., & Neupane, P. R. (2018, 1). An international comparison of rice consumption behaviours and greenhouse gas emissions from rice production. *Journal of Cleaner Production*, 172, 2288-2300. doi: 10.1016/j.jclepro.2017.11.182

- Marrero, G. A. (2010, 11). Greenhouse gases emissions, growth and the energy mix in europe. *Energy Economics*, 32, 1356-1363. doi: 10.1016/j.eneco.2010.09.007
- Menyah, K., & Wolde-Rufael, Y. (2010, 6). Co2 emissions, nuclear energy, renewable energy and economic growth in the us. *Energy Policy*, 38, 2911-2915. doi: 10.1016/j.enpol.2010.01.024
- Mitova, S., & Kahsar, R. (2022). Growing pv generation and ev loads in a world of declining net metering: impacts on household-level economic, emission, and energy benefits. *International Journal of Environmental Studies*. doi: 10.1080/00207233.2022.2126116
- Nori, H., Jenkins, S., Koch, P., & Caruana, R. (2019). Interpretml: A unified framework for machine learning interpretability. *arXiv preprint arXiv:1909.09223*.
- Ozturk, I., & Acaravci, A. (2010). *Co2 emissions, energy consumption and economic growth in turkey* (Vol. 14). Elsevier Ltd. doi: 10.1016/j.rser.2010.07.005
- Pao, H. T., & Tsai, C. M. (2010, 12). Co2 emissions, energy consumption and economic growth in bric countries. *Energy Policy*, 38, 7850-7860. doi: 10.1016/j.enpol.2010.08.045
- Pao, H. T., & Tsai, C. M. (2011b). Modeling and forecasting the co2 emissions, energy consumption, and economic growth in brazil. *Energy*, 36, 2450-2458. doi: 10.1016/j.energy.2011.01.032
- Pao, H. T., Yu, H. C., & Yang, Y. H. (2011a). Modeling the co2 emissions, energy use, and economic growth in russia. *Energy*, 36, 5094-5100. doi: 10.1016/j.energy.2011.06.004
- Paramati, S. R., Mo, D., & Gupta, R. (2017, 8). The effects of stock market growth and renewable energy use on co2 emissions: Evidence from g20 countries. *Energy Economics*, 66, 360-371. doi: 10.1016/j.eneco.2017.06.025
- Park, J. K., Das, A., & Park, J. H. (2018, 4). Integrated model for predicting rice yield with climate change. *International Agrophysics*, 32, 203-215. doi: 10.1515/intag-2017-0010

- Pedroni, P. (1999). Critical values for cointegration tests in heterogeneous panels with multiple regressors. *Oxford Bulletin of Economics and statistics*, 61(S1), 653–670.
- Pedroni, P. (2001). Fully modified ols for heterogeneous cointegrated panels. In *Nonstationary panels, panel cointegration, and dynamic panels* (pp. 93–130). Emerald Group Publishing Limited.
- Pesaran, H. H., & Shin, Y. (1998). Generalized impulse response analysis in linear multivariate models. *Economics letters*, 58(1), 17–29.
- Pesaran, H. H., Yongcheol, S., & P, S. R. (1999). Pooled mean group estimation of dynamic heterogeneous panels. *Journal of the American statistical Association*, 94(446), 621–634.
- Poullikkas, A., Kourtis, G., & Hadjipaschalis, I. (2013). *International journal of energy and environment a review of net metering mechanism for electricity renewable energy sources* (Vol. 4). Online. Retrieved from www.IJEE.IEEFoundation.org
- Qian, H., Zhu, X., Huang, S., Linquist, B., Kuzyakov, Y., Wassmann, R., . . . Jiang, Y. (2023, 10). *Greenhouse gas emissions and mitigation in rice agriculture* (Vol. 4). Springer Nature. doi: 10.1038/s43017-023-00482-1
- Rehman, W. U., Bhatti, A. R., Awan, A. B., Sajjad, I. A., Khan, A. A., Bo, R., . . . Oboreh-Snapps, O. (2020). *The penetration of renewable and sustainable energy in asia: A state-of-the-art review on net-metering* (Vol. 8). Institute of Electrical and Electronics Engineers Inc. doi: 10.1109/ACCESS.2020.3022738
- REN21 Secretariat. (2020). Renewables 2020 global status report. *Rep. Paris: REN12*. Retrieved from https://www.ren21.net/wp-content/uploads/2019/05/gsr_2020_full_report_en.pdf
- Ritchie, H. (2019). Who has contributed most to global co2 emissions? *Our World in Data*. Retrieved from <https://ourworldindata.org/contributed-most-global-co2>

- Ros, A. J., & Sai, S. S. (2023, 2). Residential rooftop solar demand in the u.s. and the impact of net energy metering and electricity prices. *Energy Economics*, 118. doi: 10.1016/j.eneco.2022.106491
- Salari, M., Javid, R. J., & Noghanibehambari, H. (2021, 3). The nexus between co2 emissions, energy consumption, and economic growth in the u.s. *Economic Analysis and Policy*, 69, 182-194. doi: 10.1016/j.eap.2020.12.007
- Schelly, C., Louie, E. P., & Pearce, J. M. (2017, 12). Examining interconnection and net metering policy for distributed generation in the united states. *Renewable Energy Focus*, 22-23, 10-19. doi: 10.1016/j.ref.2017.09.002
- Sioshansi, F., & Pfaffenberger, W. (2006). *Electricity market reform: an international perspective*. Elsevier.
- Smith, K. M., Koski, C., & Siddiki, S. (2021, 3). Regulating net metering in the united states: A landscape overview of states' net metering policies and outcomes. *Electricity Journal*, 34. doi: 10.1016/j.tej.2020.106901
- Statista. (2024). *Principal rice exporting countries worldwide in 2023/2024*. Retrieved from <https://www.statista.com/statistics/255947/top-rice-exporting-countries-worldwide-2011/> (Accessed: 2024-10-07)
- Sudjianto, A., Zhang, A., Yang, Z., Su, Y., & Zeng, N. (2023). Piml toolbox for interpretable machine learning model development and validation. *arXiv preprint arXiv:2305.04214*.
- Sun, Di, L., Sun, Z., Shen, Y., & Lai, Z. (2019, 10). County-level soybean yield prediction using deep cnn-lstm model. *Sensors (Switzerland)*, 19. doi: 10.3390/s19204363
- Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of econometrics*, 225(2), 175–199.

- Szmolyan, A. (2020). *Disunity among the united states: Navigating net-metering disunity among the united states: Navigating net-metering without getting electrocuted without getting electrocuted*. Retrieved from <https://digitalcommons.pepperdine.edu/jbel/vol13/iss1/5>
- Trevino-Martinez, S., Sawhney, R., & Sims, C. (2022, 12). Energy-carbon neutrality optimization in production scheduling via solar net metering. *Journal of Cleaner Production*, 380. doi: 10.1016/j.jclepro.2022.134627
- U.S. Energy Information Administration. (2020). *Residential energy consumption survey (recs)*. Retrieved from <https://www.eia.gov/consumption/residential/faqs.php> (Accessed: 2024-11-04)
- U.S. Environmental Protection Agency. (2024). *Ghg emission factors hub*. Retrieved from <https://www.epa.gov/climateleadership/ghg-emission-factors-hub> (Accessed: 2024-11-04)
- USARice. (2024). *U.s. rice facts*. Retrieved from <https://www.usarice.com/thinkrice/discover-us-rice/us-rice-facts> (Accessed: 2024-02-29)
- USDA Economic Research Service. (2024). *Rice sector at a glance*. Retrieved from <https://www.ers.usda.gov/topics/crops/rice/rice-sector-at-a-glance/> (Accessed: 2024-02-29)
- USDA National Agricultural Statistics Service. (2024). *Cropland data layer*. Retrieved from <https://croplandcros.scinet.usda.gov/> (Accessed: 2024-05-31)
- Vermote, E. (2019a). *Noaa climate data record (cdr) of avhrr leaf area index (lai) and fraction of absorbed photosynthetically active radiation (fapar), version 5. [2008-2022]*. Retrieved from <https://doi.org/10.7289/V5TT4P69> (Accessed: 2024-05-28)

- Vermote, E. (2019b). *Noaa climate data record (cdr) of avhrr normalized difference vegetation index (ndvi), version 5. [indicate subset used]*. Retrieved from <https://doi.org/10.7289/V5ZG6QH9> (Accessed: 2024-05-28)
- Wang, S. S., Zhou, D. Q., Zhou, P., & Wang, Q. W. (2011, 9). Co2 emissions, energy consumption and economic growth in china: A panel data analysis. *Energy Policy*, 39, 4870-4875. doi: 10.1016/j.enpol.2011.06.032
- Westerlund, J. (2005). New simple tests for panel cointegration. *Econometric Reviews*, 24(3), 297–316.
- You, J., Li, X., Low, M., Lobell, D., & Ermon, S. (2023). *Deep gaussian process for crop yield prediction based on remote sensing data*. Retrieved from www.aaai.org
- Zaman, K., & el Moemen, M. A. (2017). *Energy consumption, carbon dioxide emissions and economic development: Evaluating alternative and plausible environmental hypothesis for sustainable growth* (Vol. 74). Elsevier Ltd. doi: 10.1016/j.rser.2017.02.072

APPENDIX A

CHAPTER 2

Table A.1: Short-run Granger Causality/Block Exogeneity Wald Tests Summary for the Base Model.

From	To						
	ΔTOT_CO2	$\Delta ENER_USE$	ΔGDP	ΔGDP_SQ	$\Delta ENER_PRICE$	ΔH_DD	$\Delta RENER_USE$
ΔTOT_CO2	-	16.368***	1.910	1.823	12.977***	27.649***	2.089
$\Delta ENER_USE$	7.243**	-	14.443***	15.343***	36.528***	5.260*	7.023**
ΔGDP	4.247	3.485	-	27.329***	13.396***	16.639***	5.665*
ΔGDP_SQ	3.634	2.858	13.921***	-	14.953***	17.740***	4.717*
$\Delta ENER_PRICE$	3.068	11.844***	0.250	0.435	-	150.494***	10.582***
ΔH_DD	18.295***	39.496***	3.439	3.445	29.011***	-	10.208***
$\Delta RENER_USE$	5.786*	11.322***	8.451**	8.358**	37.147***	10.676***	-

The values in the table report Chi-square statistics from VEC Granger causality/block exogeneity tests with two degrees of freedom for each analysis. These Chi-square statistics are from Wald-style exclusion tests on the lagged difference terms. Significance levels are marked as * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$.

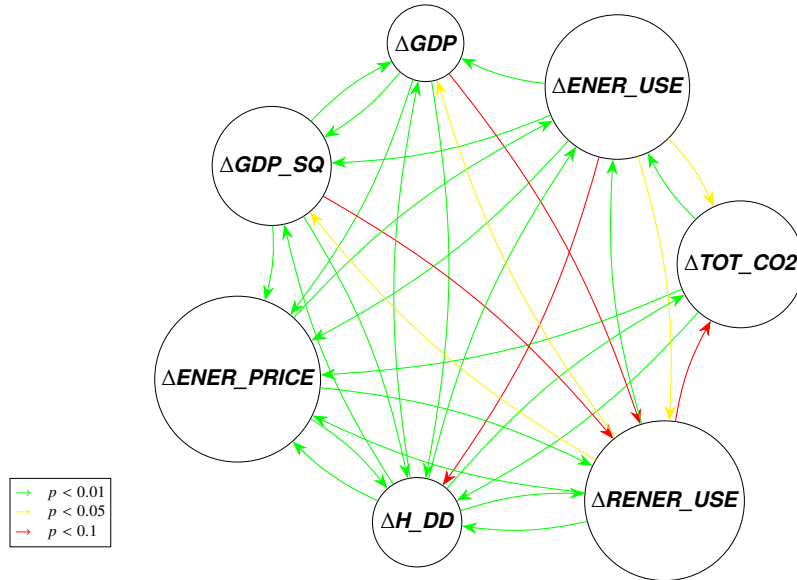


Figure A.1: Causal relationships based on VEC Granger causality tests for the Base Model.
Note: The arrows represent the direction of causality, and the colors indicate the level of statistical significance: green for $p < 0.01$, yellow for $p < 0.05$, and red for $p < 0.1$.



Figure A.2: Generalized Impulse Response Functions for the Base Model.

Table A.2: FMOLS Results for the Base Model.

Variable	Estimate
lnNONRENEWpc	0.885*** (0.043)
lnRENEWpc	-0.051*** (0.008)
lnPRC	-0.011 (0.013)
HDD	0.022*** (0.007)
lnRGDPpc	0.624*** (0.148)
lnGDP2pc	-0.393*** (0.063)

Note: The table presents the estimates and standard errors for the variables impacting $\ln CO2pc$, derived from panel FMOLS regression. *** p<0.01, ** p<0.05, * p<0.1.

B.1 A Note on Residential GHG Emissions

Residential GHG emissions represent a distinct category from other sectoral emissions and include both direct emissions from on-site fuel combustion (e.g., natural gas for heating and cooking) and indirect emissions from electricity consumption generated off-site. These emissions differ fundamentally from those in the transportation, industrial, commercial, and agricultural sectors, each with unique energy usage patterns and emission intensities. This study focuses on residential emissions because net metering policies primarily target household-level renewable energy adoption. The calculation of residential GHG emissions comes from comprehensive data collection by the EIA (U.S. Energy Information Administration, 2020). This process combines household surveys conducted every four years, which gather detailed information about housing characteristics and energy usage patterns, with actual energy supplier billing data. This dual-source approach is then enhanced through engineering-based modeling of various household energy end-uses, such as heating, cooling, and appliances. The EIA then calibrates the model outputs against actual billing records to ensure the modeled estimates align with real-world data. The final step involves calculating residential GHG emissions by combining residential energy consumption data with related emission factors. Direct emissions are determined by multiplying on-site fuel consumption by the U.S. EPA emission factors for specific fuels (U.S. Environmental Protection Agency, 2024). Indirect emissions, resulting from electricity consumption, are calculated by applying regional grid emission factors to the amount of electricity used.

C.0.1 Extended Diagnostic Tests

Univariate and Bivariate Analysis

Figure C.1 shows the subplots for the univariate analysis. In the climate variables group, most features exhibit relatively symmetric distributions, with some variables showing slight skewness. The temperature-related variables appear to be approximately normally distributed, which suggests a balanced range of temperature conditions across counties. The heating degree day variable has an astronomical number of zeros since rice is grown in the southern part of the U.S. where it tends to be warmer the vast majority of the time. We exclude this feature from the models due to lack of variation.

In the soil properties group, the soil pH variables at different depth levels display positively skewed distributions, with a higher concentration of values on the lower end of the pH scale. This suggests that the majority of the soil has relatively low pH values, with fewer instances of high pH levels. The water content variables also exhibit a slight positive skewness, indicating that most soil samples have lower water content, with a smaller proportion of samples having higher water retention capacities. The soil texture variables show an uneven distribution across different texture classes, with classes such as Silt Loam (SiLo) being more prevalent up to 10 cm depth, Loam (Lo) at 30 cm depth, Sand Clay Loam (SaClLo) between 60 and 100 cm depth and Loam around 200 cm depth. In the rice data group, the production and area-related variables are positively skewed, with a long right tail. This suggests the presence of a few large-area rice production sites, while the majority of the locations have lower production and smaller cultivation areas.

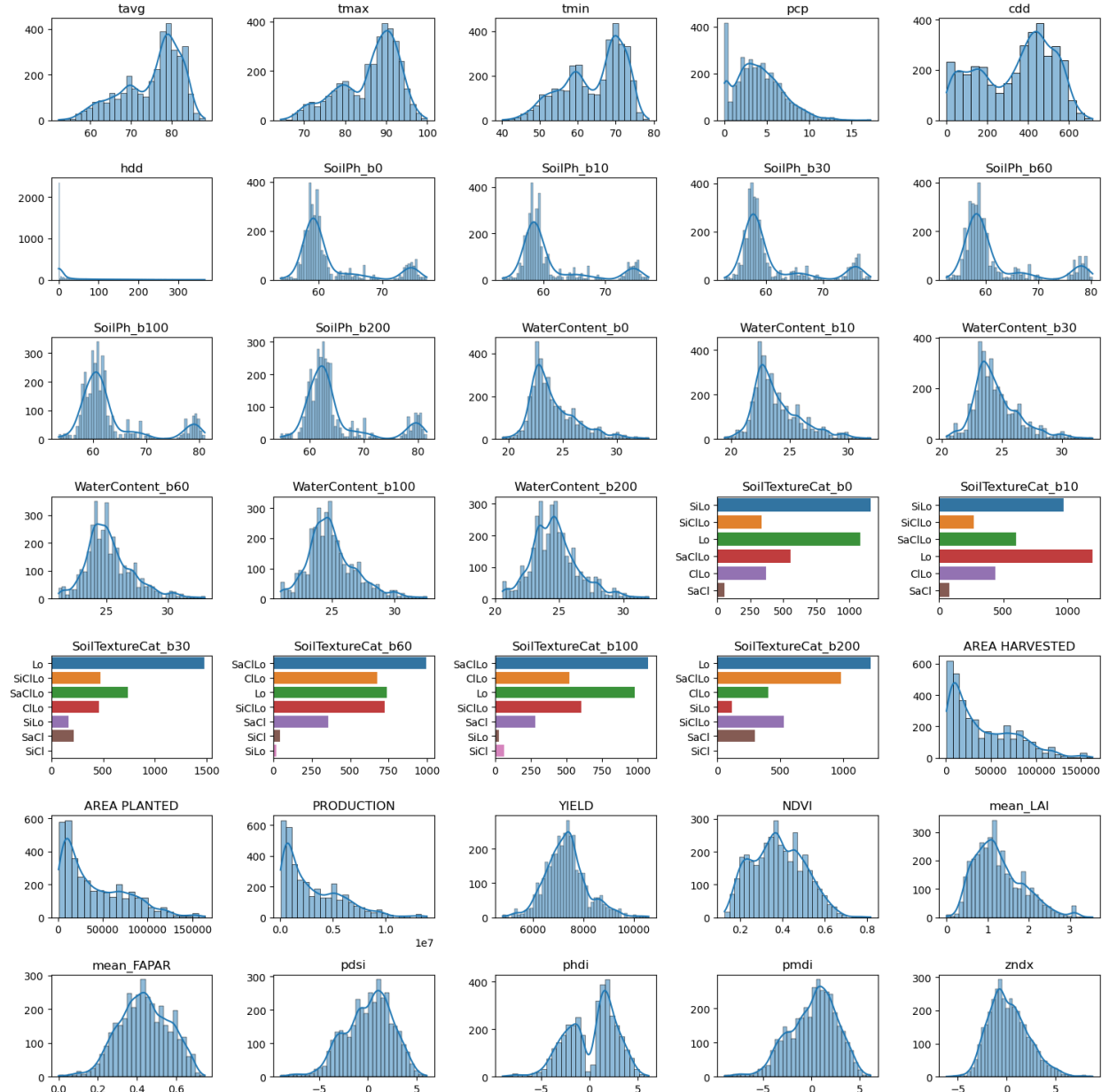


Figure C.1: Matrix of Plots for Univariate Analysis.

Notes: The x-axis for numeric variables represents the range of values observed for each variable, while the y-axis represents the frequency of occurrences of these values in the dataset. For categorical variables, the x-axis corresponds to the number of observations or counts for each category, while the y-axis represents the different categories or classes.

Lastly, in the satellite data group, the NDVI variable exhibits a relatively symmetric distribution, indicating a balanced range of vegetation greenness across the study area. The LAI variable shows a slight positive skewness, suggesting that most locations have lower vegetation density, with fewer instances of high LAI values. On the other hand, the FAPAR variable exhibits a slight negative skewness, indicating that a higher proportion of locations have relatively high canopy absorption, with fewer instances of low FAPAR values. This difference in skewness between LAI and FAPAR highlights the complex relationship between vegetation structure and its ability to absorb photosynthetically active radiation, emphasizing the importance of considering both variables in understanding the dynamics of rice growth and yield formation. The drought indices are fairly normally distributed, except for *phdi*, which is bimodal, likely reflecting predominantly either dry or wet seasons.

Figure C.2 shows the correlation heatmap, which provides a visual representation of the pairwise correlations between the variables. As expected, within the climate variables group, there are strong positive correlations among temperature-related variables and cooling degree days. Precipitation exhibits moderate negative correlations with temperature variables and cooling degree days, while showing positive correlations with drought indices. Heating degree days have negative correlations with temperature variables and cooling degree days.

The soil properties group shows a moderately strong correlation between soil texture and soil pH, and a relatively weak correlation between water content and soil pH. The strong positive correlation between soil texture and water content can be attributed to the fact that soil texture significantly influences the soil's ability to retain water. Intuitively, soils with a higher clay content should exhibit increased water retention due to the smaller particle sizes and greater surface area, while sandy soils should facilitate rapid drainage. The moderately strong correlation between soil texture and soil pH may be due to varying mineral compositions and buffering capacities inherent in different soil textures. For instance, clay-rich soils often have higher cation exchange capacities,

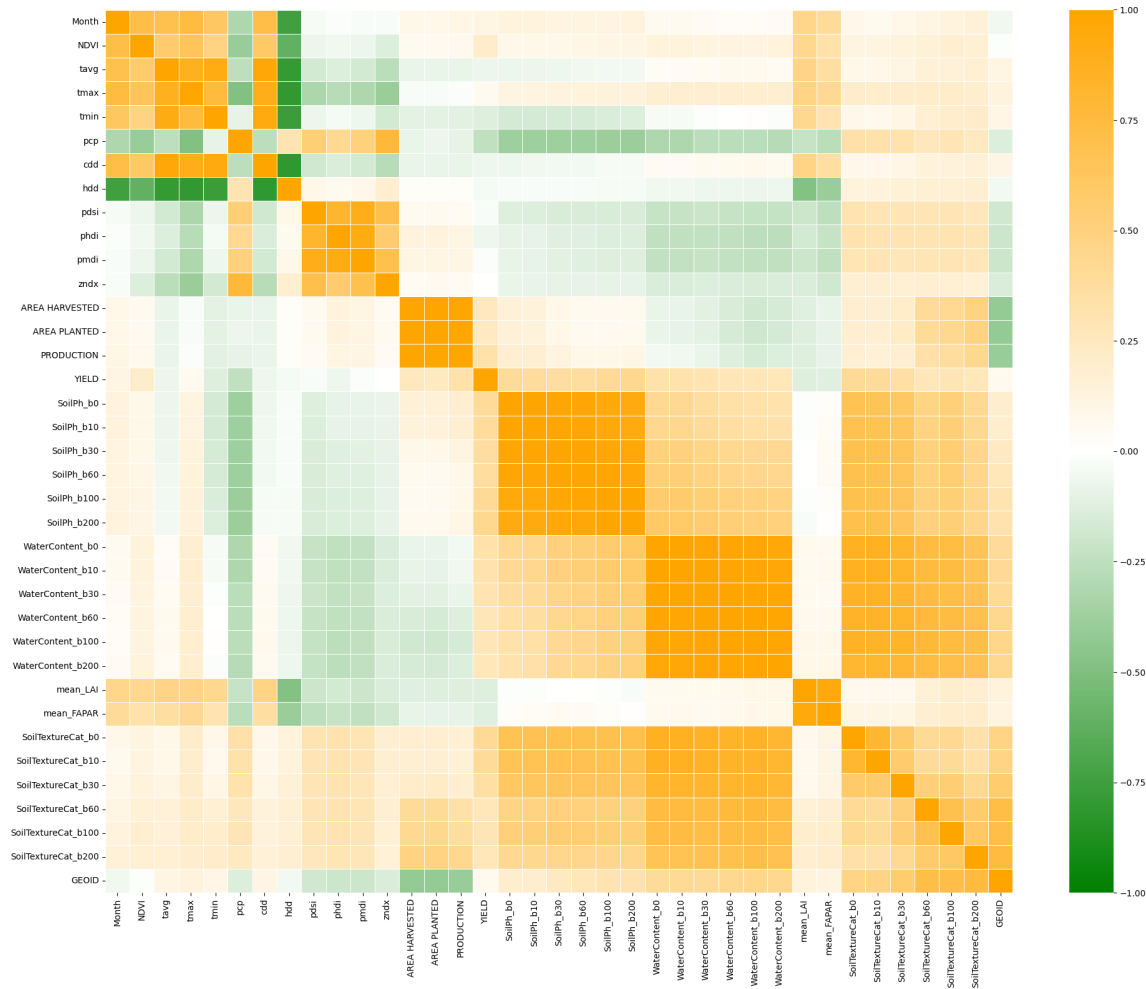


Figure C.2: Correlation Heatmap.

which can influence pH levels.¹ Conversely, the relatively weak correlation between water content and soil pH suggests that while water retention is influenced by soil texture, pH levels are more significantly affected by other factors such as organic matter content.

In the rice data group, the area-related variables and production exhibit very strong positive correlations with each other. Yield, however, has a surprisingly weak positive correlation with the area and production variables, indicating that increases in area harvested and planted do

¹Cation exchange capacity is the soil's ability to hold and exchange positively charged ions. Higher CEC indicates greater nutrient retention and pH buffering capacity, typically found in clay-rich soils due to their larger surface area compared to sandy soils.

not proportionately translate to higher yields, possibly due to variations in farming practices or environmental conditions. We drop the variables area planted and production from the analysis to prevent potential model leaking since these two are used to calculate yield. There is a positive correlation between the satellite-derived variables. NDVI has a moderate positive correlation with LAI and a weaker positive correlation with FAPAR. LAI and FAPAR have a strong positive correlation, suggesting they capture similar aspects of vegetation growth and productivity.

Across the different feature groups, climate variables, particularly temperature-related variables and cooling degree days, have moderate positive correlations with NDVI, LAI, and FAPAR. This implies that favorable temperature conditions are associated with increased vegetation greenness and productivity. Soil pH variables show weak to moderate positive correlations with yield, indicating that higher soil pH levels might be associated with better crop yields, potentially due to optimal nutrient availability. Lastly, water content variables have weak positive correlations with NDVI, LAI, and FAPAR, suggesting that, although higher soil moisture content supports vegetation growth, other factors likely play more significant roles.

Outlier Detection and Handling

The boxplots in Figure C.3 highlight several significant outliers across various groups. In the climate variables group, the heating degree days variable contains a lot of zeros that could distort the analysis results. Consequently, we drop it as previously mentioned.

For the remaining variables, a few outliers are present but are not overly concerning given the number of observations in the dataset and the values of these outliers relative to their interquartile ranges. To address these outliers and avoid magnitude issues (such as differences between Yield and NDVI values), all features are rescaled to standardize their magnitudes. This normalization ensures that outliers do not disproportionately affect the analysis. Additionally, tree-based ML methods, such as XGBoost, robustly handle outliers due to their regularization and inherent resistance to the influence of extreme values.

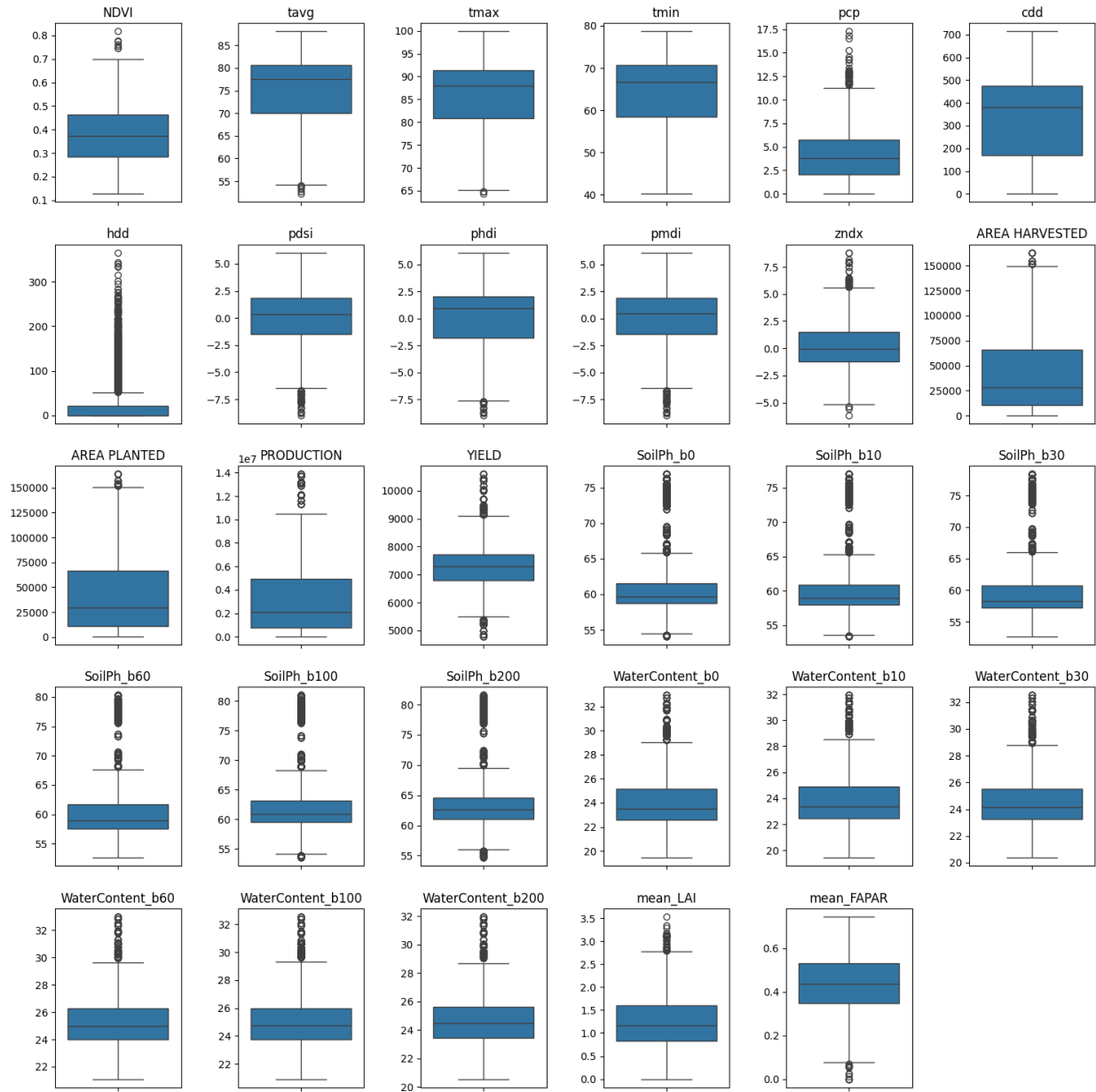
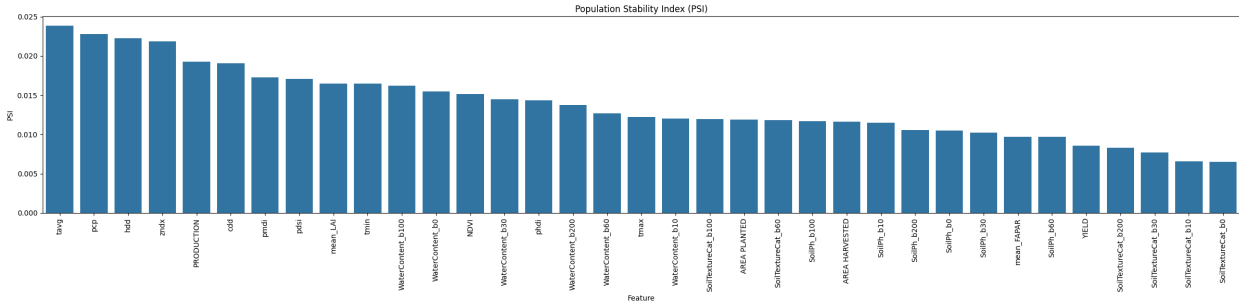


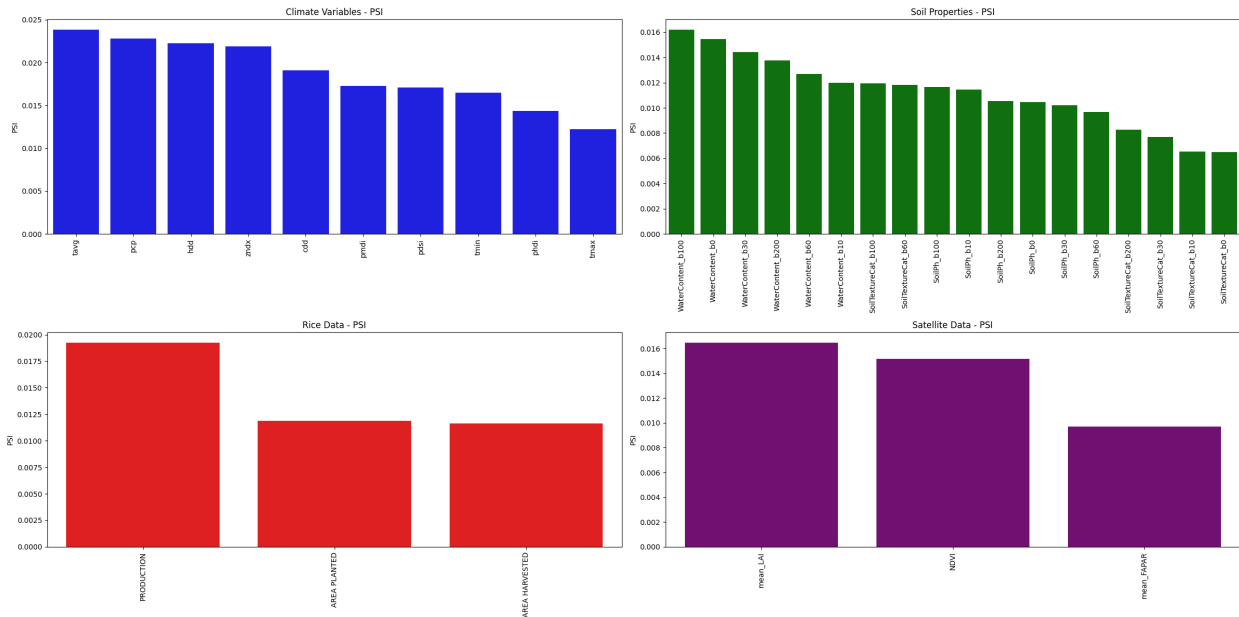
Figure C.3: Boxplot-based Outlier Detection.

Data Drift Test

The results of the data drift test, as illustrated in Figure C.4, reveal varying degrees of marginal distribution drift across the features. The PSI is used to quantify these differences, with higher values indicating greater data drift.



(a) Data Drift Tests - All Features



(b) Data Drift Tests - By Category

Figure C.4: Data Drift Tests.

The feature *avg* exhibits the highest PSI value of 0.024, followed by *pcp* (0.023) and *hdd* (0.022). These relatively higher PSI values suggest significant changes in their marginal distributions

between the training and testing datasets. As mentioned before, such drifts can impact the model's performance and its ability to generalize effectively.

Notably, climate-related variables are among those with higher PSI values, indicating substantial changes in weather patterns or measurement conditions over time. The *production* variable also shows a notable drift (0.019), which might reflect changes in agricultural practices, yield improvements, or external factors affecting production levels. Conversely, soil texture features show the least drift, implying stable distributions over time. Soil texture categories generally remain consistent, which could be due to the inherent stability of soil characteristics compared to more dynamic environmental or climatic factors.