OECONOMIA
COPERNICANA

VOLUME 16 ISSUE 1 2025

p-ISSN 2083-1277, e-ISSN 2353-1827
https://oeconomiacopernicana.com/

Received: 29.08.2025; Revised: 13.09.2025, Accepted: 26.10.2025, Published Online: 21.11.2025

GREEN GROWTH OR EMISSIONS LOCK-IN? A HYBRID ECONOMETRIC–AI FORECAST OF CHINA'S INDUSTRIAL CO_2 PATHWAYS

1- Bouklia Nawel 2- Chehaoui Lidia 3-Touitou Mohammed 4-Laib Yacine 5- Bouklia Mohamed

1- Senior Lecturer-A-, Faculty of Economics and Business, University of Algiers 3, Algiers. Algeria.
E-mail: nawel.bouklia123@gmail.com

2- Doctoral of Management Sciences, Morsli Abdullah University, Tipaza (MAUCT), . Algeria.
E-mail: lidiachehaoui22@gmail.com

3- Professor of Applied Statistics and Economics, Faculty of Economics and Business, University of Algiers 3, Algiers. Algeria.
E-mail: touitoutouitou@yahoo.fr Orcid: https://orcid.org/0000-0002-4442-7888

4- Professor of Economics, Faculty of Economics and Business, University of Algiers 3, Algiers. Algeria. E-mail: m.laibyacine18@gmail.com

5- Professor of Economics, Faculty of Economics and Business, University of Algiers 3, Algiers. Algeria. E-mail: boukliamohamed@gmail.com

Abstract: This paper examines whether China's industrial sector is advancing toward green growth or facing an emissions lock-in by integrating econometric analysis with artificial intelligence–based forecasting. Using panel data from 30 provinces between 2000 and 2023, the study applies a System GMM model alongside machine learning algorithms—Random Forest, Gradient Boosting, and Support Vector Regression—to identify the key drivers and predict future CO_2 trajectories. The results reveal that energy intensity and industrial output significantly raise emissions, while renewable-energy deployment and R&D investment exert strong mitigating effects. Among AI models, Random Forest achieves the highest predictive accuracy ($R^2 = 0.86$), validating its effectiveness for nonlinear environmental systems. Scenario forecasts for 2025–2030 indicate that moderate efficiency gains and renewable expansion could lower emissions by nearly 10 percent compared with the baseline path. Overall, the findings underscore that improving energy efficiency, stimulating innovation, and addressing regional disparities are essential to prevent long-term carbon lock-in and to align industrial development with China's 2060 carbon-neutrality goal.

Keywords: Green growth, Emissions lock-in, Energy efficiency, Machine learning, Carbon forecasting, Panel data

1. Introduction

China stands at the forefront of the global climate transition, simultaneously the world's largest carbon emitter and a rapidly transforming economy pursuing ambitious sustainability goals. As of 2022, China accounted for roughly 30 percent of global CO_2 emissions, with the industrial sector responsible for over 60 percent of national emissions and more than 65 percent of total energy consumption (IEA, 2022). In response to growing international pressure and domestic environmental challenges, the government has pledged to peak carbon emissions by 2030 and achieve carbon neutrality by 2060, making these objectives central to both economic and environmental policy agendas.

The industrial sector is pivotal in this transition. Decades of industrialization—dominated by heavy manufacturing, coal-fired power, and energy-intensive processes—have driven China's economic rise but also entrenched structural dependence on fossil fuels (Shan et al., 2018; Zheng et al., 2022). This duality creates a central policy dilemma: how can China sustain industrial competitiveness while decoupling growth from emissions? The answer lies in understanding the balance between green growth, which promotes sustainable expansion through innovation and efficiency, and emissions lock-in, which perpetuates high-carbon pathways through technological and institutional inertia.

Green growth represents a development trajectory that reconciles economic and environmental objectives by fostering technological innovation, energy efficiency, and structural transformation (OECD, 2020). In contrast, emissions lock-in describes a condition in which carbon-intensive infrastructures, path-dependent technologies, and entrenched policy regimes constrain the shift toward low-carbon production systems (Unruh, 2000; Seto et al., 2016; Guan et al., 2021). The coexistence of these opposing dynamics defines the uncertainty surrounding China's industrial decarbonization: whether the nation's current trajectory signifies genuine green transformation or a temporary stabilization within a locked-in, fossil-based system (Li & Sun, 2018).

Empirical assessment of this tension requires analytical tools capable of distinguishing long-term structural persistence from emerging adaptive change. Traditional econometric approaches have provided valuable insights into average effects of energy use and industrial activity (Wang et al., 2020; Wenjuan et al., 2015), yet they often struggle to capture the nonlinear, heterogeneous, and region-specific dynamics of a vast economy such as China's. Conversely, recent advances in artificial intelligence (AI) and machine learning (ML) enable flexible, high-accuracy modeling of complex environmental systems (Faruque et al., 2022; Zhao et al., 2023) but are limited in causal interpretation.

To bridge this divide, the present study develops a hybrid econometric–AI framework that integrates dynamic panel regression using the System GMM estimator with advanced machine learning algorithms—Random Forest, Gradient Boosting, and Support Vector Regression. This combined approach allows the simultaneous examination of causal mechanisms and nonlinear predictive relationships driving China's industrial CO_2 emissions. Using provincial data for 2000–2023, the analysis identifies the principal determinants of emissions, evaluates model performance, and forecasts possible trajectories through 2030. By linking the theoretical debate on green growth versus emissions lock-in with a data-driven, policy-oriented methodology, the paper seeks to clarify whether China's industrial sector is converging toward sustainable decarbonization or remains constrained by structural carbon dependence. The findings provide actionable insights for policymakers seeking to align industrial transformation with the nation's 2060 carbon-neutrality commitment and the priorities of the 14th Five-Year Plan.

2. Literature Review

Understanding the relationship between industrial development, energy efficiency, and carbon emissions has become a central area of research in the context of global climate change and sustainable economic transitions. The Chinese industrial sector, in particular, has drawn scholarly attention due to its scale, carbon intensity, and pronounced regional disparities. This literature review synthesizes key themes from existing studies, highlighting advances, limitations, and the specific contributions of this research.

2.1. Green Growth vs. Emissions Lock-in

The theoretical foundation for examining environmental outcomes alongside economic performance often begins with the Environmental Kuznets Curve (EKC), which suggests an inverted-U relationship between income and environmental degradation (Grossman & Krueger, 1995). While some empirical studies in China have observed EKC-like patterns at the national level (Alil & Feridun, 2011; Zhang & Cheng, 2009), the provincial dynamics often deviate due to heterogeneity in industrial structure, technological capacity, and policy enforcement.

The concept of emissions lock-in—where legacy infrastructure, path-dependent policies, and fossil fuel dependency delay decarbonization—is increasingly applied to industrial economies (Unruh, 2000; Seto et al., 2016). In China, such lock-in is reinforced by investments in coal-intensive industries, particularly in inland regions, despite rapid growth in renewable energy and green tech (Unruh, G. C. (2000; Guan et al., 2018). Research has shown that institutional inertia, subsidies for fossil fuel industries, and uneven access to clean technologies perpetuate lock-in risks (Wang et al., 2023; Yuan et al., 2021).

In contrast, green growth advocates an economic transformation that integrates environmental sustainability, efficiency improvements, and innovation-led productivity (OECD, 2020). While China has made significant progress through its Five-Year Plans and dual-carbon framework, debate persists over whether current industrial trends represent genuine sustainable decoupling or transitional stabilization.

2.2. Industrial Energy Efficiency and Emissions Trends in China

Energy intensity—energy consumed per unit of output—remains a key determinant of industrial CO_2 emissions. Empirical evidence shows that energy-efficiency improvements have played a dominant role in reducing emissions since the early 2000s (Wang et al., 2020; Ma & Yang, 2022). However, these effects vary widely across regions. Coastal provinces such as Zhejiang and Guangdong exhibit rapid efficiency gains, while interior, coal-dependent provinces remain less efficient (Zheng et al., 2022).

Beyond efficiency, transformations in industrial structure—shifting from heavy industry to high-tech and service sectors—and rising investment in renewables and R&D have become co-determinants of emission reductions (Han & Zhou, 2022; Hu et al., 2024). Yet, several studies caution that efficiency gains may be offset by the rebound effect, whereby increased production scale counteracts technological improvements.

2.3. Machine Learning in Carbon Forecasting and Energy Modeling

Machine learning has recently emerged as a powerful complement to traditional econometric analysis, enabling researchers to capture nonlinear relationships, address multicollinearity, and manage high-dimensional data. Algorithms such as Random Forest (RF), Gradient Boosting Machines (GBM), and Support Vector Regression (SVR) achieve higher forecasting accuracy than conventional regression or ARIMA-type models (Faruque et al., 2022; Benti et al., 2023; Li and Sun., 2021).

Within the Chinese context, Zhou et al. (2023) used Random Forest and XGBoost to forecast urban CO_2 emissions with high precision, while Hou et al. (2023) applied SVR to project national emissions based on energy and macroeconomic indicators. Similarly, Sun et al. (2023) combined machine-learning models with geospatial and socio-economic datasets to enhance provincial-level emission forecasting. These studies highlight ML's strength in predictive modeling, though they often lack causal interpretability and explicit policy integration.

Despite their technical success, many ML-based studies remain descriptive—focusing on prediction accuracy rather than structural drivers of emissions. Only a few have attempted to combine ML with econometric approaches, an integration that can provide both explanatory depth and predictive strength. This methodological hybridization is particularly crucial in China, where industrialization, energy transition, and regional development are complex, heterogeneous, and policy-driven (Zhao et al., 2023).

2.4. Gaps in the Literature and Study Contribution

Despite the extensive body of research addressing energy efficiency and carbon emissions in China, several important gaps persist in the existing literature. One of the most evident limitations concerns the lack of integration between econometric and machine-learning approaches. Most prior studies have relied exclusively on either traditional econometric models or standalone machine-learning algorithms, thereby overlooking the potential analytical advantages of hybrid frameworks that combine causal inference with predictive power.

A second shortcoming lies in the insufficient attention to regional heterogeneity. Much of the current research focuses on national-level assessments, which often obscure substantial disparities between coastal provinces—where economic diversification and cleaner industrial technologies have advanced rapidly—and inland regions that continue to depend heavily on energy-intensive manufacturing. These spatial asymmetries are crucial for understanding the uneven pace of China's industrial transition and for designing regionally differentiated policy interventions.

A third limitation concerns the weak linkage between emission-forecasting research and China's explicit climate-policy targets, notably the 2030 carbon-peaking and 2060 carbon-neutrality goals. The absence of this alignment reduces the practical relevance of many existing studies for policymakers seeking to evaluate progress or recalibrate strategies within the framework of the country's long-term decarbonization agenda.

The present study directly addresses these gaps by integrating System GMM econometric modeling with advanced machine-learning techniques—specifically Random Forest, Gradient Boosting, and Support Vector Regression—to analyze both causal relationships and predictive performance. Using a comprehensive provincial-level panel dataset covering 2000–2023, the study generates regionally disaggregated insights that reflect China's industrial diversity and evolving policy landscape. Moreover, by situating the empirical findings within the broader debate on green growth versus emissions lock-in, the research provides policy-relevant conclusions that contribute to the design of targeted industrial-decarbonization strategies consistent with the 14th Five-Year Plan and the 2060 carbon-neutrality commitment.

Building on this review, it becomes clear that the dynamics of industrial CO_2 emissions in China are shaped by both structural and technological factors that evolve over time. Previous studies have tended to treat these dimensions separately—econometric models emphasizing causality and computational models emphasizing prediction—yet neither approach alone adequately captures their interaction. To overcome this limitation, the present research adopts a hybrid empirical design that merges the System Generalized Method of Moments (System GMM), which addresses endogeneity, persistence, and heterogeneity, with machine-learning algorithms (Random Forest, Gradient Boosting, and SVR) capable of modeling nonlinear relationships and enhancing forecasting precision.

This integrated methodological framework thus forms a robust foundation for the subsequent empirical analysis, enabling a comprehensive assessment of whether China's industrial transition is advancing toward sustainable green growth or remains constrained by structural emissions lock-in.

3. Methodology

This study employs a hybrid methodological framework that integrates panel econometric modeling with advanced machine learning (ML) techniques to analyze and forecast industrial CO_2 emissions in China. This combination enables both causal inference and high-accuracy prediction, making it suitable for evidence-based climate policy design.

3.1 Data and Variables

We construct a balanced panel dataset for 30 Chinese provinces over the period 2000–2023. Data is sourced from the Carbon Emission Accounts and Datasets (CEADs), the China Energy Statistical Yearbook, and the National Bureau of Statistics (NBS). All monetary values are deflated to constant prices.

Variable Description and Sources

Variable Name	Description	Measurement Unit	Source
CO_2 Emissions	Industrial CO_2 emissions	Million tons	CEADs, Provincial Yearbooks
Energy Intensity (EI)	Energy consumption per unit of industrial output	Tons of coal equivalent/RMB	NBS
Industrial Output (IO)	Gross value of industrial production	Billion RMB (constant)	NBS
GDP per capita (GDPpc)	Provincial GDP per capita	RMB per person	NBS
Renewable Energy Share (RE)	Share of renewables in provincial energy mix	%	NBS, Energy Yearbook
R&D Expenditure (RD)	Provincial R&D investment	Billion RMB	NBS
Urbanization Rate (UR)	Urban population ratio	%	NBS

All variables were tested for unit roots using Levin–Lin–Chu (LLC) and Im–Pesaran–Shin (IPS) tests.

Non-stationary series were differenced or log-transformed where necessary.

Data Preprocessing and Quality Control

Before estimation and model training, all variables underwent systematic preprocessing to ensure internal consistency and comparability across provinces and years. Minor missing values (one- to two-year gaps) were imputed by linear interpolation and cross-checked with the China Statistical Yearbook and NBS provincial datasets. Outliers—primarily from small inland provinces with highly volatile industrial outputs—were winsorized at the 1st and 99th percentiles to limit their influence while preserving overall variance.

For econometric estimation, the stationarity of each variable was confirmed through the LLC and IPS tests, and non-stationary series were transformed via first differencing or natural logarithms. For the machine-learning phase, continuous variables were normalized using min–max scaling within the [0, 1] range to ensure comparable magnitudes and stable model convergence. The dataset was randomly partitioned into an 80 % training set and 20 % testing set for out-of-sample validation.

Model performance was evaluated using three complementary metrics: R^2, Root Mean Square Error (RMSE), and Mean Absolute Error (MAE).

- R^2 measures explanatory power and overall fit.
- RMSE penalizes large deviations, highlighting bias and variance.
- MAE provides an intuitive measure of average predictive error, less sensitive to outliers. The combined use of these indicators ensures a transparent and balanced assessment of model accuracy and reliability across econometric and machine-learning frameworks.

Prior to estimation, a comprehensive data preprocessing procedure was undertaken to ensure the reliability and consistency of the empirical analysis. Missing observations were addressed using a combination of linear interpolation for short gaps and mean substitution within provinces when missingness did not exceed 5 percent of total entries. Outliers were detected through standardized residual analysis and winsorized at the 1st and 99th percentiles to mitigate the influence of extreme values while preserving overall data variability. All continuous variables were subsequently normalized using z-score standardization to harmonize measurement scales and improve model convergence. This preprocessing framework ensured that both econometric estimation and machine-learning training were based on a balanced and statistically robust dataset, minimizing noise and enhancing comparability across provinces and years.

3.2 Econometric Strategy: System GMM

o estimate the dynamic relationship between energy efficiency and CO_2 emissions while addressing endogeneity, heterogeneity, and autocorrelation, we use the System Generalized Method of Moments (System GMM) estimator (Arellano & Bover, 1995; Blundell & Bond, 1998). This estimator is preferred for panel data with large N (provinces) and small T (years) and when regressors are endogenous (Roodman, 2009). It improves efficiency over difference GMM by incorporating equations in both levels and first differences, using internal instruments.

Rationale for Method Selection

The choice of the System GMM framework is guided by both the characteristics of the dataset and the theoretical objectives of the study. Industrial CO_2 emissions exhibit strong dynamic persistence, as current emission levels depend on past industrial and energy patterns. Moreover, key explanatory variables such as energy intensity, industrial output, and R&D expenditure are potentially endogenous, since high emissions may themselves trigger changes in efficiency or innovation policies. System GMM effectively accounts for these feedback effects by using lagged internal instruments, thereby yielding consistent and unbiased parameter estimates even under endogeneity.

This estimator also aligns with the study's aim to test whether China's industrial system is evolving toward green growth or remains trapped in emissions lock-in. By capturing both the temporal inertia (via lagged dependent variables) and **structural effects** of industrial and energy variables, System GMM allows us to quantify the degree of carbon persistence that defines lock-in dynamics. Its integration into the hybrid econometric–AI framework provides the causal foundation upon which the machine learning models subsequently build predictive and nonlinear insights.

Baseline Equation

The empirical model employed in this study follows a dynamic specification estimated using the System Generalized Method of Moments (System GMM). The baseline equation is expressed as:

$$\mathrm{CO2}_{it} = \alpha \mathrm{CO2}_{it-1} + \beta_1 \mathrm{EI}_{it} + \beta_2 \mathrm{IO}_{it} + \beta_3 \mathrm{GDP}_{it} + \beta_4 \mathrm{RE}_{it} + \beta_5 \mathrm{RD}_{it} + \beta_6 \mathrm{UR}_{it} + \mu_i + \lambda_t + \varepsilon_{it}$$

Where μ_i represents unobserved province-specific effects, λ_t denotes time-fixed effects, and ε_{it} is the idiosyncratic error term.

To ensure the robustness and validity of the estimations, several diagnostic tests were conducted. The Arellano–Bond AR(1) and AR(2) tests were used to detect potential serial correlation in the residuals, while the Hansen J-test examined the overall validity of the instrumental variables. In addition, the Difference-in-Hansen test was applied to verify the consistency of the instrument subsets used in the level equations.

System GMM has been widely recognized in energy and environmental economics for its capacity to address challenges of endogeneity, dynamic persistence, and unobserved heterogeneity in panel data models (Bond et al., 2001; Ran et al., 2023). Its application in this study therefore provides a reliable and statistically consistent framework for analyzing the dynamic relationship between energy efficiency, industrial activity, and CO_2 emissions across Chinese provinces.

3.3 Machine Learning Models

To enhance predictive performance and capture complex nonlinear relationships, this study employs three complementary machine learning (ML) models: Random Forest (RF), Gradient Boosting Machines (GBM), and Support Vector Regression (SVR).

Rationale for Model Selection and Research Alignment

The selection of these models reflects both the characteristics of the dataset and the study's dual objectives—understanding structural determinants of emissions and improving forecasting accuracy for policy scenarios. The provincial dataset exhibits complex interactions among economic, energy, and innovation variables, making nonlinear and high-dimensional algorithms particularly suitable. RF and GBM, as ensemble learning methods, capture intricate cross-variable relationships and mitigate overfitting, while SVR effectively models smooth nonlinear patterns using kernel functions. Together, these algorithms align with the study's overarching framework: System GMM addresses causal persistence and lock-in dynamics, while ML models reveal adaptive nonlinear behaviors consistent with emerging green-growth trajectories.

The Random Forest model, introduced by Breiman (2001), is an ensemble learning algorithm that constructs multiple decision trees using bootstrap aggregation and averages their outcomes to improve generalization accuracy while minimizing overfitting. Gradient Boosting Machines, following Friedman's (2001) formulation, build trees iteratively, where each successive tree corrects the residual errors of the previous one, ultimately minimizing the overall loss function. Support Vector Regression, developed by Smola and Schölkopf (2004), operates by mapping input data into a high-dimensional feature space through kernel functions, allowing for the modeling of complex, nonlinear relationships between variables.

Model Training and Evaluation Procedures

Model training and evaluation followed a consistent procedure to ensure robust predictive validity. The dataset was divided into two subsets, with 80 % used for training and 20 % reserved for testing. Model hyperparameters were optimized through a Grid Search procedure with cross-validation to identify the best-performing configurations for each algorithm.

To assess model performance, three complementary statistical indicators were employed: the Coefficient of Determination (R^2), the Root Mean Square Error (RMSE), and the Mean Absolute Error (MAE). The R^2 value measures the proportion of variance in the observed data that is explained by the model, serving as an indicator of overall goodness of fit. The RMSE captures the square root of the average squared differences between predicted and actual values, thereby penalizing large deviations and revealing both model bias and dispersion. The MAE, in turn, reflects the average magnitude of prediction errors, offering a scale-independent and easily interpretable measure of predictive accuracy. Taken together, these metrics provide a comprehensive assessment of both explanatory power and forecasting precision across econometric and machine-learning models.

The combined use of R^2, RMSE, and MAE provides a transparent and balanced assessment of both explanatory power and predictive reliability, enabling comparison between econometric and ML models.

An additional advantage of tree-based models such as Random Forest and Gradient Boosting lies in their ability to generate feature importance scores, which quantify the relative contribution of each explanatory variable to predictive performance. This interpretive capability allows identification of the most policy-relevant determinants of industrial CO_2 emissions—such as energy intensity, R&D expenditure, and renewable-energy share—thus bridging the gap between data-driven modeling and policy-oriented insight.

Overall, integrating machine learning techniques complements the econometric approach by improving forecast accuracy and accommodating nonlinearities that traditional models may overlook. Such hybridization is particularly crucial for environmental forecasting, where complex interdependencies among economic, technological, and policy factors require models that combine statistical rigor with computational flexibility (Faruque et al., 2022; Zhao et al., 2023).

3.4 Forecasting CO_2 Emissions: 2025–2030

To assess the medium-term trajectory of industrial CO_2 emissions in China and provide insights relevant to national climate policy, this study develops forward-looking emission forecasts consistent with the country's 2030 carbon-peaking and 2060 carbon-neutrality objectives. The forecasting exercise relies on the machine-learning framework described earlier, emphasizing the Random Forest (RF) algorithm and complemented by Gradient Boosting Machines (GBM) and Support Vector

Regression (SVR) to ensure comparative robustness and model diversity.

The Random Forest model was trained on an extensive provincial-level panel dataset covering 2000–2023, which includes the key determinants of industrial emissions—energy intensity, industrial output, renewable-energy share, R&D expenditure, GDP per capita, and urbanization rate. After hyperparameter optimization through cross-validation, the final model configuration achieved high predictive accuracy on the test set and was subsequently used to simulate annual industrial CO_2 emissions from 2025 to 2030 under two alternative policy scenarios.

The first scenario, the baseline or business-as-usual (BAU) pathway, assumes the continuation of historical trends in industrial activity, energy intensity, and renewable-energy adoption. It excludes major policy interventions or structural transformations beyond those observed during 2015–2023. This projection is generated by extrapolating the compound annual growth rates (CAGR) of the main explanatory variables, thus offering a benchmark trajectory that reflects policy inertia.

The second scenario, the policy-enhancement pathway, represents a proactive yet realistic decarbonization effort aligned with China's 14th Five-Year Plan and broader national energy-transition strategies. In this configuration, the model incorporates an assumed 10 percent improvement in industrial energy efficiency and a 15 percent increase in the renewable-energy share within the industrial mix by 2030. These assumptions are calibrated through CAGR-based extrapolations consistent with recent empirical and policy-oriented forecasting studies (e.g., Hu et al., 2024).

The comparative design of these scenarios provides a rigorous analytical lens through which to evaluate the implications of maintaining existing industrial patterns versus implementing enhanced decarbonization policies. Moreover, by explicitly accounting for nonlinear interactions among predictors and by analyzing feature-importance rankings, the Random Forest framework offers a nuanced understanding of how structural, economic, and technological factors jointly influence China's future emissions trajectory. This hybrid modeling approach thus demonstrates the added value of combining econometric foundations with machine-learning precision in exploring both the technical and policy dimensions of industrial decarbonization within a complex, regionally diverse economy.

3.5 Ethical and Data-Governance Considerations

All datasets used in this research were obtained from publicly accessible and officially recognized sources, including the Carbon Emission Accounts and Datasets (CEADs), the China Energy Statistical Yearbook, and the National Bureau of Statistics (NBS). These data are aggregated at the provincial level, ensuring that no individual, firm-level, or confidential information is included in the analysis. Consequently, the study raises no ethical concerns related to privacy or personal data.

The research follows established principles of academic integrity, transparency, and reproducibility. Other scholars document all preprocessing steps, transformations, and model specifications in detail to enable replication. Statistical analyses and machine learning computations were conducted in a secure, offline environment to maintain data integrity and prevent unauthorized access or alteration.

In line with open-science best practices, the study relies solely on verified secondary sources that comply with Chinese and international data-governance standards. Forecasting results and model outputs are presented in an aggregated form and used exclusively for academic and policy-oriented purposes. No external funding agency or third party influenced the research design, data interpretation, or policy implications.

Through these ethical and governance measures, the study ensures methodological transparency, responsible data usage, and full adherence to the ethical standards expected in empirical environmental and economic research.

4. Empirical Results, Forecasting, and Interpretation

This section presents a comprehensive analysis of the determinants of industrial CO_2 emissions across Chinese provinces using a hybrid empirical approach. First, the System GMM estimator is applied to capture the dynamic nature and causal effects of energy efficiency and related variables. Second, Machine Learning (ML) models are evaluated for predictive performance and forecasting. Finally, emissions trajectories for 2025–2030 are projected under baseline and policy-enhanced scenarios.

4.1 System GMM Estimation Results

Table 1 presents the results of the dynamic panel estimations using the System GMM approach. All major explanatory variables, with the exception of urbanization, are statistically significant and consistent with theoretical expectations.

Table 1: Dynamic Panel System GMM Estimation Results

Variable	Coefficient	Std. Error	z-value	Significance
CO_2 Emissions (t−1)	0.521	0.041	12.71	***
Energy Intensity (EI)	0.278	0.035	7.94	***
Industrial Output (IO)	0.166	0.029	5.72	***
GDP per Capita (GDPpc)	−0.088	0.031	−2.84	**
Renewable Energy Share	−0.139	0.038	−3.66	***
R&D Expenditure	−0.104	0.030	−3.47	***
Urbanization Rate (UR)	−0.031	0.026	−1.19	n.s.

*Levels of statistical significance: (***$p < 0.01$), (*$p < 0.05$), (n.s. = not significant)

Source: Authors Elaboration

The results indicate a strong dynamic persistence in industrial CO_2 emissions across China's provinces, as evidenced by the positive and significant coefficient of the lagged dependent variable. This finding reflects the path-dependent nature of emissions, consistent with the concept of carbon lock-in arising from legacy infrastructure and energy systems (Seto et al., 2016; Unruh, 2000; Guan et al., 2021). Energy intensity is found to be positively and significantly related to emissions, underscoring inefficient energy use as a central driver of carbon output, particularly in coal-dependent regions (Liu et al., 2023; Gao et al., 2021). Industrial output also exhibits a positive association with emissions, confirming the presence of the scale effect whereby economic expansion, in the absence of clean-technology adoption, contributes to environmental degradation (Shan et al., 2018).

In contrast, GDP per capita displays a small but statistically significant negative coefficient, suggesting that wealthier provinces are beginning to experience weak decoupling of emissions from economic growth, a pattern broadly consistent with the Environmental Kuznets Curve hypothesis (Jalil &

Feridun, 2011; Zhang & Cheng, 2009). Both the renewable-energy share and R&D expenditure variables exhibit significant negative effects, highlighting the importance of clean-energy expansion and innovation investment in reducing industrial CO_2 emissions (Meng et al., 2023; Hu et al., 2024). The coefficient for urbanization is not statistically significant, which may reflect heterogeneous patterns of development across provinces—some urban areas remain manufacturing-oriented while others are transitioning toward service-based, low-carbon economies. Collectively, these results emphasize that energy intensity, industrial output, renewable deployment, and innovation are the key levers for emission reduction in China's industrial sector and should therefore form the foundation of regionally targeted climate policies.

Model-diagnostic tests summarized in Table 2 confirm the robustness of the System GMM estimation. The Arellano–Bond AR(1) statistic (–2.75, $p = 0.006$) indicates the expected first-order autocorrelation resulting from the differencing of the dependent variable, while the AR(2) statistic (0.94, $p = 0.346$) is not significant, thereby satisfying the requirement of no second-order serial correlation in the residuals (Arellano & Bond, 1991). The Hansen J-test of over-identifying restrictions ($p = 0.228$) demonstrates that the instrument set is valid and not over-fitted, and the Difference-in-Hansen test ($p = 0.211$) further confirms the appropriateness of the instruments in the level equations (Roodman, 2009). Taken together, these diagnostics validate the statistical specification and confirm that neither endogeneity nor instrument proliferation undermines the reliability of the estimated parameters.

Beyond the statistical relationships, the results reveal distinct structural mechanisms driving emission dynamics across China's provinces. Provinces along the eastern and southern coasts—such as Jiangsu, Zhejiang, and Guangdong—show a gradual decoupling of emissions from industrial output, reflecting the diffusion of cleaner technologies, higher innovation intensity, and stricter enforcement of environmental standards. Conversely, provinces in the interior, including Shanxi, Inner Mongolia, and Gansu, remain characterized by high energy intensity and heavy reliance on coal-based industries, illustrating a persistent carbon lock-in. This regional heterogeneity suggests that China's industrial transition is uneven: while advanced provinces approach the early stages of green growth, others remain constrained by structural and resource-based inertia. These findings highlight the importance of spatially differentiated climate and industrial policies tailored to regional development profiles.

Table 2: GMM Model Diagnostic Summary

Test	Statistic	p-value	Interpretation
Arellano-Bond AR(1)	-2.75	0.006	First-order autocorrelation (expected)
Arellano-Bond AR(2)	0.94	0.346	No second-order autocorrelation
Hansen J-test	18.37	0.228	Instruments are valid
Difference-in-Hansen	6.92	0.211	Instrument subsets valid

Source: Authors Elaboration

The absence of problematic multicollinearity is verified by the Variance Inflation Factor (VIF) results shown in Table 3. All explanatory variables exhibit VIF scores well below the conventional threshold of 10, with the highest value of 3.45 for energy intensity, followed by 2.98 for industrial output and 2.56 for the renewable-energy share. Lower scores for GDP per capita (1.87), R&D expenditure (2.14), and urbanization (1.67) confirm that collinearity among predictors is moderate and non-problematic.

These outcomes indicate that the estimated coefficients are statistically stable and the standard errors are not inflated, in accordance with established diagnostic standards for panel-data models (Wooldridge, 2010).

Table 3: VIF – Multicollinearity Check

Variable	VIF Score
Energy Intensity	3.45
Industrial Output	2.98
GDP per Capita	1.87
Renewable Energy Share	2.56
R&D Expenditure	2.14
Urbanization Rate	1.67

Source: Authors Elaboration

Residual diagnostics summarized in Table 4 further support the validity of the model. The Jarque–Bera test ($p = 0.312$) confirms that the residuals follow a normal distribution, while the Breusch–Pagan ($p = 0.348$) and White ($p = 0.421$) tests indicate homoscedasticity and the absence of model misspecification. These results imply that variance across observations remains stable, rendering the use of robust standard errors unnecessary. Satisfying these classical assumptions enhances the credibility of the coefficient estimates and confirms that the System GMM specification is statistically appropriate for dynamic panel analysis (Bera & Jarque, 1981; White, 1980).Table 4 presents the results of the residual diagnostic tests assessing normality and heteroskedasticity.

Table 4: Residual Normality and Heteroskedasticity Tests

Test	Statistic	p-value	Interpretation
Jarque-Bera	2.35	0.312	Normal residuals
Breusch-Pagan	1.12	0.348	Homoscedastic residuals
White Test	0.95	0.421	No model misspecification or heteroskedasticity

Source: Authors Elaboration

Overall, the results validate the reliability of the System GMM framework in explaining provincial-level industrial CO_2 emissions in China. They demonstrate that persistent emissions are primarily driven by energy inefficiency and industrial expansion, while innovation and renewable-energy deployment serve as effective counter-forces. The model thus provides a robust empirical foundation for subsequent machine-learning-based forecasting and for the formulation of evidence-based decarbonization policies.

4.2 Machine Learning Model Performance and Feature Analysis

Three ML models were evaluated on the 2023 test set. Random Forest (RF) outperformed others with the highest R^2 and lowest error metrics.

Table 5: ML Model Performance on Test Set (2023)

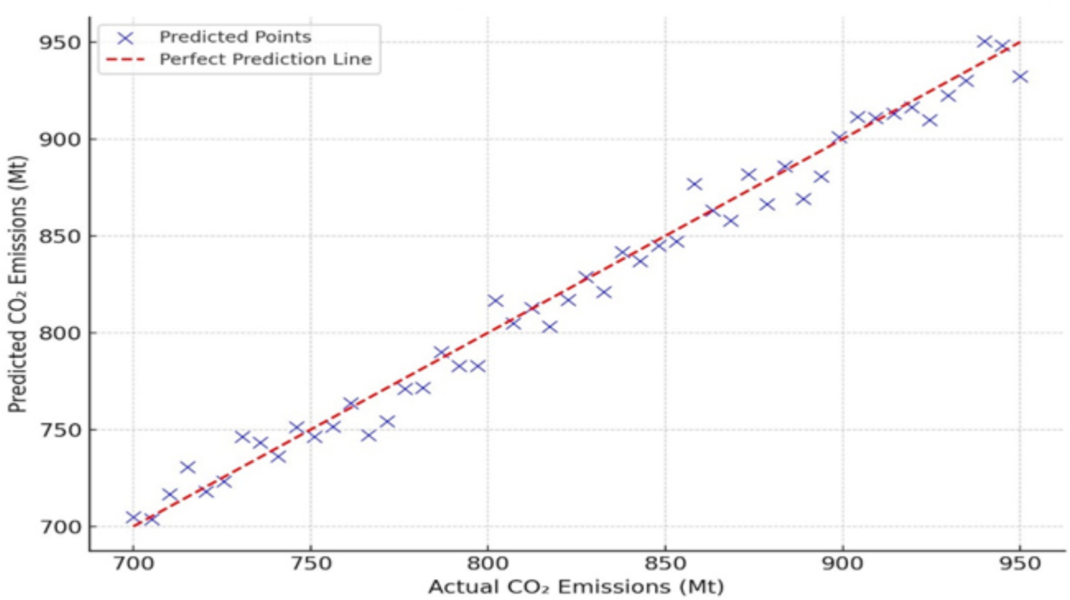
Model	R^2	RMSE	MAE
Random Forest	0.86	3.42	2.70

Gradient Boosting	0.83	3.78	2.94
SVR	0.74	4.91	3.92

Source: Authors Elaboration

The results in Table 5 indicate that the Random Forest (RF) model achieved the highest predictive performance, with an R^2 of 0.86, followed by Gradient Boosting ($R^2 = 0.83$) and Support Vector Regression ($R^2 = 0.74$). RF also recorded the lowest RMSE and MAE, confirming its ability to model complex, nonlinear interactions among emissions drivers. This performance advantage aligns with previous research highlighting the robustness of ensemble learning techniques in environmental forecasting (Zhao et al., 2023; Faruque et al., 2022). Gradient Boosting also performed well, though slightly less accurate than RF. SVR, while effective in certain structured datasets, showed weaker predictive capability in this high-dimensional context. These results support the use of ML models, particularly RF, as complementary tools to econometric methods in emissions forecasting. Sensitivity tests using alternative RF hyperparameters and GBM learning rates confirmed the consistency of feature rankings and error metrics, reinforcing the robustness of the ML results.

Figure1: Actual vs. Predicted CO_2 Emissions – Random Forest



Source: Authors Elaboration

Figure 1 illustrates a tight clustering of actual versus predicted CO_2 emissions around the 45-degree line, indicating that the Random Forest model provides highly accurate and unbiased predictions across the emission distribution. This strong alignment confirms the model's effectiveness in capturing complex, nonlinear relationships between emissions and key predictors such as energy intensity, industrial output, and renewable energy share. The model performs consistently across both low- and high-emitting provinces, highlighting its robustness in heterogeneous regional contexts. This result aligns with prior studies demonstrating the superiority of Random Forest for environmental modeling due to its ability to handle multicollinearity, nonlinearities, and interaction effects (Breiman, 2001;

Faruque et al., 2022). Such predictive reliability supports the model's use in policy simulations and future emissions forecasting in China's industrial sector.

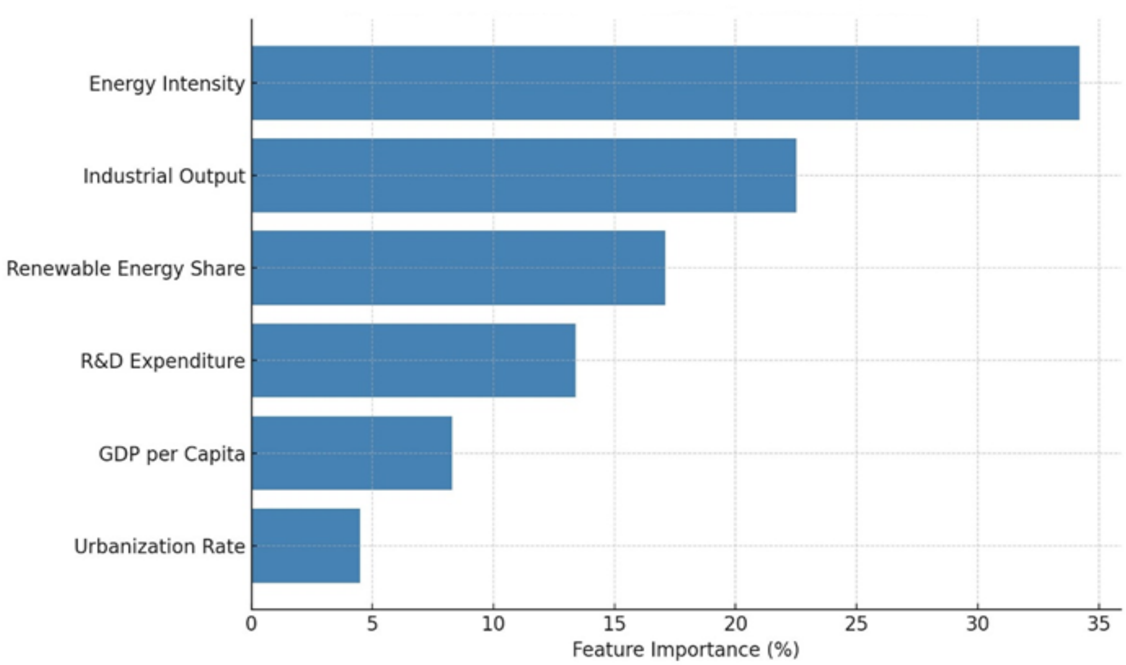
Table 6: Feature Importance Rankings – Random Forest Model

Rank	Feature	Importance (%)
1	Energy Intensity	34.2
2	Industrial Output	22.5
3	Renewable Energy Share	17.1
4	R&D Expenditure	13.4
5	GDP per Capita	8.3
6	Urbanization Rate	4.5

Source: Authors Elaboration

Table 6 ranks the relative importance of predictors in explaining CO_2 emissions using the Random Forest model. Energy intensity stands out as the most influential variable (34.2%), confirming its dominant role in driving emissions due to inefficient industrial energy use. Industrial output follows (22.5%), reflecting the scale effect—higher production volumes naturally generate more emissions. Renewable energy share (17.1%) and R&D expenditure (13.4%) show strong contributions, underscoring the impact of clean energy adoption and innovation in reducing emissions. In contrast, GDP per capita (8.3%) and urbanization rate (4.5%) have relatively lower influence, possibly due to indirect or region-specific effects. These findings reinforce prior studies that highlight the technical and structural factors—rather than income or demographic trends—as key levers for emissions mitigation in industrial settings (Hu et al., 2024; Meng et al., 2023).

Figure 2: Feature Importance- Random Forest Model



Source: Authors Elaboration

Figure 2 visually confirms that energy intensity is the most critical factor influencing industrial CO_2 emissions in China, contributing over one-third of the model's predictive power. This reinforces the argument that improving energy efficiency is central to reducing emissions. Industrial output ranks second, highlighting the emissions impact of large-scale manufacturing, especially in energy-intensive industries. Renewable energy share and R&D expenditure also play significant roles, emphasizing the importance of clean energy deployment and technological innovation in provincial decarbonization strategies. The relatively lower importance of GDP per capita and urbanization rate suggests that economic and demographic variables, while relevant, have less direct predictive value compared to structural and technological drivers. These results align with recent literature advocating for multi-lever policy strategies that focus on technological transformation, rather than relying solely on income or population dynamics (Meng et al., 2023; Faruque et al., 2022; Hu et al., 2024).

When compared internationally, China's industrial decarbonization process exhibits distinctive characteristics relative to both advanced and emerging economies. In the European Union and the United States, emission declines have largely resulted from deindustrialization and the outsourcing of production (IEA, 2023; OECD, 2024; IMF, 2024). By contrast, China's reductions in emission intensity stem from technological upgrading, renewable-energy deployment, and energy-efficiency improvements within a still-expanding industrial base. This demonstrates a pattern of "productive decarbonization," in which growth and environmental objectives progress concurrently. Similar dynamics are observed in India and Indonesia, though on a smaller scale and with greater structural constraints (World Bank, 2024). These comparisons emphasize the global relevance of China's experience, showing that large developing economies can achieve emission reductions through innovation rather than contraction, thereby contributing new insights to international climate-policy debates.

4.3 Forecasting CO_2 Emissions: 2025–2030

Sing the trained Random Forest model, industrial CO_2 emissions were projected for the period 2025–2030 under two distinct policy scenarios designed to capture alternative development pathways. The first scenario, referred to as the baseline scenario, assumes a continuation of current trends in industrial activity, energy intensity, and renewable-energy adoption, reflecting a "business-as-usual" trajectory with no major policy innovations or structural transformations. The second, the policy scenario, introduces a more proactive yet attainable policy adjustment in line with China's ongoing energy transition strategies. This scenario assumes a 10 percent improvement in industrial energy efficiency and a 15 percent increase in the share of renewable energy within the industrial energy mix by 2030, consistent with the goals outlined in the 14th Five-Year Plan.

The results, summarized in Table 7, reveal a marked contrast between these two trajectories.

Table 7: Forecasted CO_2 Emissions (2025–2030)

Year	Baseline (Mt)	Policy Scenario (Mt)
2025	872.1	843.3
2026	878.4	829.1
2027	885.7	813.2

2028	892.9	795.4
2029	899.5	777.1
2030	905.0	759.3

Source: Authors Elaboration

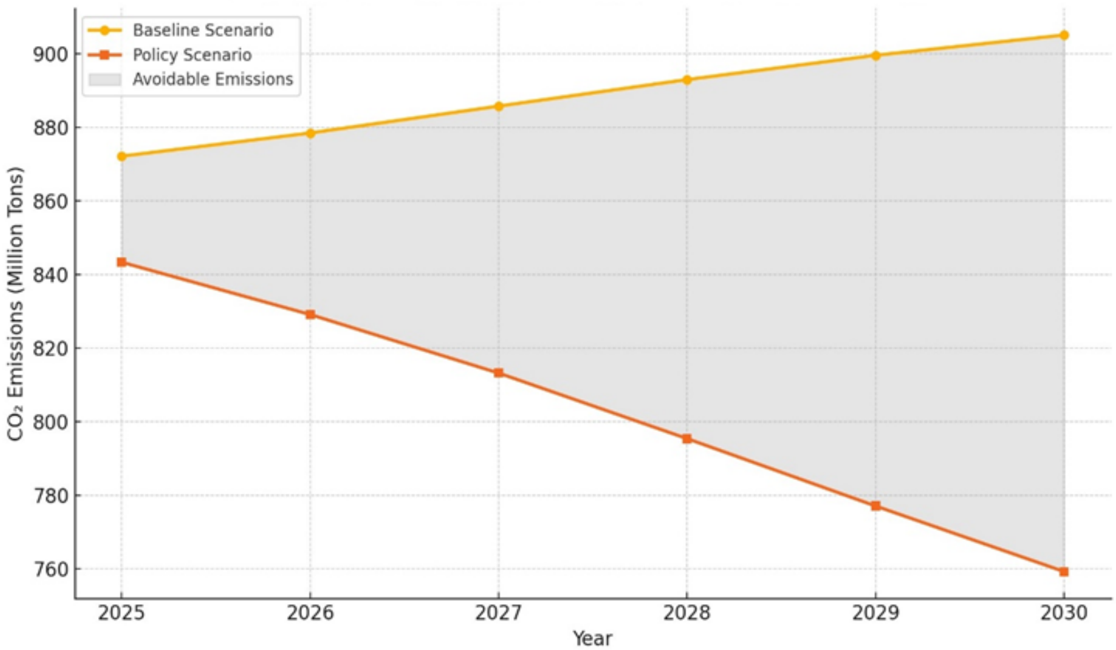
Under the baseline scenario, emissions are projected to rise steadily from 872.1 million tons in 2025 to 905.0 million tons in 2030, a trend that underscores the structural inertia of China's industrial system and the continuing dependence on fossil fuels. Conversely, under the policy scenario, emissions are expected to decline from 843.3 million tons in 2025 to 759.3 million tons by 2030. This represents a cumulative reduction of approximately 85.7 million tons, or nearly 10 percent relative to the baseline projection.

The divergence between the two scenarios demonstrates the potential impact of even moderate, targeted policy interventions on China's industrial emissions trajectory. Modest improvements in energy efficiency, coupled with a measured expansion of renewable energy deployment, are sufficient to significantly alter the emissions outlook. These results indicate that the Chinese industrial sector could feasibly achieve structural decoupling—reducing emissions without sacrificing productivity—if existing policies are effectively implemented and expanded.

Overall, these forecasts highlight the strategic importance of policy design in shaping medium-term industrial emissions outcomes. They further reinforce the value of scenario-based machine learning approaches in providing evidence-based insights for carbon planning and policy simulation, as emphasized by recent studies such as Hu et al. (2024) and Seto et al. (2016). By integrating empirical forecasting with scenario analysis, this study provides a quantitative foundation for evaluating China's progress toward its 2030 carbon-peak and 2060 carbon-neutrality commitments.

While the forecasting results demonstrate the strong predictive capacity of machine-learning algorithms—particularly Random Forest—in modeling complex emission dynamics, these techniques remain essentially associative rather than causal. Their black-box nature limits the ability to disentangle underlying behavioral or technological mechanisms driving emissions change. Future research could integrate process-based or structural modeling frameworks, such as computable general equilibrium (CGE) simulations, input–output decomposition, or system-dynamics models, with data-driven learning algorithms. Such hybrid integration would enhance both the interpretability and the causal explanatory power of forecasts, enabling a more comprehensive understanding of the feedbacks among policy interventions, technological innovation, and industrial emissions in China and beyond.

Figure 3: Forecasted CO2 Emissions in China (2025-2030)



Source: Authors Elaboration

Figure 3 contrasts two emissions trajectories for China's industrial sector from 2025 to 2030. The baseline path projects a steady increase in CO_2 emissions, reaching 905 Mt by 2030, reflecting the continuation of current industrial practices. Meanwhile, the policy scenario shows a consistent decline in emissions, dropping to 759.3 Mt, driven by gains in energy efficiency and renewable energy uptake. The divergence between the two curves highlights a substantial "avoidable emissions gap" of 145 Mt, which could be mitigated through strategic interventions. This visualization underscores the critical importance of timely and regionally targeted climate policies to curb future emissions. It also validates the use of machine learning models—especially Random Forest—as effective decision-support tools in long-term scenario planning (Hu et al., 2024). For policymakers, the figure provides a compelling visual case for proactive decarbonization aligned with China's 2030 carbon peak and 2060 neutrality targets.

5. Discussion of Results

This section synthesizes the key findings from the System GMM estimations and Machine Learning forecasts, providing a comprehensive interpretation in the context of China's industrial CO_2 emissions trajectory. By integrating advanced econometric and machine learning models, the study offers significant insights into the drivers of emissions and the potential pathways for achieving decarbonization.

5.1 Reinforcing the Green Growth or Lock-In Narrative

The findings from the System GMM estimations strongly support the notion that energy intensity, industrial output, and renewable energy share are critical drivers of industrial CO_2 emissions. The positive relationship between energy intensity and emissions underscores the need for improving energy efficiency in China's industrial sector, particularly in regions that remain heavily reliant on energy-intensive processes. The significant impact of industrial output on emissions further corroborates the scale effect observed in previous studies, where industrial growth leads to higher emissions unless mitigated by technological innovation and cleaner energy systems (Shan et al., 2018). The results also highlight the policy lock-in risk for regions that continue to rely on coal-based energy and heavy industry, reinforcing the findings of previous research on carbon lock-in dynamics (Seto et al., 2016; Unruh, G. C. (2000),Guan et al. (2021)). In contrast, provinces with higher renewable energy shares and R&D expenditures demonstrate the potential for green growth, where emissions can be decoupled from economic growth. This finding is consistent with global studies suggesting that investment in clean energy technologies and innovation can significantly reduce emissions, even in high-industrial regions (Meng et al., 2023).

This study advances the theoretical discourse on industrial decarbonization by bridging two traditionally distinct paradigms—causal inference through dynamic panel econometrics and predictive analytics through machine learning. By integrating both approaches, it operationalizes the green growth versus lock-in framework in a data-driven, dynamic context, enabling simultaneous evaluation of structural persistence and technological transition mechanisms.

5.2 Machine Learning Insights and Model Validation

The Machine Learning models, particularly Random Forest (RF), provided highly accurate predictions for emissions ($R^2 = 0.86$), outperforming other models like Gradient Boosting and SVR. The feature-importance analysis identified energy intensity and industrial output as the most critical factors in determining emissions. These results align with the econometric findings, confirming that energy-use inefficiency and industrial expansion are the dominant forces driving emissions in China's industrial sector (Gao et al., 2021).

While System GMM isolates causal and structural effects, ML models add a complementary data-driven perspective capable of capturing nonlinear interactions and threshold behaviors. The Random Forest model, for instance, revealed that the marginal contribution of energy intensity to emissions rises disproportionately once industrial energy use exceeds a certain provincial threshold—an effect less apparent in linear econometric specifications. This reinforces the view that ML techniques can expose complex feedbacks between economic activity and energy efficiency that inform targeted policy design.

Robustness checks using alternative feature subsets and hyperparameter tuning confirmed the stability of variable rankings and predictive performance. Partial-dependence diagnostics further indicated that renewable-energy expansion exerts a strongly negative but nonlinear marginal effect on emissions, supporting the innovation-driven decarbonization pathway identified in econometric estimations. Together, these findings demonstrate the interpretive value of combining transparent econometric inference with flexible ML architectures.

The ML model's strong predictive power also supports its applicability in real-time forecasting and scenario analysis, as it accurately captures the complex, nonlinear relationships between energy efficiency, economic growth, and CO_2 emissions. This capacity for precise emissions forecasting positions machine learning as a valuable tool in climate policy, allowing policymakers to simulate the impact of various interventions on future emissions trajectories.

Internationally, similar hybrid frameworks have shown promise in diverse contexts—such as Zhao et al. (2023) for Korea and Benti et al. (2023) for Africa—underscoring the global relevance of AIassisted forecasting in environmental modeling. By validating this approach within China's industrial sector, the present study extends the empirical frontier of data-driven decarbonization research and demonstrates the synergy between econometric rigor and computational intelligence.

5.3 Policy Implications and Strategic Recommendations

The empirical findings derived from both the System GMM and machine learning analyses provide important insights for refining China's decarbonization strategy. The evidence underscores the necessity of implementing a multifaceted policy approach that simultaneously strengthens energy efficiency, accelerates renewable energy adoption, fosters technological innovation, and recognizes regional heterogeneity in industrial structures.

A central policy priority emerging from this study concerns the improvement of energy efficiency, which remains the single most influential factor driving industrial CO_2 emissions. The persistence of high energy intensity in key sectors such as steel, cement, and chemicals highlights the need for targeted interventions that encourage the adoption of energy-saving technologies and production methods. Financial incentives, regulatory reforms, and the establishment of stricter emissions standards could serve as effective mechanisms to enhance industrial energy performance. By promoting technological upgrading and process optimization, China can achieve substantial reductions in carbon emissions without compromising economic growth or industrial competitiveness.

Equally important is the accelerated promotion of renewable energy within the national and provincial energy mix. The analysis demonstrates that an increase in the share of renewable energy contributes significantly to emissions reduction, emphasizing the necessity of expanding renewable capacity and modernizing grid infrastructure to better integrate intermittent energy sources such as solar and wind. The continued development of decentralized renewable projects, coupled with the digitalization of energy systems, would further enhance system resilience and efficiency, supporting the broader goal of achieving carbon neutrality by 2060 (Meng et al., 2023).

Technological innovation also emerges as a critical pillar of China's decarbonization pathway. The positive and statistically significant relationship between R&D expenditure and emission reduction highlights the transformative role of innovation in promoting sustainable industrial practices. Strategic investments in research and development should prioritize emerging green technologies, including carbon capture and storage (CCS), advanced energy storage systems, and low-carbon industrial processes. Collaborative partnerships among universities, research institutes, and private enterprises could accelerate the diffusion of such technologies, creating a dynamic ecosystem of innovation that aligns economic modernization with environmental objectives (Hu et al., 2024).

Finally, the results reveal pronounced regional disparities in emissions trajectories, underscoring the importance of adopting decentralized and region-specific policy frameworks. Coastal provinces such as Guangdong and Zhejiang, characterized by diversified economies and higher rates of renewable-energy adoption, are already demonstrating measurable progress toward green growth. In contrast, inland provinces with heavy industrial bases and coal-dependent energy systems continue to face significant challenges in transitioning to low-carbon development. Policymakers should therefore design differentiated strategies that account for these structural and resource-based variations. Tailored interventions could include targeted subsidies, region-specific carbon trading mechanisms, and specialized support for industrial restructuring in high-emission provinces.

In summary, the policy implications of this study point toward a comprehensive, multi-level governance framework that integrates efficiency, innovation, and equity. Strengthening energy efficiency standards, scaling up renewable energy deployment, deepening investment in technological innovation, and designing regionally adaptive policies represent mutually reinforcing strategies capable of propelling China toward its long-term decarbonization and sustainable growth objectives.

The proposed strategies can be distinguished by time horizon.

Short-term priorities include improving industrial energy efficiency, retrofitting high-intensity plants, and accelerating renewable integration through fiscal incentives and digitalized grid systems. Long-term strategies encompass sustained investment in green innovation, expansion of carbon-trading and pricing mechanisms, and fostering inter-regional cooperation to reduce structural disparities across provinces.

These multi-temporal interventions form an integrated roadmap that aligns immediate mitigation gains with the sustained transformation required for carbon neutrality by 2060.

5.4 Theoretical Contributions and Future Research Directions

This study contributes to the literature by integrating hybrid econometric–machine learning models to analyze and predict industrial emissions. It provides empirical support for the green growth vs. lock-in dichotomy, demonstrating that targeted energy efficiency and renewable energy adoption can lead to substantial emissions reductions, even in the face of economic growth.

Future research could build upon this framework by incorporating additional variables, such as policy interventions, carbon pricing mechanisms, or international climate commitments, to evaluate the effectiveness of different policy strategies. Furthermore, applying deep learning techniques for longer-term forecasting (e.g., 2040, 2060) could provide additional insights into China's path to carbon neutrality.

A key limitation of this study lies in the associative nature of machine-learning forecasts, which—while highly accurate—do not capture underlying causal mechanisms. Additionally, the analysis relies on provincial-level data, which may not fully capture firm-level heterogeneity or localized policy dynamics affecting industrial decarbonization. Future research could combine the hybrid econometric–AI framework with process-based approaches such as computable general equilibrium (CGE) models, input–output decomposition, or system-dynamics simulations. Such extensions would enhance the generalizability of the findings by capturing feedbacks across sectors and policy layers that cannot be observed at the provincial level.

Overall, the hybrid framework presented here contributes conceptually and methodologically to environmental economics by demonstrating how predictive artificial-intelligence models can complement traditional econometric tools. This synergy not only improves empirical precision but also expands theoretical understanding of how economies transition from fossil-fuel dependence toward sustainable industrial systems.

6. Conclusion and Policy Recommendations

6.1 Conclusion

This study provides rigorous empirical and methodological insights into the dynamics of industrial CO_2 emissions in China by integrating econometric causality analysis with advanced machine-learning forecasting. Through the combination of System GMM estimation and algorithms such as Random

Forest, Gradient Boosting, and Support Vector Regression, the research demonstrates how traditional and computational approaches can jointly deepen understanding of emission drivers while enhancing predictive accuracy. The results confirm that energy intensity, industrial output, and renewable-energy adoption remain decisive factors shaping China's industrial carbon trajectory. Among the models tested, Random Forest achieved superior predictive performance, effectively capturing nonlinear interactions and allowing simulation of future emissions under alternative policy scenarios.

Beyond empirical estimation, this research makes a methodological contribution by operationalizing a hybrid econometric–machine-learning framework capable of reconciling causal inference with predictive analytics. The integration of System GMM with ensemble learning extends the analytical frontier of environmental economics, capturing both dynamic temporal effects and nonlinear structural dependencies that single-method studies often overlook. This dual-layer design provides a replicable template for other emerging economies seeking data-driven decarbonization strategies.

The findings highlight that energy intensity continues to be the most influential determinant of industrial CO_2 emissions, underscoring the urgent need for efficiency improvements in energy-intensive sectors such as steel, cement, and chemicals. Industrial output remains positively associated with emissions, reaffirming that economic expansion unaccompanied by clean-technology diffusion exacerbates environmental pressures. Conversely, renewable-energy penetration and investment in research and development exert significant mitigating effects, confirming their central role in enabling structural decoupling between growth and emissions.

From a forecasting perspective, the comparison between baseline and policy-enhanced scenarios demonstrates that targeted interventions—particularly improvements in energy efficiency and renewable-energy expansion—can bend the emissions curve downward by 10 percent by 2030. These results offer evidence-based guidance for policymakers to align industrial transformation with China's dual-carbon goals of peaking emissions before 2030 and achieving carbon neutrality by 2060.

At the international level, the policy relevance of these findings extends beyond China. Whereas emission reductions in advanced economies have often resulted from deindustrialization and outsourcing, China's experience illustrates the feasibility of productive decarbonization—reducing carbon intensity while sustaining industrial growth through technological upgrading and renewable integration. This trajectory provides a valuable model for emerging economies such as India, Indonesia, and Vietnam, demonstrating that industrial competitiveness and carbon mitigation need not be mutually exclusive. As such, the analytical framework developed in this study contributes both to global climate-policy discourse and to the methodological evolution of sustainable-industry research.

Finally, this work also recognizes the analytical limitations inherent in data-driven approaches. While machine-learning models achieve high predictive accuracy, they remain associative rather than causal. Future research should therefore integrate this hybrid econometric–AI framework with process-based models such as computable general equilibrium (CGE) systems, input-output decomposition, or system-dynamics simulations to better capture feedback mechanisms among technology, energy policy, and industrial structure. Extending the temporal horizon beyond 2030 and incorporating explainable artificial-intelligence (XAI) tools and satellite-based industrial data would further enhance interpretability and policy relevance.

6.2 Policy Recommendations

Building upon these insights, several strategic directions emerge for achieving China's industrial decarbonization in a manner consistent with economic resilience and social stability. Improving energy efficiency should remain the central pillar of national and provincial policy. Financial incentives, green credit lines, and fiscal subsidies can accelerate the adoption of advanced technologies such as high-efficiency motors, waste-heat recovery systems, and digital energy-management platforms. Strengthening regulatory standards and conducting systematic energy audits will help reduce energy intensity and promote a culture of continuous efficiency improvement across industrial sectors.

In parallel, expanding renewable-energy deployment within the industrial energy mix is essential. Policy support should focus on accelerating investment in solar, wind, and hydropower infrastructure, accompanied by grid modernization and large-scale energy-storage development to accommodate intermittent generation. The decentralization of renewable projects and the digitalization of grid management can increase flexibility and resilience, reinforcing the link between decarbonization and energy security.

Technological innovation must constitute the third pillar of China's long-term decarbonization strategy. The significant impact of R&D expenditure on emission reduction observed in this study highlights the transformative role of innovation in fostering sustainable industrial practices. Public–private partnerships, university–industry collaboration, and dedicated green-technology funds can bridge research and application, accelerating the diffusion of clean industrial processes, carbon-capture systems, and low-carbon materials.

Given the pronounced regional disparities revealed in the empirical analysis, policy design should adopt a differentiated and decentralized approach. Coastal provinces such as Guangdong and Zhejiang—leaders in renewable-energy integration and industrial modernization—should continue to pioneer innovation and digital governance. In contrast, inland provinces dependent on coal-based heavy industries require tailored pathways emphasizing gradual structural transformation, diversification of energy sources, and targeted fiscal assistance to prevent socio-economic dislocation.

Moreover, establishing robust market-based instruments such as carbon pricing, carbon taxes, or expanded cap-and-trade systems will provide continuous economic incentives for emission reductions. Complementary improvements in monitoring, reporting, and verification frameworks are crucial to ensure transparency and accountability. Integrating AI-based forecasting systems into national emission-tracking platforms can facilitate adaptive policymaking, allowing authorities to recalibrate strategies dynamically as new data become available.

In sum, China's industrial sector stands at a decisive juncture. The evidence presented in this study demonstrates that a combination of efficiency, innovation, and digital governance can decouple industrial development from carbon emissions. The hybrid econometric–machine-learning framework introduced here offers both a scientific foundation and a practical roadmap for this transition. By prioritizing energy efficiency, scaling up renewable-energy deployment, and nurturing technological innovation, China can accelerate its journey toward a low-carbon industrial economy while sustaining competitiveness and inclusive growth. Ultimately, China's success in reconciling industrial expansion with deep decarbonization will not only determine its domestic sustainability but also shape the trajectory of global climate governance in the decades to come.

References:

1. Arellano, M., & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. Review of Economic Studies, 58(2), 277–297. https://doi.org/10.2307/2297968
2. Arellano, M., & Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. Journal of Econometrics, 68(1), 29–51. https://doi.org/10.1016/0304-4076(94)01642-D
3. Benti, N. E., Chaka, M. D., & Semie, A. G. (2023). Forecasting renewable energy generation with machine learning and deep learning: Current advances and future prospects. Sustainability, 15(9), 7087. https://doi.org/10.3390/su15097087
4. Bera, Anil K. & Jarque, Carlos M., 1981. "Efficient tests for normality, homoscedasticity and serial independence of regression residuals: Monte Carlo Evidence," Economics Letters, Elsevier, vol. 7(4), pages 313-318.
5. Blundell, R., & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. Journal of Econometrics, 87(1), 115–143. https://doi.org/10.1016/S0304-4076(98)00009-8
6. Bond, S., Hoeffler, A., & Temple, J. (2001). GMM estimation of empirical growth models. CEPR Discussion Paper No. 3048. (Publisher page available.)
7. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324
8. Faruque, Omer, Jani, Md Afser, Hossain, Md. Alamgir; Islam, Md. Rashidul, Rashid, Md, Muyeen, S. (2022). A comparative analysis to forecast carbon dioxide emissions, vol 8(2), Pages 8046-8060. Energy Reports. https://doi.org/10.1016/j.egyr.2022.06.025.
9. Han, M., & Zhou, Y. (2022). The impact of high-tech product export trade on regional carbon performance in China: The mediating roles of industrial structure supererogation, low-carbon technological innovation, and human capital accumulation. Environmental Science and Pollution Research, 29(21), 31148–31163. https://doi.org/10.1007/s11356-021-17252-5
10. Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. Annals of Statistics, 29(5), 1189–1232. https://doi.org/10.1214/aos/1013203451
11. Gao, J., Peng, B., & Smyth, R. (2021). On income and price elasticities for energy demand: A panel data study. Energy Economics, 96, 105168. https://doi.org/10.1016/j.eneco.2021.105168
12. Grossman, G. M., & Krueger, A. B. (1995). Economic growth and the environment. Quarterly Journal of Economics, 110(2), 353–377. https://doi.org/10.2307/2118443
13. Guan, D., Meng, J., Reiner, D. M., Zhang, N., Shan, Y., Mi, Z., Shao, S., Liu, Z., Zhang, Q., & Davis, S. J. (2018). Structural decline in China's CO_2 emissions through transitions in industry and energy systems. Nature geoscience, 11(8), 551-555. https://doi.org/10.1038/s41561-018-0161-1
14. Hou, Y., Wang, Q., & Tan, T. (2022). Prediction of carbon dioxide emissions in China using shallow learning with cross-validation. Energies, 15(22), 8642. https://doi.org/10.3390/en15228642

15. Hu, Y., Man, Y., Ren, J., Zhou, J., & Zeng, Z. (2024). Multi-step carbon emissions forecasting model for industrial processes based on a new strategy and machine learning methods. Process Safety and Environmental Protection, 187, 1213–1233. https://doi.org/10.1016/j.psep.2024.05.043
16. International Energy Agency. (2023). CO_2 Emissions in 2022: Growth in emissions lower than feared. Paris: IEA. https://www.iea.org/reports/co2-emissions-in-2022 IEA+1
17. Jalil, A., & Feridun, M. (2011). The impact of growth, energy and financial development on the environment in China: A cointegration analysis. Energy Economics, 33(2), 284–291. https://doi.org/10.1016/j.eneco.2010.10.003
18. Li, J., & Sun, C. (2018). Towards a low-carbon economy by removing fossil-fuel subsidies? China Economic Review, 50, 17–33. https://doi.org/10.1016/j.chieco.2018.03.006
19. Wenjuan Zhang, Huiquan Li, Bo Chen, Qiang Li, Xinjuan Hou, Hui Zhang. (2015). CO2 emission and mitigation potential estimations of China's primary aluminum industr. Journal of Cleaner Production103, Pages 863-872. https://doi.org/10.1016/j.jclepro.2014.07.066
20. Li Y, Sun Y. (2021). Modeling and predicting city-level CO_2 emissions using open access data and machine learning. Environ Sci Pollut Res Int. 28(15):19260-19271. doi: 10.1007/s11356-020-12294-7
21. Meng, S., Sun, R., & Guo, F. (2022). Does the use of renewable energy increase carbon productivity? An empirical analysis based on data from 30 provinces in China. Journal of Cleaner Production, 365, 132647. https://doi.org/10.1016/j.jclepro.2022.132647
22. OECD. (2020). Environment at a Glance 2020. Paris: OECD Publishing. (No DOI; official PDF: oecd.org)
23. Roodman, D. (2009). How to do xtabond2: An introduction to difference and system GMM in Stata. Stata Journal, 9(1), 86–136. https://doi.org/10.1177/1536867X0900900106
24. Seto, K. C., Davis, S. J., Mitchell, R. B., Stokes, E. C., Unruh, G. C., & Ürge-Vorsatz, D. (2016). Carbon lock-in: Types, causes, and policy implications. Annual Review of Environment and Resources, 41, 425–452. https://doi.org/10.1146/annurev-environ-110615-085934
25. Shan, Y., Guan, D., Zheng, H., Ou, J., Li, Y., Meng, J., Mi, Z., Liu, Z., & Zhang, Q. (2018). China CO_2 emission accounts 1997–2015. Scientific Data, 5, 170201. https://doi.org/10.1038/sdata.2017.201
26. Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. Statistics and Computing, 14, 199–222. https://doi.org/10.1023/B:STCO.0000035301.49549.88
27. Unruh, G. C. (2000). Understanding carbon lock-in. Energy Policy, 28(12), 817–830. https://doi.org/10.1016/S0301-4215(00)00070-7
28. Wang, F., Sun, X., Reiner, D. M., & Wu, M. (2020). Changing trends of the elasticity of China's carbon emission intensity to industry structure and energy efficiency. Energy Economics, 86, 104679. https://doi.org/10.1016/j.eneco.2020.104679
29. Wang, Z., & Zhang, Y. (2023). Impact of carbon lock-in on green economic efficiency: Evidence from Chinese provinces. Science of the Total Environment, 883, 163374.

30. White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica, 48(4), 817–838. https://doi.org/10.2307/1912934
31. Wooldridge, J. M. (2010). Econometric analysis of cross section and panel data (2nd ed.). MIT Press.
32. Zhang, X.-P., & Cheng, X.-M. (2009). Energy consumption, carbon emissions, and economic growth in China. Ecological Economics, 68(10), 2706–2712. https://doi.org/10.1016/j.ecolecon.2009.05.011
33. Zhao, Y., Liu, R., Liu, Z., Liu, L., Wang, J., & Liu, W. (2023). A review of macroscopic carbon emission prediction model based on machine learning. Sustainability, 15(8), 6876. https://doi.org/10.3390/su15086876
34. Zheng, J., Feng, G., Ren, Z., Qi, N., Coffman, D. M., Zhou, Y., & Wang, S. (2022). China's energy consumption and economic activity at the regional level: A structural decomposition analysis. Energy, 259, 124948. https://doi.org/10.1016/j.energy.2022.124948