ASM Sc. J., 20(2), 2025

https://doi.org/10.32802/asmscj.2025.0171

Towards Carbon Neutrality: A Novel STIRPAT-XGBoost-SHAP Framework for Provincial Energy Transition Prediction in China

Q. Lu[1], K.W. Khaw[1*], X.Y. Chew[2], W.C. Yeong[3] and W.C. Ng[1]

[1]*School of Management, Universiti Sains Malaysia, 11800 Penang, Malaysia*

[2]*School of Computer Sciences, Universiti Sains Malaysia, 11800 Penang, Malaysia*

[3]*Faculty of Computing & Informatics, Multimedia University, Persiaran Multimedia, 63100 Cyberjaya, Selangor, Malaysia*

China's commitment to carbon neutrality by 2060 demands accurate and interpretable forecasting tools at the provincial level. This study develops and validates a hybrid framework integrating the STIRPAT model, XGBoost machine learning, and SHAP interpretability analysis to forecast coal and total energy consumption across six representative Chinese provinces (2005-2021). The STIRPAT model reveals industrial structure as the dominant driver of coal dependence, while SHAP confirms structural consistency and highlights nonlinear effects of urbanisation and income. The XGBoost model achieves competitive forecasting performance (Mean Absolute Percentage Error (MAPE): 5.26% for coal and 3.02% for total energy) and effectively captures regional disparities in coal transition trajectories. These results support differentiated and structurally grounded policy interventions, offering practical guidance for subnational energy planning under China's dual-carbon strategy. The framework offers broad applicability to other structurally diverse and data-constrained contexts.

Keywords: STIRPAT; XGBoost; SHAP; energy transition; coal consumption; subnational forecasting; green finance; China

I. INTRODUCTION

China aims to achieve peak carbon emissions by 2030 and carbon neutrality by 2060. This commitment has increased the need for forecasting frameworks that are both technically accurate and structurally interpretable at the subnational level. While the central government has outlined ambitious decarbonization targets, provincial-level trajectories remain highly heterogeneous due to differing industrial legacies, energy structures, and institutional capacities. For instance, provinces such as Shaanxi and Inner Mongolia continue to rely on coal for over 70% of their primary energy consumption, while eastern coastal regions like Guangdong and Zhejiang have diversified their energy mixes with coal shares below 35% (Liu, Zhang, & Wang, 2021). These disparities challenge unified policy design. They also necessitate the development of tailored forecasting models capable of capturing cross-regional structural variations.

The challenges associated with subnational energy planning are not exclusive to China. In the United States, for example, California has implemented far-reaching clean energy initiatives within a decentralised policy system, while resource-dependent states such as Wyoming have progressed more slowly despite national climate targets (Carley & Konisky, 2020). A similar divergence is observed in Germany, where individual federal states (Länder) follow distinct decarbonisation pathways under a shared Energiewende framework, illustrating how local economic structures and resource endowments shape transition trajectories (Ohlhorst, 2015). These international experiences highlight the importance of developing

*Corresponding author's e-mail: khaiwah@usm.my

forecasting tools that are capable of addressing multi-level governance dynamics and regional heterogeneity in energy systems.

However, conventional modelling approaches often struggle to balance predictive accuracy with interpretability. Econometric approaches, including the STIRPAT model (Stochastic Impacts by Regression on Population, Affluence, and Technology), provide a solid theoretical foundation for understanding environmental drivers. However, they often fall short in representing complex nonlinear relationships and can be sensitive to data limitations (York, Rosa & Dietz, 2003). Meanwhile, machine learning algorithms such as XGBoost (Extreme Gradient Boosting) are highly effective at detecting intricate, high-dimensional patterns. A recurring criticism, however, concerns their opaque nature and limited capacity for providing clear explanations. Such drawbacks reduce their practical usefulness in policy-making contexts, where both numerical robustness and interpretive clarity are essential.

In light of these methodological trade-offs, a growing body of work has sought to merge machine learning techniques with theory-guided modelling frameworks. Such integrations strive to improve not only predictive performance but also the interpretability of model structures. For example, Wu *et al.* (2023) incorporated XGBoost with spatial indicators of urban form to study energy intensity variations among Chinese cities, achieving gains in both accuracy and explanatory insight. Other scholars, such as Lundberg and Lee (2017), have adopted SHAP (Shapley Additive Explanations) to interpret gradient boosting models, using game-theoretic principles to attribute predictions to individual input features. Despite these advances, provincial-scale energy transition forecasting still lacks a unified framework that incorporates these methodological innovations within a consistent theoretical structure.

In this study, we develop and validate a hybrid modelling approach that integrates the elasticity-based interpretability of STIRPAT, the nonlinear predictive power of XGBoost, and the transparent attribution mechanisms of SHAP. Moving beyond prior work that uses machine learning merely as a black-box extension of conventional models, our research emphasises theoretical consistency, model interpretability, and relevance to policy applications. We test the proposed framework using data from six Chinese provinces over the period 2005-2021, assessing its utility along three key dimensions: (1) predictive performance under realistic constraints; (2) coherence of structural attributions across different model components; and (3) applicability to policy-focused energy analysis at the subnational level.

Beyond its empirical focus, the hybrid framework offers tangible value for engineering applications in regional energy infrastructure planning. Accurate forecasts of coal share trajectories can support power grid design, fossil fuel asset retirement scheduling, and renewable capacity deployment, particularly in provinces undergoing industrial restructuring. Moreover, the SHAP-based interpretability enhances the model's suitability for engineering decision support systems by enabling transparent diagnostic insights into structural drivers. These features make the framework readily integrable into energy system modelling workflows, such as those supporting distributed grid design or renewable planning tools.

II. MATERIALS AND METHOD

A. Study Area and Data Sources

This analysis examines six Chinese provinces selected to represent diverse regional characteristics and ensure data completeness: Guangdong and Jiangsu (eastern developed regions), Henan and Hubei (central transitional areas), and Sichuan and Shaanxi (western resource-abundant provinces). These provinces were chosen based on three criteria: (1) availability of complete annual energy and socioeconomic data from 2005 to 2021; (2) structural diversity in terms of industrialisation, urbanisation, and economic development; and (3) policy relevance under China's dual carbon strategy. The study period yields 102 province-year observations, capturing both pre-crisis stability and recent carbon policy intensification.

All data were sourced from official Chinese statistical publications, specifically the China Statistical Yearbook and China Energy Statistical Yearbook (National Bureau of Statistics of China & National Energy Administration, 2006-2022). Data cleaning procedures included cross-checking across years, interpolation for occasional missing entries, and standardisation of energy units using official conversion coefficients. All variables were reviewed for internal consistency and temporal comparability to ensure panel reliability.

The core variables used in the modelling framework are defined as follows:

(1) CoalConsumption: Annual coal consumption (10,000 tonnes)

(2) TotalEnergyConsumption:Total primary energy consumption (10,000 tonnes standard coal equivalent)

(3) CoalShare:Coal consumption as percentage of total energy consumption (calculated as CoalConsumption/TotalEnergyConsumption×100)

(4) UrbanRate: Urban population as percentage of total population

(5) GDPperCap: Per capita gross domestic product (yuan, constant 2005 prices)

(6) IndusShare: Secondary industry value-added as percentage of total GDP

All continuous variables were log-transformed to reduce skewness and allow elasticity interpretation in the econometric component. These structural variables were selected based on their theoretical roles in driving energy consumption: urbanisation reflects population density and infrastructure demand; income captures affluence-driven energy intensity changes; and industrial structure captures sectoral energy dependency, particularly coal-intensive manufacturing (Zhang *et al.*, 2021). Variable descriptions and sources are listed in Appendix A.

B. STIRPAT Model Specification

The theoretical foundation of the hybrid framework employs the STIRPAT model, widely used for analysing environmental impacts under structural change. We implement a fixed-effects panel specification to control for unobserved provincial heterogeneity, as specified in Eq. (1):

$$\ln(CoalShare_{it}) = \alpha_i + \beta_1 \ln(UrbanRate_{it}) + \beta_2 \ln(GDPperCap_{it}) + \beta_3 \ln(IndusShare_{it}) + \beta_4 Year_t + \varepsilon_{it} \quad (1)$$

where represents the coal share in province i during, captures province-specific fixed effects, and is a linear time trend capturing national policy effects. The model is estimated using fixed-effects regression with cluster-robust standard errors to address heteroskedasticity and within-group autocorrelation.

Multicollinearity diagnostics confirm that all variance inflation factors (VIF) remain below 5.0, ensuring parameter stability (Appendix C). The log-linear specification enables direct interpretation of coefficients as elasticities, facilitating comparison with machine learning attributions.

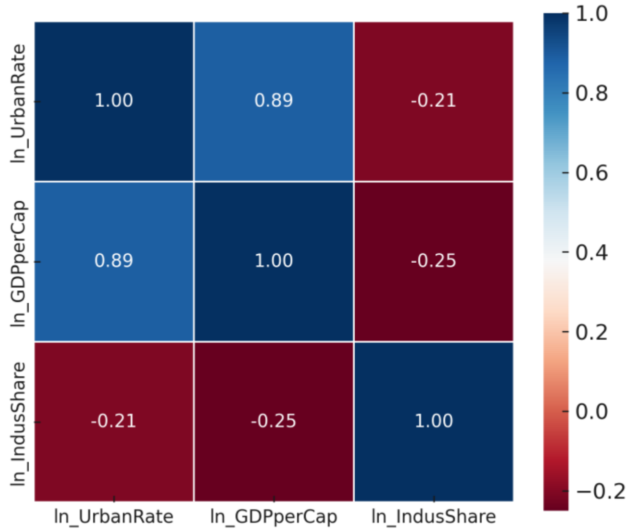

Figure 1. Pearson correlation heatmap of the log-transformed predictor variables.

To further assess potential collinearity among the explanatory variables, Figure 1 presents a Pearson correlation heatmap for the log-transformed predictors. The observed maximum pairwise correlation (0.89) occurs between ln_UrbanRate and ln_GDPperCap, which is theoretically expected given the co-movement of urbanisation and affluence in economic transitions. However, this correlation remains just below the multicollinearity threshold of 0.90, and the remaining pairwise correlations (e.g., with ln_IndusShare) are notably lower and negative. These results are consistent with the VIF diagnostics, which all fall under the standard cutoff of 5.0, indicating no significant risk of multicollinearity. As such, the inclusion of all three variables is statistically appropriate and structurally justified.

This econometric formulation, using panel fixed-effects and heteroskedasticity-robust inference, is widely accepted for environmental impact decomposition models with repeated cross-sectional data (Bergmeir *et al*., 2018).

C. XGBoost Forecasting Strategy

To address the STIRPAT model's limitations in capturing nonlinear relationships and interactions, the hybrid framework incorporates XGBoost (Extreme Gradient Boosting) for predictive modeling. Rather than directly forecasting CoalShare, the framework separately predicts CoalConsumption and TotalEnergyConsumption, then calculates the coal share ratio to ensure consistency with energy accounting identities. This two-stage structure prevents statistical distortions from modelling compositional ratios directly and improves the model's physical interpretability.

For each prediction target

$Y_{it} \in \{ CoalConsumption_{it}, TotalEnergyConsumption_{it} \}$, the model estimates a nonlinear function as shown in Eq. (2):

$$\hat{Y}_{it} = f(X_{it}) + \varepsilon \qquad (2)$$

Where X_{it} includes the same log-transformed structural variables from the STIRPAT specification:

- ln(UrbanRate $_{it}$)
- ln(GDPperCap it)
- In(IndusShare it)
- Yeart

To capture latent nonlinear interactions, three interaction terms are additionally constructed, as shown in Eqs. (3)-(5):

$$Interact1_{it} = ln(UrbanRate_{it}) \times ln(IndusShare_{it}) \quad (3)$$
$$Interact2_{it} = ln(GDPperCap_{it}) \times ln(IndusShare_{it}) \quad (4)$$
$$Interact3_{it} = ln(UrbanRate_{it}) \times Year_t \quad (5)$$

These interactions allow the model to encode region-specific elasticities that vary with structural characteristics such as industrial intensity or urban development stage, which would be infeasible under a linear framework.

The model was trained on data from 2005 to 2019 (N=90), while the period 2020-2021 (N=12) was held out for out-of-sample testing. This temporal hold-out approach follows common protocols in machine learning with time-series data (Chen & Guestrin, 2016), helping improve the model's generalisability. To strengthen robustness and reduce overfitting, we implemented a five-fold cross-validation procedure along with a comprehensive grid search over 108 hyperparameter combinations within the training set. This tuning process entailed 540 training iterations in total, aiming to minimise out-of-fold prediction error (Hutter *et al*., 2019). Additional details regarding hyperparameter tuning can be found in Appendix D. The final model configuration was chosen to balance complexity and predictive performance, taking advantage of XGBoost's inherent regularisation properties (Liang *et al*., 2021), and is outlined below:

- ☐n_estimators = 700
- ☐max_depth = 7
- ☐learning_rate = 0.05
- ☐subsample = 0.8
- ☐colsample_bytree = 0.8

This configuration, which combines a moderate learning rate and tree depth with a large number of estimators, enables the model to gradually capture complex patterns, thereby reducing the risk of overfitting often associated with small panel datasets (Molnar, 2022). The use of subsampling parameters-specifically subsample and colsample_bytree-further improves generalisation by introducing stochasticity into the training process, a fundamental regularisation feature of the algorithm.

D. SHAP-Based Model Interpretation

To enhance the interpretability of the XGBoost model and evaluate its alignment with the econometric findings, this study employs SHAP (Shapley Additive Explanations)-a model-agnostic interpretation method based on cooperative game theory (Aas *et al*., 2021). SHAP decomposes predictions into contributions from individual features by calculating each feature's average marginal contribution across all possible subsets of features. This ensures both local accuracy and global consistency, addressing the transparency limitations of black-box models. Global SHAP values represent average feature importance across the dataset, while local values provide province-year-specific explanations.

This dual-level attribution allows validation of theoretical consistency by comparing SHAP rankings with STIRPAT coefficients. It also reveals conditional effects (e.g., nonlinear urbanisation-industry interactions) that linear regressions cannot capture. The approach thus bridges accuracy and structure, reinforcing the coherence of the hybrid model (Hyndman & Athanasopoulos, 2018). The SHAP analysis was implemented using the Python shap package (version 0.41) in combination with the TreeExplainer module designed for tree-based models such as XGBoost. This ensures compatibility between predictive outputs and post-hoc interpretability layers.

E. Model Validation and Performance Evaluation

The model is trained on 2005-2019 data and validated on 2020-2021 to ensure temporal robustness. For benchmarking, Ordinary Least Squares (OLS) and Auto regressive Integrated Moving Average (ARIMA) (1,1,1) models are fitted on 2005-2017 data and tested on 2018-2019. This design reflects realistic forecasting conditions and aligns with best practices in time-series learning (International Energy Agency, 2022).

Forecast accuracy is evaluated using three standard metrics: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE), as shown in Eqs. (6)-(8):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\widehat{Y_{it}} - Y_i\right| \tag{6}$$

$$RSME = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{Y_{it}} - Y_i\right)^2} \tag{7}$$

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{\widehat{Y_{it}} - Y_i}{Y_i}\right| \tag{8}$$

These indicators collectively assess both absolute and relative forecast deviation, enabling consistent comparison across methods under limited-data constraints.

III. RESULT AND DISCUSSION

A. Results

1. STIRPAT Model: Structural Driver Analysis

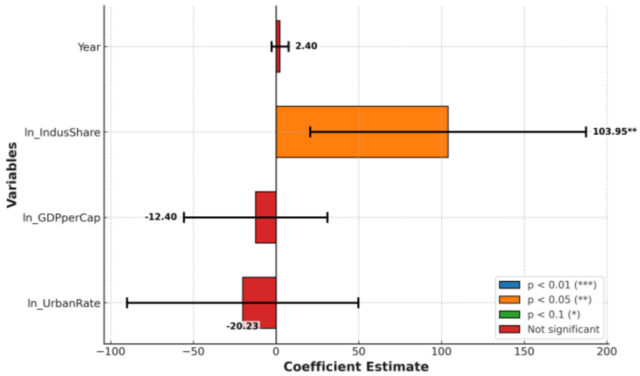

Figure 2. Coefficient estimates from the STIRPAT model for provincial coal share (2005-2021).

The fixed-effects panel regression results reveal the structural determinants of provincial coal consumption patterns (Figure 2Figure 2). The model demonstrates strong overall explanatory power with an adjusted R^2 of 0.885 and a highly significant F-statistic ($F = 77.23$, $p < 0.001$), indicating that the selected structural variables jointly explain the majority of variation in coal share across provinces and time. Full regression results are available in Appendix F.

The fixed-effects panel regression results identify industrial structure as the primary driver of provincial coal consumption patterns (Figure 2). The coefficient for ln(IndusShare) is 103.95 ($p < 0.01$), indicating a strong positive relationship with coal share. It should be noted that this elasticity estimate may be influenced by the bounded nature of the CoalShare variable, particularly in provinces where coal dependence approaches upper limits.

While the direction and significance of this relationship are clear, the precise magnitude of the elasticity should be interpreted with caution. The results confirm that provinces with manufacturing-intensive economies maintain higher coal consumption levels but also suggest that the relationship may not be fully captured by a simple log-linear specification.

The coefficients for urbanisation (-20.23) and per capita income (-12.40) are negative but statistically insignificant, indicating that their effects may be conditional or nonlinear. The time trend coefficient (2.40) is also statistically insignificant, suggesting no uniform temporal decline in coal share after controlling for structural factors.

These findings highlight the need for more flexible modelling approaches to capture the complex relationships between structural factors and energy consumption patterns, which we address through the XGBoost framework in the following section.

2. XGBoost Model: Predictive Performance Assessment

Table 1. Forecasting performance of XGBoost, OLS, and ARIMA models (2020-2021)

Variable	Model	MAE	RMSE	MAPE (%)
CoalConsumption	XGBoost	1154.61	1403.33	5.26
	OLS	514.83	592.85	3.79
	ARIMA (1,1,1)	810.71	958.58	3.79
TotalEnergyConsumption	XGBoost	980.47	1211.08	3.02
	OLS	579.06	601.35	2.61
	ARIMA (1,1,1)	453.98	481.02	2.07

The XGBoost forecasting model demonstrates competitive performance across both target variables, with results benchmarked against traditional statistical approaches (Table 1). For coal consumption forecasting, XGBoost achieves an MAE of 1154.61 (10,000 tonnes), RMSE of 1403.33, and MAPE of 5.26%. While OLS regression shows lower MAE (514.83) and ARIMA demonstrates comparable MAPE (3.79%), XGBoost provides superior handling of structural nonlinearities and missing data patterns.

For total energy consumption, XGBoost records an MAE of 980.47, RMSE of 1211.08, and MAPE of 3.02%, representing strong predictive accuracy. Traditional methods show mixed performance: ARIMA achieves the lowest MAPE (2.07%) for total energy consumption, while OLS demonstrates intermediate accuracy levels.

Importantly, XGBoost's primary advantage lies not in raw forecasting accuracy but in its capacity to incorporate multiple structural features simultaneously while providing interpretable output through SHAP analysis. The model successfully captures complex interactions between urbanisation, economic development, and industrial structure that linear approaches cannot accommodate.

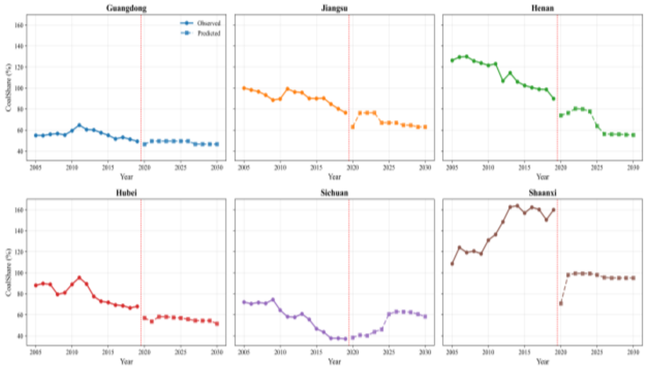

Figure 3. Observed and predicted coal share trajectories for the six provinces (2005-2030).

Figure 3 illustrates the model's capability to reproduce historical CoalShare trajectories and generate plausible future projections through 2030. A vertical red line marks the division between the training period (2005-2019) and the forecast horizon (2020-2030). Predictions for 2020–2021 align closely with observed values across all six provinces, indicating reliable near-term forecasting performance.

Provincial trends display clear divergence: Guangdong and Jiangsu sustain consistently low coal shares with a gradual decline, whereas Shaanxi maintains a high dependence on coal with limited structural shift. Henan and Hubei show intermediate levels of coal use, accompanied by moderate fluctuations, consistent with their transitional energy mix characteristics.

3. SHAP Analysis: Feature Importance and Structural Consistency

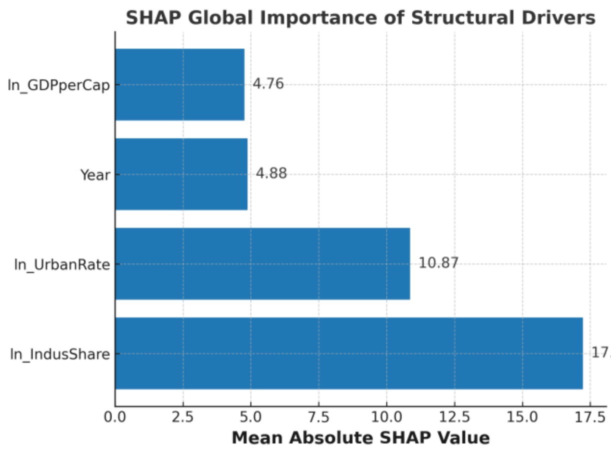

Figure 4. Global SHAP feature importance for the coal consumption prediction model.

SHAP global importance analysis confirms the structural insights derived from STIRPAT regression while revealing the relative significance of each factor in the machine learning context (Figure 4). Industrial share (ln_IndusShare) dominates feature importance with a mean absolute SHAP value of 17.23, consistent with its statistical significance in the econometric analysis.

Urbanisation (ln_UrbanRate) is identified as the second most influential factor, with a SHAP value of 10.87-notably exceeding what its econometric significance would imply. This divergence suggests that the impact of urbanisation may involve strong nonlinear or interactive effects, which machine learning methods are better suited to capture. In comparison, the time trend (Year) and income level (ln_GDPperCap) show more modest contributions, with SHAP values of 4.88 and 4.76, respectively.

The agreement between STIRPAT and SHAP results regarding the role of industrial structure reinforces the internal consistency of the hybrid framework. Conversely, the discrepancies observed for other variables underscore the benefit of integrating both linear and nonlinear modelling approaches. Moreover, province-specific driver rankings (Appendix B) and normalised contribution analysis (Appendix E) point to substantial regional heterogeneity, emphasising that factors such as urbanisation and industrial structure exert differentiated influence across China.

4. SHAP-Based Interaction Effects

While global SHAP values provide insight into the average contribution of each structural variable, they do not capture how these effects vary under different contextual conditions. To address this, SHAP interaction plots are generated to examine how the marginal impact of one variable is conditioned by another. These visualisations uncover structural dependencies that are often masked in additive models.

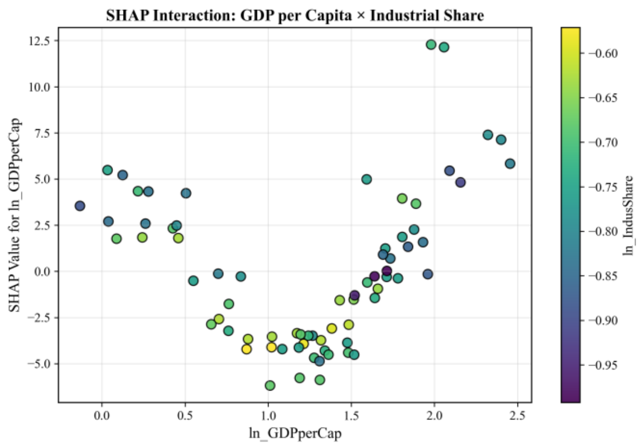

Figure 5. SHAP interaction dependence plot: GDP per capita and industrial share.

SHAP interaction plots reveal complex conditional relationships that linear models cannot capture.

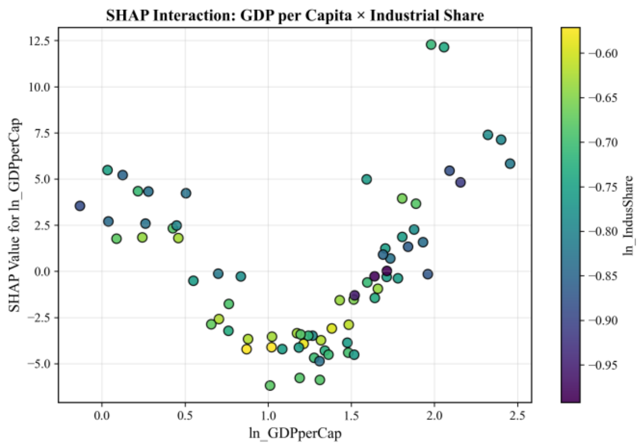

Figure 5 demonstrates the interaction between GDP per capita and industrial share, where the effect of income on coal consumption varies substantially with industrial intensity. In provinces characterised by lower industrial shares (represented by yellow data points), elevated per capita GDP correlates with a reduction in coal consumption. Conversely, in regions with higher industrial intensity (indicated by blue points), this relationship reverses and turns positive. This suggests that economic affluence may in fact reinforce coal dependency within manufacturing-intensive regions, a nuance effectively captured by the STIRPAT-model-informed analysis.

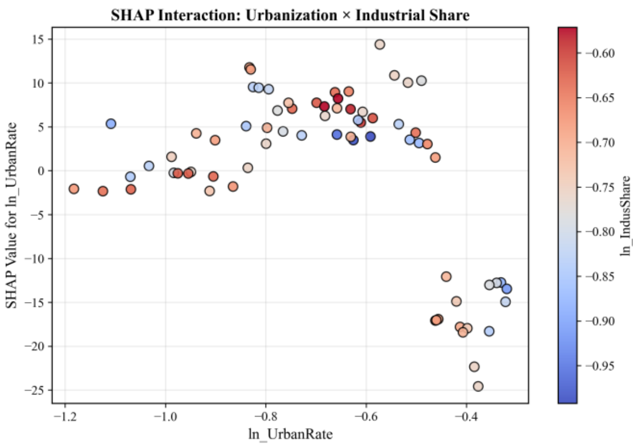

Figure 6. SHAP interaction dependence plot: urbanisation rate and industrial share.

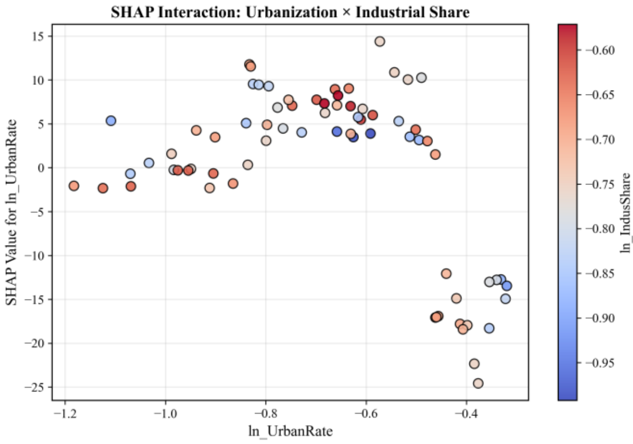

Figure 6 illustrates how urbanisation effects depend on industrial structure. When industrial share is low (blue regions), urbanisation shows positive SHAP values, potentially reflecting energy-intensive suburban expansion. Conversely, in highly industrialised provinces (red regions), urbanisation demonstrates negative or neutral effects, indicating structural efficiency gains or service sector transitions.

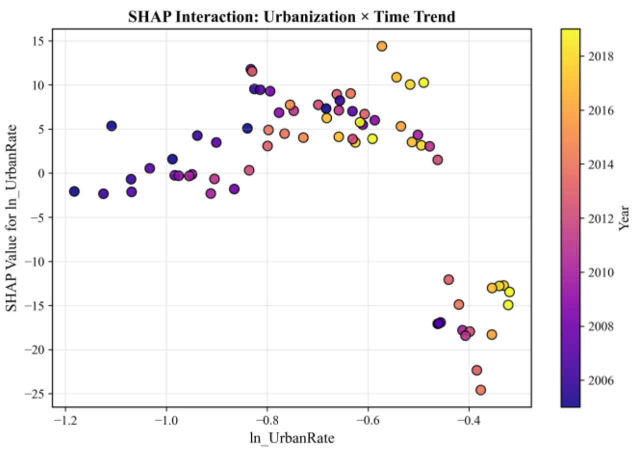

Figure 7. Temporal SHAP value distribution for the urbanisation rate variable (2005-2021).

Figure 7 illustrates the shifting influence of urbanisation over time. In earlier years (depicted by purple points), its impact was primarily positive, whereas in more recent periods (shown in yellow), this effect has diminished or reversed. This temporal pattern indicates a fundamental shift in how urbanisation relates to coal consumption, likely driven by policy changes, technological advances, or broader economic restructuring.

These interactive dynamics highlight that the drivers of coal consumption are not only interconnected but also exhibit substantial variability over time and space, underscoring the importance of flexible modelling frameworks capable of capturing such complexity.

5. Provincial Trajectory Analysis: Regional Heterogeneity

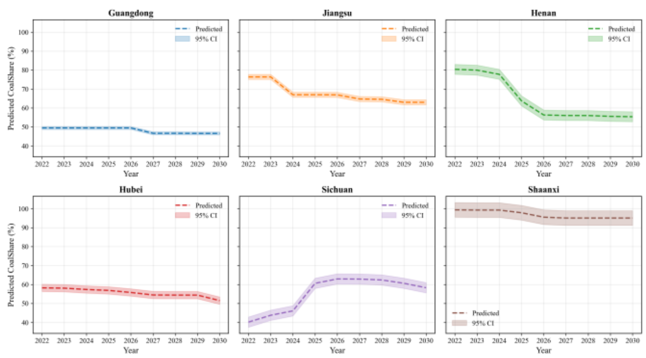

Figure 8. Forecasted provincial coal share trajectories from the hybrid framework (2022-2030).

The projected coal share trajectories up to 2030 reveal starkly different patterns across provinces (Figure 8), underscoring the regional heterogeneity in China's energy transition. Appendix D details the SHAP-derived contributions of structural factors for each province. Based on these projections, we categorise the six provinces into three distinct groups:

Stable High-Coal Regions: Shaanxi's coal share is projected to remain high with little decline, indicating strong structural inertia. Its flat trajectory and narrow confidence intervals suggest limited capacity for diversification, a common trait among resource-dependent western provinces.

Rapid Transition Regions: Henan and Jiangsu show a clear and accelerating decline in coal share, especially after 2025. This trend aligns with substantial structural adjustments in these regions, likely driven by policy-led industrial upgrading and diversification.

Plateau Regions: Sichuan, Hubei, and Guangdong exhibit relatively stable trends with minor fluctuations. Guangdong sustains a low coal share, consistent with its advanced service economy, while Sichuan's profile is shaped by a balanced combination of hydroelectric and thermal power generation.

These heterogeneous pathways highlight that China's energy transition is far from uniform. Our findings stress the necessity for province-specific strategies that account for local economic structures and developmental stages.

B. Discussion

1. Methodological Contributions and Framework Validation

This study proposes a practical and adaptable forecasting framework that integrates econometric and machine learning approaches to address subnational energy forecasting under data scarcity. The framework combines STIRPAT, XGBoost, and SHAP, leveraging their complementary strengths: theoretical grounding through elasticity analysis, flexible pattern recognition via tree-based algorithms, and post-hoc interpretability based on game-theoretic feature attribution.

The strong agreement between STIRPAT coefficients and SHAP-derived feature importance confirms the internal consistency of the framework, particularly in identifying industrial structure as the primary driver of coal consumption. At the same time, machine learning reveals nuanced, context-dependent effects of urbanisation and income-effects that traditional models often overlook. These divergences do not reflect contradiction but rather demonstrate SHAP's capacity to uncover conditional relationships that are masked in global regression estimates.

Beyond internal validation, the framework is designed for transferability to other data-constrained settings, especially in developing countries with fragmented or heterogeneous energy systems. Unlike conventional econometric models that require long time series, our approach remains robust even with short panel data, enhancing its practicality in regions with limited official statistics.

Benchmarking results indicate that while XGBoost does not always outperform traditional models in sheer predictive accuracy, it offers a superior ability to incorporate multiple structural predictors while maintaining interpretability. This aligns with the study's focus on methodological transparency and policy relevance, rather than optimization alone. In contrast to black-box AI tools, the hybrid framework emphasises clarity, traceability, and diagnostic value for decision-making.

2. Policy Implications for Provincial Decarbonisation

Our empirical findings underscore the necessity of region-specific strategies to steer China's coal transition. In eastern provinces such as Guangdong and Jiangsu—where coal shares are already low and declining—policy efforts should focus on sustaining this momentum through market-based renewable energy incentives, support for digital grid modernisation, and electrification of transport and services. Their diversified economic structures enable these regions to pilot next-generation energy technologies without compromising supply stability.

In contrast, central provinces like Henan and Hubei exhibit steady but slower decarbonisation trajectories and require targeted policy interventions to overcome structural inertia. Although less coal-dependent than western provinces, they often lack the fiscal and institutional capacity to accelerate transitions independently. Measures such as tiered electricity pricing, performance-linked subsidies, and regional carbon trading pilots could improve the cost-effectiveness and implementation feasibility of such interventions. Coupling these with industrial upgrading initiatives may produce more durable outcomes than standardised policies.

Western provinces-particularly Shaanxi-remain constrained by structural dependencies on coal-related employment, infrastructure, and fiscal revenue. Here, rapid coal substitution is impractical, necessitating longer-term strategies. Central government support is crucial, both in financing renewable energy infrastructure and promoting economic diversification to reduce reliance on extractive industries. Without dedicated investments and institutional capacity-building, market-based reforms alone are unlikely to spur substantial change (Wilson *et al.*, 2012).

SHAP interaction analysis further reveals that the effects of urbanisation and income are highly contextual, implying that uniform policies may lead to divergent provincial outcomes. For example, income growth can reinforce fossil energy use in industry-intensive provinces, while potentially supporting efficiency gains in service-oriented regions. These insights reinforce the importance of tailored strategies that reflect local economic structures and development stages. International experience, such as the state-level energy transition paths in the United States and Germany's regionally varied coal phase-out plans, further illustrates that subnational differentiation is not only appropriate but necessary for feasible decarbonisation (Steckel *et al.*, 2013).

3. Limitations and Future Research Directions

Several limitations constrain the generalisability and scope of these findings. The sample includes only six provinces, potentially limiting representativeness across China's diverse regional landscape. Future research should expand coverage to include additional provinces and autonomous regions, particularly those with unique economic structures or energy endowments.

Furthermore, this study relies on official statistical yearbooks, which, despite being the most comprehensive source, may contain inherent data biases that could influence the results. A significant concern is the systematic under-reporting or misclassification of coal consumption, particularly in regions with heavy reliance on small, informal, or inefficient industrial boilers and furnaces, which are often poorly monitored (Liu *et al.*, 2015). Such non-random measurement errors could lead to an underestimation of the true coal share and attenuate the measured strength of the relationship between industrial structure (ln_IndusShare) and coal dependence. Consequently, the SHAP-derived feature importance from the XGBoost model might also be affected, potentially altering the perceived ranking of key drivers. Future research should aim to integrate alternative data sources, such as remote sensing data for nighttime lights or atmospheric pollutants, to cross-validate official statistics and improve the accuracy of energy transition forecasts.

The analysis excludes several potentially important factors such as technological innovation indicators, policy enforcement strength, and renewable energy capacities due to data availability constraints. Incorporating such variables, where measurable, could enhance both forecasting precision and policy targeting.

Although the hybrid framework performs reliably within the observed sample, several limitations remain. First, the model does not account for external disruptions such as the COVID-19 pandemic, volatile global energy prices, or sudden shifts in regulatory policy. These shocks can introduce structural breaks that are difficult to anticipate using trend-based models alone. Incorporating scenario-based simulations or variables that proxy such disruptions would improve the model's adaptability under uncertainty. Second, the current study period (2005-2021) predates major developments under China's 14th Five-Year Plan and the full-scale implementation of the dual carbon strategy. As post-2022 data becomes available, re-estimating the model will be important for capturing new dynamics, validating projection accuracy, and refining long-term transition trajectories.

Third, while this study focuses on forecasting coal share at the provincial level, it does not encompass broader aspects of energy system transformation. Key dimensions such as electrification progress, renewable energy integration, and power market reform are not yet reflected in the model structure. Future research could expand this hybrid approach to address multi-sector interactions or adopt a whole-system modelling perspective, enabling a more comprehensive assessment of China's energy transition pathway.

4. Applicability and Sample Scope Considerations

Although this study covers only six provinces, the hybrid STIRPAT-XGBoost-SHAP framework is structurally modular and adaptable, making it suitable for application in other subnational contexts where similar energy and socioeconomic indicators are available. The model architecture is modular and can be retrained with minimal modifications across different regions, assuming availability of basic structural indicators such as industrial share, urbanisation rate, and income level. This makes the framework highly transferable to other emerging economies or decentralised energy systems facing similar challenges in coal share and structural heterogeneity.

Nonetheless, the limited sample scope introduces constraints on statistical generalisability and scenario diversity. While the selected provinces span a range of development stages and energy profiles, the framework may not fully capture outlier dynamics (e.g., in highly autonomous regions or special economic zones). To mitigate small-sample limitations, the study leverages panel fixed effects, robust standard errors, and SHAP-based disaggregation to ensure model stability and interpretability under constrained data conditions. Future work will aim to expand the provincial panel, incorporate more recent data, and test cross-country applicability to validate external robustness and policy relevance in broader engineering applications.

IV. CONCLUSION

In response to the need for accurate and interpretable subnational energy forecasting, this study has developed and validated a hybrid STIRPAT-XGBoost-SHAP framework for provincial coal consumption prediction in China. The integration of econometric theory, machine learning prediction, and interpretable attribution provides a comprehensive approach to subnational energy forecasting that balances structural understanding with predictive capability.

Key findings include:

(1) Industrial structure emerges as the dominant driver of provincial coal consumption patterns, confirmed across both econometric and machine learning analyses;

(2) Urbanisation and income effects are highly nonlinear and context-dependent, requiring flexible modelling approaches to uncover their true influence;

(3) Forecasted coal transition trajectories display significant heterogeneity across provinces, underscoring the need for differentiated and locally grounded policy responses;

(4) The hybrid framework demonstrates internal consistency and practical feasibility for policy-oriented energy modelling in data-constrained environments.

The methodology contributes to the energy forecasting literature by bridging traditional econometric frameworks with modern machine learning techniques, while retaining interpretability essential for policy formulation. Unlike black-box models that emphasise predictive accuracy at the expense of transparency, this study highlights the value of structural interpretability for informing real-world decision-making. The findings affirm that decarbonisation strategies must be tailored to regional conditions to effectively support China's dual carbon goals, especially in light of pronounced differences across provinces in industrial structure, urbanisation pathways, and governance capabilities.

Integrating SHAP into the forecasting workflow enables province-level diagnostic insights, revealing the relative influence of key drivers and offering actionable guidance for policymakers designing local energy transition roadmaps. From an engineering standpoint, these insights help prioritise infrastructure investments-such as aligning capacity expansion with regional demand patterns, optimising the sequence of grid upgrades, and improving system resilience amid uncertainty.

Future studies should extend the analysis to include more provinces with varying developmental backgrounds. Incorporating additional indicators-such as renewable energy adoption rates, regulatory effectiveness, or industrial modernisation metrics-could strengthen the model's policy relevance and explanatory power. This hybrid framework may also be extended to simulate system-wide transitions, including renewable integration, power market restructuring, and carbon pricing mechanisms.

In summary, this study provides a scalable and transferable framework for subnational energy modelling under data constraints. With improved data availability and computational tools, the approach can serve as an effective decision-support tool for devolving climate planning and infrastructure strategies in other emerging economies undergoing similar energy transitions.

V. ACKNOWLEDGEMENT

This work was supported by Universiti Sains Malaysia, Bridging with Project No: R501-LR-RND003-0000000913-0000.

VI. REFERENCES

Aas, K, Jullum, M & Løland, A 2021, ‘Explaining individual predictions when features are dependent: More accurate approximations to Shapley values’, Artificial Intelligence, vol. 298, p. 103502. Doi: 10.1016/j.artint.2021.103502

Bergmeir, C, Hyndman, RJ & Koo, B 2018, ‘A note on the validity of cross-validation for evaluating autoregressive time series prediction’, Computational Statistics & Data Analysis, vol. 120, pp. 70-83. Doi: 10.1016/j.csda.2017.11.003

Carley, S & Konisky, DM 2020, ‘The justice and equity implications of the clean energy transition’, Nature Energy, vol. 5, no. 8, pp. 569-577. Doi: 10.1038/s41560-020-0641-6

Chen, T & Guestrin, C 2016, ‘XGBoost: A scalable tree boosting system’, in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York: ACM, pp. 785-794. Doi: 10.1145/2939672.2939785

Hutter, F, Kotthoff, L & Vanschoren, J (eds.) 2019, ‘Automated Machine Learning: Methods, Systems, Challenges’, Cham: Springer Nature.

International Energy Agency 2022, World Energy Outlook 2022, Paris: IEA, <https://www.iea.org/reports/world-energy-outlook-2022 >.

Liang, Y, Li, J, Zhou, H & Liu, L 2021, ‘Comparison of machine learning algorithms for predicting residential electricity consumption in China’, Energy, vol. 229, p. 120719. Doi: 10.1016/j.energy.2021.120719

Liu, J, Zhang, Z & Wang, Y 2021, ‘Regional differences in coal dependency across China: A provincial-level comparison’, Energy Policy, vol. 158, p. 112528. Doi: 10.1016/j.enpol.2021.112528

Liu, Z, Guan, D, Wei, W, Davis, SJ, Ciais, P, Bai, J & He, K 2015, ‘Reduced carbon emission estimates from fossil fuel combustion and cement production in China’, Nature, vol. 524, no. 7565, pp. 335-338. Doi: 10.1038/nature14677

Lundberg, SM & Lee, S-I 2017, ‘A unified approach to interpreting model predictions’, in Advances in Neural Information Processing Systems 30 (NIPS 2017), <https://proceedings.neurips.cc/paper/2017>.

Molnar, C 2022, ‘Interpretable Machine Learning: A Guide for Making Black Box Models Explainable’, 2nd edn, <https://christophm.github.io/interpretable-ml-book/>.

National Bureau of Statistics of China & National Energy Administration (2006-2022) China Statistical Yearbook & China Energy Statistical Yearbook, Beijing: China Statistics Press, <https://www.stats.gov.cn>.

Ohlhorst, D 2015, ‘Germany's energy transition policy and regional implementation’, Energy, Sustainability and Society, vol. 5, no. 1, pp. 1-13. Doi: 10.1186/s13705-015-0041-6

Steckel, JC, Jakob, M & Edenhofer, O 2013, ‘From carbonization to decarbonization? Past trends and future scenarios for China's CO_2 emissions’, Energy Policy, vol. 60, pp. 759-769. Doi: 10.1016/j.enpol.2013.05.002

Wilson, C, Grubler, A, Gallagher, KS & Nemet, GF 2012, ‘Marginalization of end-use technologies in energy innovation for climate protection’, Nature Climate Change, vol. 2, no. 11, pp. 780-788. Doi: 10.1038/nclimate1576

Wu, X, Zhou, K & Lin, B 2023, ‘Urban form, machine learning, and energy intensity: An interpretable modeling framework’, Energy Economics, vol. 120, p. 106664. Doi: 10.1016/j.eneco.2023.106664

Yang, H, Liu, J & Tang, X 2022, ‘Improving the explainability of energy forecasting using SHAP and interpretable ensembles’, Applied Energy, vol. 313, p. 118858. Doi: 10.1016/j.apenergy.2022.118858

York, R, Rosa, EA & Dietz, T 2003, ‘STIRPAT, IPAT and ImPACT: analytic tools for unpacking the driving forces of environmental impacts’, Ecological Economics, vol. 46, no. 3, pp. 351-365. Doi: 10.1016/S0921-8009(03)00188-5

Zhang, Y, Bai, H & Liu, Y 2021, ‘Decoupling and coordination between economic development and carbon emissions in China's provinces: Spatial-temporal heterogeneity and influencing factors’, Journal of Cleaner Production, vol. 279, p. 123838. Doi: 10.1016/j.jclepro.2020.123838