World Journal of Economics and Business Research
Print ISSN: 2960-0073
Online ISSN: 2960-0081
DOI: https://doi.org/10.61784/wjebr3100

A COMBINED EWM AND MULTIVARIATE STATISTICAL APPROACH FOR EVALUATING GREEN FINANCE AND ECONOMIC SUSTAINABILITY

HongYue Deng
School of Thermal Engineering, Shandong Jianzhu University, Jinan 250101, Shandong, China.

Abstract: Unraveling the nexus between financial allocation and ecological performance demands a framework capable of dismantling complex spatial heterogeneities. This study constructs a multidimensional statistical pipeline analyzing cross-sectional data across thirty Chinese provinces. Following objective dimension reduction via the entropy weight method, spatial clustering strictly partitions the landscape into distinct developmental tiers. Canonical correlation analysis subsequently extracts orthogonal structural equations, yielding a systemic coupling coefficient of 0.976. The derived loadings explicitly quantify a dual-edged transition mechanism: ecological sustainability relies heavily on capital friction against high-energy sectors (loading: -0.869), running parallel to targeted capitalization that accelerates industrial emission reductions (loading: 0.784). Beyond immediate operational insights, this research provides a mathematically rigorous blueprint for macroeconomic restructuring. By mapping these exact directional constraints, the framework equips policymakers to abandon homogeneous interventions, facilitating precision green resource allocation across highly polarized regional economies.
Keywords: Multivariate statistical analysis; Entropy weight method; Spatial clustering; Canonical correlation analysis; Eco-sustainability evaluation

1 INTRODUCTION

Transitioning toward ecological sustainability requires a fundamental paradigm shift driven by capital reallocation. Within this macroeconomic restructuring, green finance operates as a critical institutional mechanism to divert liquidity from environmentally degrading sectors toward sustainable innovations, a pathway increasingly recognized globally for achieving industrial carbon neutrality [1]. However, decoding the precise efficacy of this mechanism remains a significant empirical challenge. Large-scale economies inherently possess vast geographical disparities, ensuring that financial resource allocation and real-economy ecological responses rarely follow simple linear trajectories. Instead, they constitute a highly complex system characterized by profound spatial heterogeneity, as regional green finance and eco-economic development frequently exhibit severe structural mismatches and polarization trends [2,3].

A systematic review of contemporary literature reveals critical methodological limitations in evaluating this ecosystem. Traditional empirical models often rely on heavily aggregated indicators. Previous studies frequently proxy green finance with singular metrics, such as green credit volumes or regional bond issuance scales, a practice that tends to simplify the multidimensional characteristics of financial inputs. The widespread application of standard panel regression also restricts analytical insights to average treatment effects. Since conventional linear frameworks are primarily designed for unidirectional relationships, they struggle to capture complex coupling mechanisms across macroeconomic sub-systems. Therefore, machine learning and data fusion methods are increasingly applied to parse these non-linear indicators [4,5]. To overcome these analytical barriers, advanced algorithmic optimizations and data-driven evaluations are required to prevent generalized conclusions that overlook regional polarization [6]. Consequently, this study constructs a multidimensional pipeline integrating spatial clustering with Canonical Correlation Analysis to accurately map these simultaneous interactions.

To systematically breach these analytical barriers, this study constructs a multidimensional multivariate statistical pipeline. By establishing a dual-system evaluation framework comprising specific financial input and ecological output vectors, the research integrates objective dimension reduction with unsupervised spatial clustering and structural equation mapping. The core objective is to mathematically dismantle the complex coupling structure between green finance and economic sustainability, explicitly quantifying the exact directional forces dictating the ecological transition across highly polarized regional landscapes.

2 METHODOLOGY

To objectively evaluate the complex interplay between green finance mechanisms and economic sustainability, this study proposes a combined multivariate statistical approach. The analytical pipeline is systematically divided into four sequential stages, ranging from dual-system panel data extraction to canonical correlation analysis.

As illustrated in Figure 1, the complete research framework operates as a closed-loop evaluation system. Initially, cross-sectional panel data representing the two distinct systems are constructed. Subsequently, an objective dimension reduction algorithm is executed to assign mathematical weights and compute comprehensive scores. Following this data processing, the third stage introduces independent spatial clustering to computationally identify regional heterogeneities and stratification across the evaluated regions. Ultimately, orthogonal structural equations are extracted to mathematically map the deep coupling mechanisms and directional loadings between the financial inputs and ecological outputs.

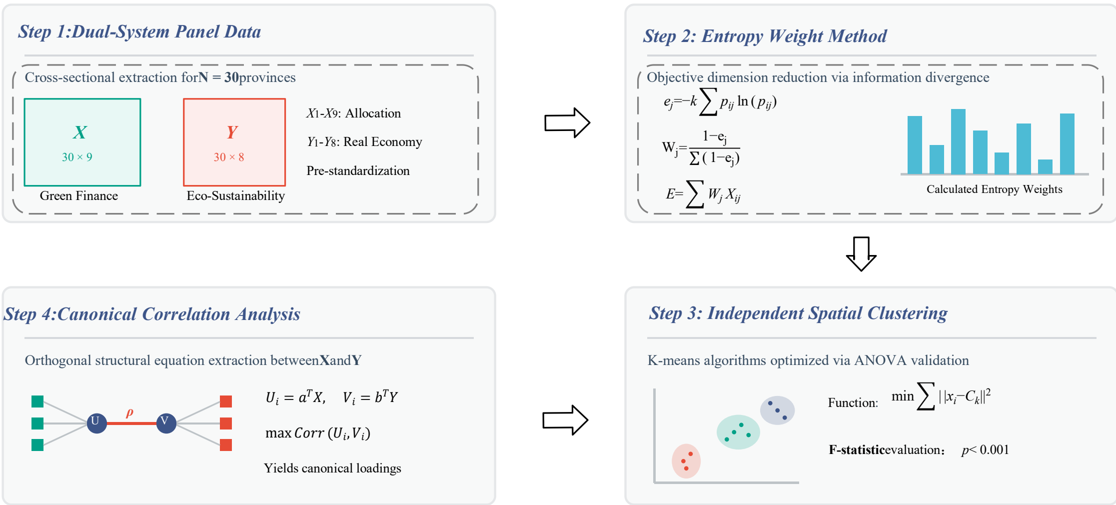

Figure 1 Research Framework of the Dual-system Evaluation Approach

2.1 Data Sources and Indicator Construction

The empirical research relies on a dual-system panel data structure comprising cross-sectional observations from thirty distinct provinces in China. To ensure the systemic integrity and computational robustness of the evaluation, a comprehensive indicator system is established.

As detailed in Table 1, the foundational variables of the dual-system evaluation framework are established.The first subsystem, designated as the allocation dimension, represents the green finance mechanism. It contains nine specific indicator vectors ranging from X_1 through X_9. These variables are strategically selected to capture the spatial distribution of financial resources across credit markets, securities, insurance, and carbon trading platforms.

Table 1 Evaluation Indicators and Polarity for the Dual Feature Space

Feature Space	Indicator Code & Description	Polarity (+/-)
Green Finance (X)	X_1: Eco-loan scale of listed companies	+
	X_2: Interest ratio of high-energy sectors	-
	X_3: Market value ratio of eco-friendly firms	+
	X_4: Market value ratio of high-energy firms	-
	X_5: Agricultural insurance payout ratio	+
	X_6: Agricultural insurance penetration depth	+
	X_7:Environmental pollution control investment ratio	+
	X_8: Ecological public expenditure ratio	+
	X_9: Carbon finance market scale	+
Economic Sustainability (Y)	Y_1: Production-level employment	+
	Y_2: Fixed asset value of large enterprises	+
	Y_3: Industrial water consumption density	-
	Y_4: Energy consumption per unit of production	-
	Y_5: Industrial value-added	+
	Y_6: Industrial SO_2 emissions	-
	Y_7: Industrial wastewater discharge volume	-
	Y_8: Industrial solid waste volume	-

Conversely, the second subsystem represents the real economy and focuses squarely on eco-sustainability. It is quantified through eight specific indicators, designated as Y_1 through Y_8. These metrics serve as the output response variables, reflecting the ecological state and operational health of regional economies through measures such as energy consumption intensities and environmental emission burdens. Prior to algorithm implementation, all raw data undergo rigorous pre-standardization to eliminate dimensional discrepancies, normalizing both positive and negative indicators into a uniform scale for mathematical consistency.

2.2 Entropy Weight Method

To avoid human bias in subjective weighting, this study employs the Entropy Weight Method, an objective dimension-reduction technique widely used to evaluate the coupling coordination between financial allocation and economic sustainability [7]. Fundamentally based on the principle of information divergence, this objective algorithm calculates the weight of each indicator according to the degree of mathematical variation inherent in the dataset [8].
First, the information entropy, denoted as e_j, for the *j*-th indicator is defined by Eq. (1):

$$e_j = -k\sum_{i=1}^{n} p_{ij} \ln(p_{ij}) \tag{1}$$

where p_{ij} represents the normalized proportion of the *i*-th province under the *j*-th indicator, and the adjustment constant k equals $1/\ln(n)$.
Subsequently, the objective weight, denoted as W_j, is derived based on the calculated information divergence by Eq. (2):

$$W_j = \frac{1-e_j}{\sum(1-e_j)} \tag{2}$$

Finally, the comprehensive evaluation score E for the overall system is aggregated by Eq. (3):

$$E = \sum W_j X_{ij} \tag{3}$$

This objective dimension reduction process ensures that indicators exhibiting higher data volatility are systematically assigned greater analytical significance in the subsequent statistical modeling.

2.3 Independent Spatial Clustering via K-means

Given the vast disparities across provinces, treating the dataset homogeneously obscures regional heterogeneity. Thus, data-driven spatial clustering approaches have been adopted to objectively identify regional development patterns and spatial agglomeration [9,10]. Therefore, this study introduces the K-means clustering algorithm to perform independent spatial profiling.
The K-means algorithm computationally partitions the thirty provinces into a predefined number of distinct clusters by minimizing the within-cluster sum of squared Euclidean distances. The objective function is formulated by Eq. (4):

$$\min \sum_{i=1}^{K} \sum_{x \in S_i} || x - \mu_i ||^2 \tag{4}$$

where S_i represents the i-th cluster, x is the specific data point representing a province profile, and μ_i serves as the centroid of cluster S_i.
To rigorously validate the optimal number of clusters and ensure that the algorithmic partitioning is statistically robust rather than arbitrary, Analysis of Variance is deployed. The clustering effectiveness is confirmed strictly through an F-statistic evaluation, requiring a highly significant threshold where the p-value is strictly less than 0.001.

2.4 Canonical Correlation Analysis

While traditional multiple regression models are restricted to exploring linear relationships between multiple independent variables and a single dependent variable, they structurally fail to map the internal coupling between two high-dimensional systems. To overcome this critical limitation, Canonical Correlation Analysis, hereafter designated as CCA, is employed as the final analytical step.
CCA is mathematically designed to extract orthogonal structural equations between the Green Finance system, denoted by the vector X, and the Eco-Sustainability system, denoted by the vector Y. It seeks to identify two linear combinations, or canonical variates, U_i and V_i, defined by Eqs. (5) and (6):

$$U_i = a^T X \tag{5}$$

$$V_i = b^T Y \tag{6}$$

where the vectors a and b represent the canonical weight matrices to be estimated.
The mathematical objective of CCA is to strictly maximize the Pearson correlation coefficient between these two canonical variates, calculated by Eq. (7):

$$\max Corr(U_i, V_i) = \frac{a^T \Sigma_{XY} b}{\sqrt{a^T \Sigma_{XX} a}\sqrt{b^T \Sigma_{YY} b}} \tag{7}$$

where Σ_{XX},Σ_{YY} , and Σ_{XY} represent the respective covariance and cross-covariance matrices of the two isolated subsystems.

By solving this specific eigenvalue problem, CCA yields the canonical structural loadings. These loadings explicitly quantify the directional impact, whether functioning as a positive promotion or a negative inhibition, of individual indicators on the overarching canonical variables. This algorithmic approach provides a deep, data-driven mapping of how financial allocation inputs structurally dictate ecological sustainability outputs.

3 RESULTS

Following the methodological pipeline established previously, the empirical analysis transitions into examining the multidimensional dataset. This section sequentially unfolds the spatial clustering heterogeneities and the deep structural coupling mechanisms between the financial and ecological systems.

3.1 Spatial Heterogeneity and Clustering Characteristics

To objectively categorize the developmental stages of the evaluated regions, the spatial clustering algorithm categorizes the specific provincial profiles into three distinct developmental tiers.

As detailed in Table 2, the rigorous statistical validation for this spatial stratification is mathematically confirmed. The Analysis of Variance results strictly confirm the algorithmic partitioning, generating highly pronounced F-statistics across the evaluation indicators. Furthermore, the corresponding significance levels universally fall strictly below the 0.001 threshold. This mathematical evidence confirms that the spatial divergence across the evaluated regions is a systemic phenomenon rather than a random statistical fluctuation.

Table 2 Key K-means ANOVA Validations for Independent Systems.

Evaluated System	High-Variance Indicator	F-statistic	df1	df2	p-value
Green Finance (X)	X_7: Pollution control investment	15.908	2	27	0.000***
	X_6: Agri-insurance depth	13.923	2	27	0.000***
	X_4: High-energy market value	12.909	2	27	0.000***
Economic Sustainability (Y)	Y_5: Industrial value-added	58.612	2	27	0.000***
	Y_1: Production-level employment	56.901	2	27	0.000***
	Y_2: Fixed asset value	41.610	2	27	0.000***

As depicted in Figure 2, the dual-color scatter plot maps the respective coordinates of the evaluat onto a two-dimensional Cartesian plane defined by green finance and eco-sustainability axes. The visual distribution exhibits severe spatial heterogeneity and reveals a pronounced polarization effect. The red data points, representing the advanced developmental tier, are distinctly concentrated in the upper-right quadrant. This quadrant features benchmark regions such as Guangdong and Beijing, which demonstrate exceptional integration of financial innovation and ecological transitioning. Conversely, a dense agglomeration of cool-toned data points occupies the lower-left sector, representing the lagging tier. This density highlights that a significant number of inland and western regions remain constrained by dual systemic deficiencies, facing severe resource allocation bottlenecks.

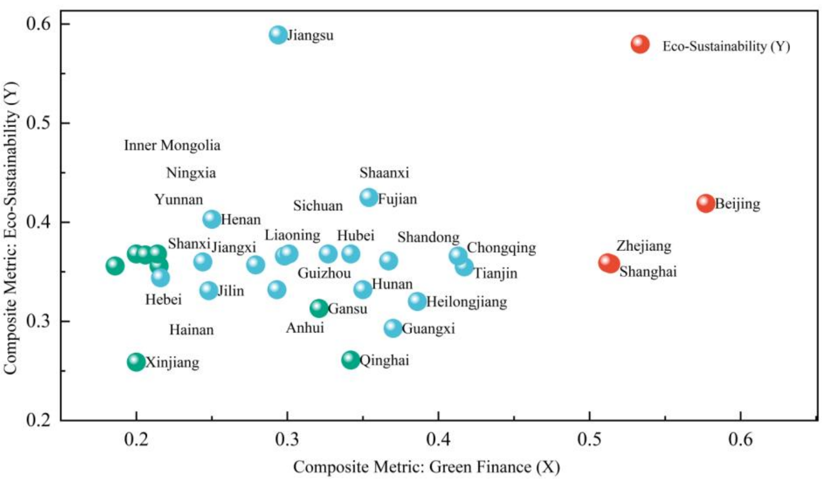

Figure 2 Spatial Clustering Distribution of Green Finance and Eco-sustainability across 30 Provinces

To deeply unpack the structural chasms driving this spatial polarization, as illustrated in Figure 3, the dual-axis centroid profile chart dissects the standardized internal metrics. By isolating the advanced tier line trajectory from the lagging tier line trajectory, a stark morphological contrast emerges. The advanced trajectory exhibits prominent peaks corresponding to sophisticated financial tools, suggesting that breakthrough progress in specific leverage mechanisms drives their overall superiority. In sharp contrast, the lagging trajectory remains consistently depressed and deeply submerged below the zero-baseline across multiple critical coordinate nodes. This visual evidence suggests that the lagging regions suffer not merely from a deficit in total financial volume, but from severe structural inadequacies across their entire resource allocation subsystems.

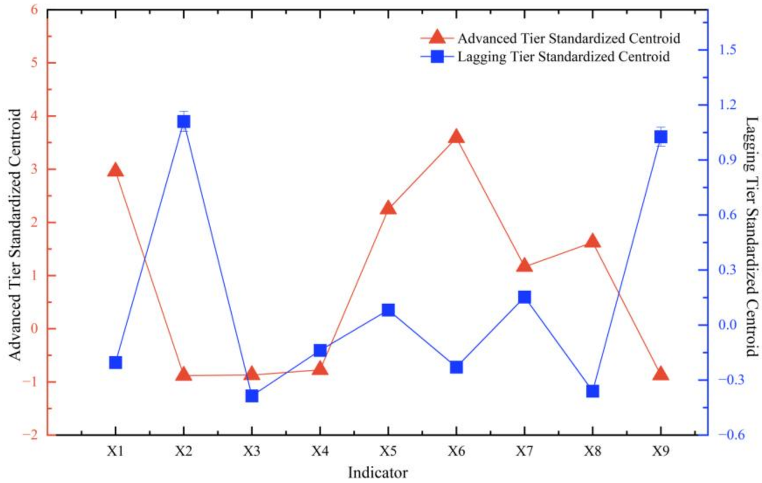

Figure 3 Dual-axis Centroid Profile Chart of Advanced and Lagging Tiers

3.2 Canonical Correlation and Structural Coupling Mechanisms

While the spatial clustering illustrates regional heterogeneity, this study primarily focuses on overall structural coupling. The clustering results are used here for descriptive purposes rather than for separate evaluations within each tier. Therefore, Canonical Correlation Analysis (CCA) is applied to the entire sample of 30 provinces to identify the general relationships between the two systems. Building localized models based on these specific sub-regions will be considered in future research. Following this whole-sample approach, CCA is employed to examine the operational linkages between the variables.

As shown in Table 3, the overall model fit validates the deployment of the structural equation extraction. The first canonical correlation function demonstrates an exceptionally high correlation coefficient. Accompanying this robust coefficient is a statistical significance value approaching zero, establishing an indestructible mathematical bridge between the green finance input vectors and the eco-sustainability output responses. This robust parameter confirms that the isolated variables indeed converge into a highly cohesive, deeply coupled macroeconomic structure.

Table 3 Canonical Correlation Coefficients and Significance Tests

Canonical Pair	Correlation Coefficient (ρ)	Eigenvalue	Wilks' Lambda (Λ)	F-statistic	df (Num)	p-value
Function 1	0.976	0.496	0.504	3.465	72.000	0.000***
Function 2	0.813	0.206	0.755	1.782	56.000	0.000***
Function 3	0.215	0.049	0.910	0.787	42.000	0.834

To precisely isolate which specific variables are driving this high-dimensional convergence, as detailed in Table 4, the structural loadings quantify the exact directional influence exerted by individual indicators. These precise numerical values serve as the blueprint for understanding the physical and economic mechanisms operating within the model.

Table 4 Canonical Structural Loadings Analysis

Feature Space	Indicator	Function 1 (U_1,V_1)	Function 2 (U_2,V_2)
Green Finance (X)	X_2: High-energy interest ratio	-0.869	0.296
	X_6: Agri-insurance depth	0.316	0.438
	X_9: Carbon finance	0.593	0.022
Economic Sustainability (Y)	Y_2: Fixed asset value	-0.181	0.315
	Y_5: Industrial value-added	0.100	0.241
	Y_6: Industrial SO_2 emissions	0.784	-0.106

Feature Space	Indicator	Function 1 (U_1,V_1)	Function 2 (U_2,V_2)
	Y_7: Industrial wastewater	0.201	0.487

As shown in Figure 4, the structural loadings are partitioned into negative inhibition and positive promotion effects.Variables with negative loadings, notably the interest share of high-energy-consuming industries, confirm a functional capital inhibition mechanism. This demonstrates that across the overall sample, restricting liquidity to heavily polluting sectors acts as a primary driver of systemic sustainability.

Conversely, positive loadings reflect effective resource allocation, closely aligning with reduced industrial emissions and carbon finance expansion. Ultimately, these results indicate that structural optimization requires the simultaneous application of financial constraints on traditional brown industries alongside targeted capitalization for ecological sectors.

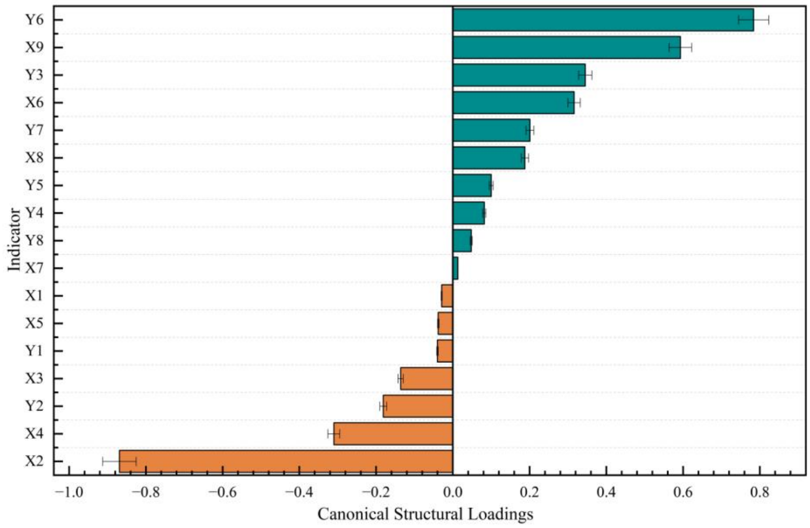

Figure 4 Canonical Structural Loadings of Green Finance and Eco-sustainability Indicators

3.3 Discussion and Theoretical Contribution

By applying this integrated statistical pipeline, this study offers a methodological contribution to macro-level sustainability evaluations. Unlike traditional regression models that test isolated variables, this overall structural mapping empirically confirms that green transition is not driven by single policy inputs, but by a dual-edged systemic mechanism. Identifying this overarching global structure provides a necessary data-driven foundation for broad macroeconomic policy design.

4 DISCUSSION AND CONCLUSION

This study evaluates the structural nexus between green finance and eco-sustainability using a multidimensional statistical pipeline. While initial spatial clustering effectively reveals severe regional heterogeneity and developmental polarization across the provinces, the Canonical Correlation Analysis successfully extracts the overarching, whole-sample mechanisms driving the ecological transition.

The overall structural findings indicate that eco-sustainability is driven by two simultaneous mechanisms. First, strict capital inhibition against high-energy-consuming industries creates necessary market friction that contracts polluting activities. Second, targeted resource allocation positively catalyzes industrial emission reductions. Based on these global findings, macro-policy design should equally prioritize restrictive financial constraints on traditional obsolete sectors and active capitalization of emerging ecological projects.

Despite these findings, this study has specific limitations. The clustering results outline spatial distribution but do not provide empirical evidence for localized, intra-cluster mechanisms. The current model relies on an aggregate evaluation and cannot verify sub-regional causalities. Therefore, future research must build upon these baseline descriptive clusters by deploying localized econometric models or correlation tests to extract specific regional mechanisms. Additionally, transitioning from static cross-sectional data to longitudinal panel datasets is necessary to map these evolutionary trajectories over time.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Li B, Wang X, Khurshid A, et al. Environmental governance, green finance, and mitigation technologies: pathways to carbon neutrality in European industrial economies. International Journal of Environmental Science and Technology, 2025, 22(15): 14899-14912.
[2] Lv C, Bian B, Lee C, et al. Regional gap and the trend of green finance development in China. Energy Economics, 2021, 102: 105476.
[3] Wang Z, Wang F, Ma S. Research on the coupled and coordinated relationship between ecological environment and economic development in China and its evolution in time and space. Polish Journal of Environmental Studies, 2025, 34(3): 3333-3342.
[4] Huang H, Gao H, Han J. A macroeconomic data fusion method based on machine learning. Statistical Research, 2022, 39(5): 134-145.
[5] Chen B, Wang F. Macroeconomic indicators, technical indicators and treasury bond futures price forecasting: an empirical test based on random forest machine learning. Statistics & Information Forum, 2019, 34(6): 29-35.
[6] Luo J, Zhuo W, Liu S, et al. The optimization of carbon emission prediction in low carbon energy economy under big data. IEEE Access, 2024, 12: 14690-14702.
[7] Meng T, Yao C. Statistical evaluation of the coupled and coordinated development of digital finance and high-quality real economy. Statistics & Decision, 2024(8).
[8] Zhang H, Deng H. Photovoltaic evaluation in buildings integrally based on entropy weight and grey relational analysis. 2025 8th International Conference on Power and Energy Applications (ICPEA). 2025: 526-531.
[9] Ji Q, Sun Y, Hu Y, et al. A survey of deep clustering algorithms. Journal of Beijing University of Technology, 2021, 47(8): 912-924.
[10] Wei Y, Wang B, Ma L. Statistical evaluation of China's digital economy development level. Statistics & Decision, 2024(10).