Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Data-Driven Governance Framework for the Cultivated Land Dilemma: Balancing Intensification and Fragmentation in the Yangtze River Delta, China

Yuelong Su
Huazhong Agricultural University
Lanjiao Wen
wenlanjiao@mail.hzau.edu.cn

Huazhong Agricultural University
Xiong Zhang
South Central University for Nationalities
Anlu Zhang
Huazhong Agricultural University

Article

Keywords: Cultivated Land Use Intensification, Cultivated Land Fragmentation, Interpretable Machine Learning, Cultivated Land Governance, Yangtze River Delta

Posted Date: March 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8166653/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

A Data-Driven Governance Framework for the Cultivated Land Dilemma: Balancing Intensification and Fragmentation in the Yangtze River Delta, China

Abstract: Cultivated land in rapidly urbanizing regions supports food production and ecological integrity, yet policies often present intensification and fragmentation as a zero-sum choice. This study examines 27 cities in the Yangtze River Delta from 2000 to 2023 and links a robust evaluation of intensification and fragmentation with a coordination assessment and a driver analysis based on spatial detection and interpretable machine learning. Findings show that intensification first rose and then declined, with higher levels at the periphery and lower levels in the interior. Levels are higher in southern Anhui and Zhejiang and lower in northern Jiangsu and northern Zhejiang. Fragmentation followed a fluctuating path with repeated rises and falls and shows a clear gradient that is higher in the south and lower in the north, with peaks in southern Jiangsu, Zhejiang, and Anhui. The two dimensions are moving into early coordination, but pronounced spatial disparities remain. Coordination is generally higher in the west, imbalance persists in the northeast, the center is relatively stable, and the south shows a structure with lower values at the core and higher values at the periphery. Industrial structure and road network density are the core drivers of the interaction between intensification and fragmentation, and urbanization and population density reinforce their effects. Building on these results, a multilevel governance architecture and a four-quadrant policy toolkit match regional interaction types with targeted interventions such as service-based consolidation, institutional innovations, and infrastructure upgrades. This data informed governance framework offers a transferable pathway for managing cultivated land systems under rapid urbanization while advancing equitable and sustainable land governance.
Keywords: Cultivated Land Use Intensification; Cultivated Land Fragmentation; Interpretable Machine Learning; Cultivated Land Governance; Yangtze River Delta.

1. Introduction

Cultivated land systems are at the heart of a critical global sustainability nexus, tasked with ensuring food security while safeguarding terrestrial ecosystems. However, governing these systems is often trapped in a persistent dilemma. On one hand, production-driven intensification aims to maximize yields through greater inputs(Hultgren et al., 2025). On the other, conservation-oriented consolidation seeks to enhance ecological integrity and achieve economies of scale by reducing fragmentation(von Jeetze et al., 2023). Conventional policy often treats these goals as a simple trade-off, forcing a false choice between productivity and sustainability. This paper challenges that dichotomy. It argues that intensification and fragmentation are not merely antagonists but two interacting facets of the same land system. Their relationship can be conflicting or synergistic, contingent on the specific socioeconomic context and governance approach.

This governance dilemma is acutely manifested in China's Yangtze River Delta (YRD), a global exemplar of rapid industrialization and urbanization. The region presents a unique natural experiment where top-down policies promoting high-intensity agriculture collide with the bottom-up realities of smallholder-dominated land tenure and relentless urban encroachment(Zhang et al., 2024). This collision has produced a complex and fragmented landscape mosaic, making the YRD an ideal case for studying the outcomes of conflicting land use pressures. Understanding this dynamic is not only vital for China's food security and rural revitalization but also offers transferable lessons for other global economic hubs navigating similar developmental pressures, from Southeast Asia to Latin America(Looga et al., 2018; Johansen et al., 2020; Oldekop et al., 2020; Prabhakar, 2021).

In recent years, both CLUI and CLF have attracted considerable scholarly attention. CLUI refers to the process of optimising resource allocation to maximise output per unit area of arable land under limited resource inputs (Liu et al., 2024). CLF refers to the fragmentation of arable land due to division or uneven distribution, leading to reduced farming efficiency and irrational land use(Ye et al., 2024). Despite growing scholarly attention to CLUI and CLF, most existing studies examine them in isolation. A large body of research explores intensification and its effects on productivity or ecology (Chen et al., 2013; Hou et al., 2016; Teffera et al., 2018; Wang et al., 2021), while another strand investigates fragmentation and its impacts on farming efficiency or rural development (Yu et al., 2021; Lyu et al., 2022; Mahanty et al., 2023). Yet very few studies analyze the interplay between intensification and fragmentation as two sides of the same system. This gap persists because conventional linear models are ill-equipped to capture the non-linear, feedback-driven dynamics inherent in the CLUI-CLF system, leading to incomplete or even

misleading policy recommendations. Even when coordination has been studied, most efforts stop at quantitative indices, lacking deeper investigation into driving mechanisms or governance strategies.

This study addresses these gaps by developing an integrated, governance-oriented framework. It moves beyond isolated metrics to achieve three primary objectives. First, it systematically evaluates the spatiotemporal dynamics of both intensification and fragmentation in the YRD using a robust, nonlinear modelling approach. Second, it uncovers the complex interplay and core socioeconomic drivers governing this relationship through a hybrid analysis combining spatial statistics and interpretable machine learning. Finally, and most importantly, it translates these empirical findings into a multilevel governance architecture, featuring a strategic framework and a tactical toolkit with differentiated interventions. This research makes three key contributions. Theoretically, it validates a context-dependent, interactive model of land use change, challenging the conventional trade-off narrative. Methodologically, it pioneers an integrated analytical pipeline that enhances the credibility of driver analysis for complex socio-ecological systems. Empirically, it provides a data-driven, actionable governance framework for managing cultivated land systems in rapidly urbanizing regions, offering insights transferable beyond China.

The research framework of this study is shown in Figure 1. The remainder of this study is organized as follows. Section 2 introduces the study area, data sources and methods. Section 3 presents the results and analysis. Section 4 discusses the theoretical and practical implications of the findings, culminating in the construction of a multilevel governance architecture based on the empirical evidence. The last section concludes this work.

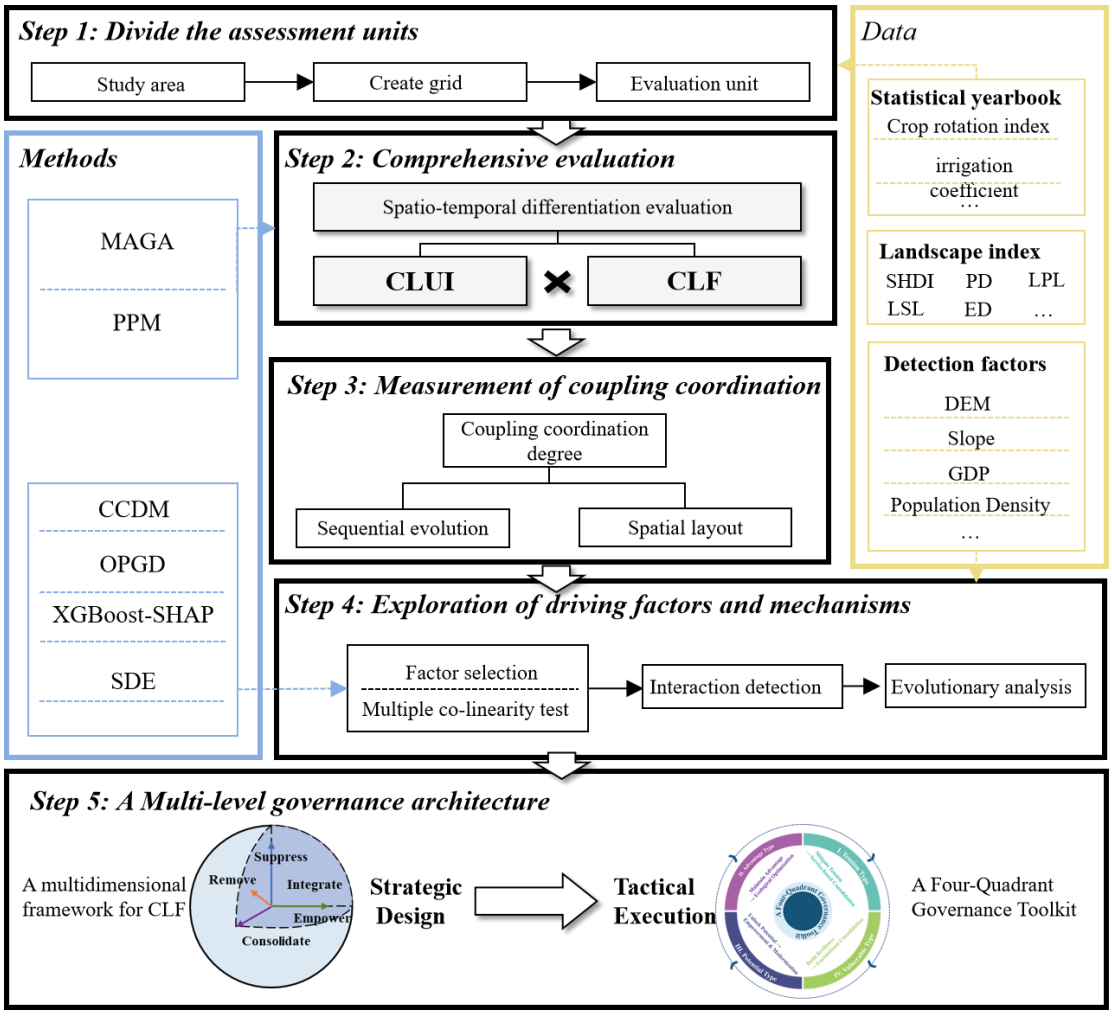


Figure 1. Research framework

2. Materials and methods

2.1 Study area

YRD is located in eastern China and includes 27 cities (Figure 2). The terrain ranges from flat to gently undulating and consists mainly of the alluvial plains of the Yangtze River. This humid subtropical region has four distinct seasons, with average annual temperatures ranging from 15°C to 18°C and precipitation ranging from 1,000 mm to 1,500 mm, making it agriculturally productive (Wang et al., 2020). The network of rivers, lakes and waterways, including the Yangtze River and Taihu Lake, is well connected and

plays a role in irrigation, fisheries and transportation. Economically, the YRD is an economic powerhouse, with its diversified economy of manufacturing, finance, technology and agriculture contributing significantly to China's GDP (Hou et al., 2021). With more than 200 million inhabitants, the region faces challenges posed by rapid urbanization and high population densities, leading to intensification of agricultural production and land fragmentation (Yang et al., 2020).

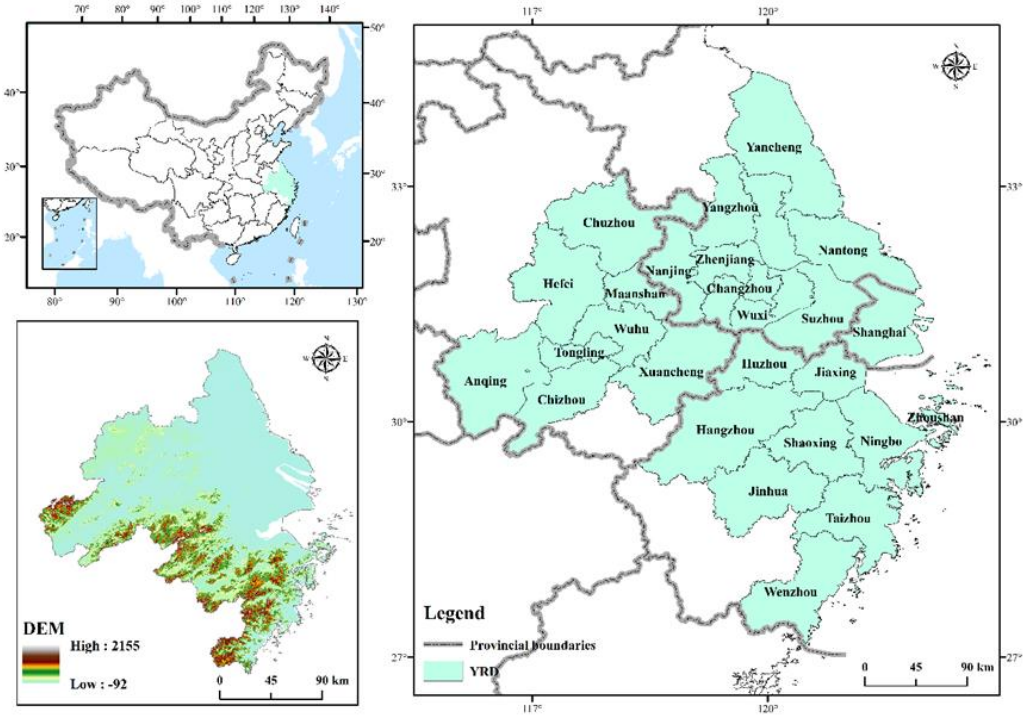


Figure 2. Study area

2.2 Data sources

In this study, the period of 2000-2023 was selected as the study period to carry out the evaluation and coupling relationship exploration between CLUI and CLF. The data of this study are as follows (Table 1).

Table 1. Information and Sources of Basic Data in YRD

Data name	Data message	Data sources
Rainfall	Daily rainfall data for 9 national meteorological stations,	National Data Center for Meteorological Sciences (http://data.cma.cn/site/index.html)
DEM	30m resolution	Geospatial Data Cloud (http://www.gscloud.cn/)
Land use	30m resolution	Resources and Environmental Sciences, Chinese Academy of Sciences (RESDC)(http://www.resdc.cn)
Remote-sensing data	Terra MODIS NDVI 250 m	Center for Resource and Environmental Sciences and Data, Chinese Academy of Sciences (http://www.resdc.cn/)
Soil	Soil type data 1km resolution	Center for Resource and Environmental Sciences and Data, Chinese Academy of Sciences (http://www.resdc.cn/)
Statistical data	Road network density, industrial structure, urbanization rate, population density, GDP per capita, etc.	YRD Statistical Yearbooks

2.3 Evaluation indicator system

2.3.1 Construction of evaluation index system of CLUI

According to the actual situation of the utilization of cultivated land in the YRD (Xu et al., 2013; Han and Chen, 2018; Ma et al., 2020; Zhu et al., 2022), this paper employs frequency analysis and a systematic review of the literature to identify four key dimensions: land input, land utilization, output efficiency, and sustainability. These 12 indicators were selected to create a holistic assessment covering inputs (e.g., labor, machinery), utilization processes (e.g., cropping intensity), economic outputs (e.g., land value), and long-term sustainability (e.g., per capita land area), reflecting the multifaceted nature of intensification (Table 2).

Table 2. Indicator System for Evaluating the CLUI in YRD

Target level	Criteria level	Criteria level	Criteria level Indicator	Direction
CLUI	Input intensity	Per capita labor input	Agricultural workers/total cultivated area (people/km^2)	+
		Gross per capita power of agricultural machinery	Total power of agricultural machinery/cultivated land (kW/km^2)	+
		Fertilizer input per plot	Agricultural fertilizer use/cultivated land area (t/km^2)	+
	Utilization level	Replanting index	Total sown area of crops/total cultivated area (%)	+
		Irrigation coefficient	Effective irrigated area/total cultivated area (%)	+
		Reclamation coefficient	Total cultivated area/regional land area (%)	+
	Output benefits	Average land value	Agricultural output value/total cultivated area (yuan/km^2)	+
		Production value per laborer	Agricultural output value/agricultural employees (yuan/person)	+
	Sustainability	Grain yield	Total grain output / sown area (kg/km^2)	+
		Cultivated land area per capita	Total area of cultivated land/total population (km^2/person)	+
		Non-agricultural index	Non-agricultural population/total population (%)	-
		Agricultural Support Index	Gross Output Value of Crop Farming / Population	+

2.3.2 Construction of the evaluation index system for CLF

Due to the influence of both natural and human factors, the selection of indicators for evaluating CLF should comprehensively reflect its spatiotemporal differentiation characteristics (Yu et al., 2018). In this study, eight landscape metrics commonly used in landscape ecology—such as patch density (PD), mean patch size (MPS), largest patch index (LPI), and landscape shape index (LSI)—were employed to analyze the spatial and temporal dynamics of CLF in the YRD. These indices, calculated using Fragstats 4.2 software, provide both quantitative and qualitative insights tailored to the region's specific land use conditions and research objectives. Among the 8 indicators, the Largest Patch Index (LPI), Mean Patch Size (MPS), and Aggregation Index (AI) are typical negative indicators, while the remaining 5 are positive indicators. These were analyzed at both municipal and provincial scales to comprehensively assess the spatiotemporal patterns of CLF across the study area.

Patch Density Index (PD) is used to calculate the number of cropland patches per unit area. A higher PD value indicates a greater number of patches and, consequently, a higher degree of cultivated land fragmentation (CLF).

$$PD = \frac{n}{A} \tag{1}$$

where A presents the total area of the landscape (hm^2), and n denotes the number of panels (n).

Largest Patch Index (LPI) refers to the proportion of the largest individual patch relative to the total landscape area, reflecting the spatial dominance and control capacity of the dominant patch type.

$$LPI = \frac{\sum_{i=1}^{n} \max(a_{ij})}{A} \tag{2}$$

Mean Patch Size (MPS) refers to the degree of CLF, where smaller MPS values more finely fragmented patches, while larger values suggest lower levels of fragmentation.

$$MPS = \frac{A}{n} \tag{3}$$

Patch shape index (LSI) measures the complexity of patch shapes and their interaction with surrounding patches. A higher LSI value indicates more complex and irregular patch shapes, while a lower value reflects simpler, more regular shapes.

$$LSI = \frac{0.25E}{\sqrt{A}} \tag{4}$$

where E is the perimeter of the plaque edge.

Area-weighted shape index (AWMSI) is the sum of the perimeter-to-area ratio of each patch multiplied by the respective area

weights. When AWMSI = 1, it indicates that the shape of the patch is the simplest square; when the value of AWMSI increases, it indicates that the shape of the patch has become more irregular and complex.

$$AWMSI = \sum_{i=1}^{n}\left[\left(\frac{0.25P_i}{\sqrt{a_i}}\right)\left(\frac{a_i}{A}\right)\right] \tag{5}$$

Edge Density Index (ED) often presents the degree of division of the edge of arable land, using the ratio of the perimeter of the edge of arable land patches to the area of the patches, and the greater the value of ED, the higher the degree of division of the boundaries of the arable land, which represents the degree of human interference.

$$ED = \frac{E}{A} \tag{6}$$

Degree of separation of arable land types (DIVISION) refers to the degree of separation of the distribution of different patches. When there is only one patch in the landscape, its value is 0. The larger its value is, the more dispersed the cultivated land patches are, the more complex the landscape is, and the higher the degree of fragmentation is. Its range is: $0 \leq \text{DIVISION} < 1$.

$$DIVISION = 1 - \sum_{j=1}^{n}\left(\frac{a_{ij}}{A}\right)^2 \tag{7}$$

Aggregation Index (AI) reflects the degree of clustering among cultivated land patches. A higher AI value indicates that cultivated land is composed of larger, more aggregated patches, implying lower fragmentation. Conversely, a lower AI value suggests that the land consists of many smaller, scattered patches, indicating a higher degree of fragmentation. The index ranges from 0 to 100.

$$AI = 100\left(1 + \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{P_{ij}\ln P_{ij}}{2\ln n}\right) \tag{8}$$

2.4 Comprehensive evaluation model

2.4.1 Projection pursuit modeling (PPM)

To comprehensively evaluate the complex, non-linear, and high-dimensional nature of CLUI and CLF, this study adopts an integrated Multi-Agent Genetic Algorithm-based Projection Pursuit Model (MAGA-PPM). This hybrid approach is chosen for its capacity to robustly capture underlying data structures that traditional linear methods often fail to identify (Loperfido, 2018; Su et al., 2024).

The model is structured in two main stages. First, the Projection Pursuit Model (PPM) is employed to formulate the evaluation task as a formal optimization problem. Its primary function is to define an objective function that seeks the most informative one-dimensional projection of the high-dimensional indicator data. An ideal projection is one that maximizes both the overall spread of the projected values and their local clustering, thereby best revealing the underlying data structure. However, the resulting objective function is highly complex and non-linear, making it prone to numerous local optima that can trap conventional optimization algorithms.

This is where the MAGA demonstrates its significant advantages. As an advanced intelligent optimization algorithm, MAGA is exceptionally well-suited to navigate this complex search space and identify the global optimum. Unlike simpler optimizers, MAGA operates on a population of potential solutions and refines them through evolutionary processes. Its distinctive multi-agent framework, which simulates collaboration and competition among individuals, enhances both its global exploration and local exploitation capabilities. This dual mechanism not only accelerates convergence but also greatly increases the probability of finding the true optimal solution while avoiding premature stagnation. Furthermore, as validated in our model tests, MAGA exhibits superior stability and robustness, making it less sensitive to initial parameter settings and ensuring the reliability of the results. Ultimately, MAGA delivers the optimal projection direction, which is then used to calculate the final, robust comprehensive evaluation scores for both CLUI and CLF. The detailed computational framework, including all specific formulas and symbol notations, is provided in Figure 3.

The MAGA-PPM was implemented using MATLAB R2022a. Based on preliminary testing and established practices, key parameters for the algorithm were set as follows: the grid size was 30, initial maximum iterations were 350, competition probability was 0.25, crossover probability was 0.2, and mutation probability was 0.1. The density window width was set to rmax/5. During execution, it was observed that sample projection values stabilized after 50 iterations, indicating effective convergence; therefore, the final maximum number of iterations was set to 50 for all runs.

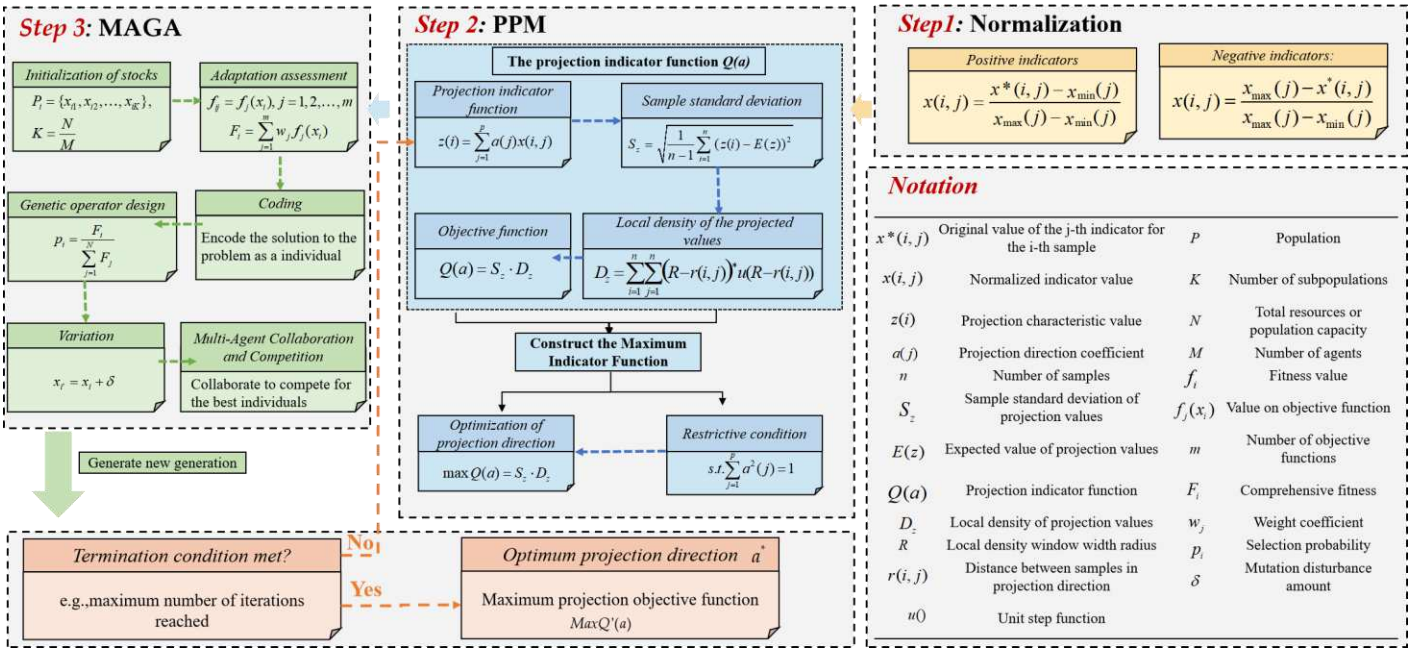


Figure 3. MAGA-PPM computational framework, specific formulas, and symbol explanations

2.5 Coupling coordination degree model (CCDM)

The CCDM is an analytical tool used to assess the level of development of interactions and coordination between two or more systems (Li et al., 2022). This model reveals the interdependencies and synergies between systems by quantifying the coupling coordination degree (CCD) between them.

(1) Coupling degree (CD)

The CD indicates the interdependence between different elements within the system. The relevant formula is:

$$C = \frac{C_1 \cdot C_2}{\sqrt{C_1^2 + C_2^2}} \tag{9}$$

where C_1 and C_2 represent the combined evaluation value of the two elements in the system, respectively.

(2) CCD

Let CCD be denoted as H; the relationship can be expressed by the following formula:

$$H = \frac{C}{C_{max}} \cdot 100\% \tag{10}$$

where C_{max} denotes the maximum coupling coefficient of the system under ideal conditions, typically attaining a value of unity.

According to previous studies (Li et al., 2012; Zuo et al., 2021), the classification criteria of CCD are shown in Table 3. The coupling and coordination degrees are combined to obtain a comprehensive CCD metric:

$$D = C \cdot H \tag{11}$$

Table 3. Classification of CCD

Interval of CCD	Level	Status of CCD
[0.0~0.1)	1	Extreme disorder
[0.1~0.2)	2	Severe disorder
[0.2~0.3)	3	Moderate disorder
[0.3~0.4)	4	Mildly disorder
[0.4~0.5)	5	Nearly disorder
[0.5~0.6)	6	Barely coordinated
[0.6~0.7)	7	Elementary coordination
[0.7~0.8)	8	Intermediate coordination
[0.8~0.9)	9	Good coordination

Interval of CCD	Level	Status of CCD
[0.9~1.0]	10	Quality coordination

2.6 Optimal parameters-based geographical detector

Geodetector is a statistical method grounded in spatial differentiation analysis (Wang et al., 2022). It offers distinct advantages by eliminating the need for a priori data distribution assumptions and enabling quantification of nonlinear interactions among multiple factors. These capabilities make it particularly suitable for uncovering the driving mechanisms underlying complex geographical phenomena(Zhao et al., 2020). However, the traditional Geodetector require manual classification of variable categories, which suffers from strong subjectivity and poor discretization effects. The Optimal Parameters-based Geographical Detector (OPGD) can overcome these shortcomings by exploring the optimal combination of spatial scale, spatial data discretization methods, and the number of spatial data partitions to achieve more precise spatial consistency analysis, thereby making the results more impartial and objective The core equations are defined as follows:

$$q = 1 - \frac{\sum_{h=1}^{L} N_h \sigma_h^2}{N\sigma^2} \tag{12}$$

In the equation, q represents the explanatory degree of the driving factor, with a value range of [0,1]. L denotes the total number of strata; N_h and N represent the sample size of the h th stratum and the total sample size, respectively; σ_h^2 and σ^2 are the variances of the h th stratum sample and the total sample, respectively.

2.7 XGBoost-SHAP

(1) XGBoost

To further accurately reveal the complex driving mechanisms of CCD affecting cultivated land use in the YRD, this study incorporates the XGBoost model. XGBoost is a powerful machine learning algorithm based on the gradient boosting framework, specifically designed to enhance predictive accuracy(Liu et al., 2025). It improves model performance by incrementally building a series of weak learners, offering advantages in handling large datasets and excelling in controlling model complexity while avoiding overfitting. Furthermore, XGBoost mitigates overfitting through mechanisms such as controlling tree complexity and incorporating regularization terms. Its objective function comprises both a loss function and a regularization term. The loss function quantifies the discrepancy between predicted and actual values, with smaller values indicating better model performance, while the regularization term constrains model complexity, thereby preventing overfitting(Zhou et al., 2025).

(2) SHAP

The XGBoost model alone cannot intuitively explain the specific direction and intensity of the impact of various driving factors on the coupling coordination degree. To address this, this study introduces the SHAP method to interpret the black-box nature of XGBoost. The SHAP method draws from the Shapley value in game theory to evaluate the relative importance of individual features in machine learning models(Tao et al., 2025). Its core lies in calculating the marginal contribution of each feature across different combinations and deriving the final impact value through weighted averaging. This method is applicable not only for global interpretation but also for local interpretation, making it a crucial tool in model interpretability research. The calculation formula is as follows(Fu et al. 2025):

$$\varphi_i = \sum_{S \subseteq H,\ \{i\}} \frac{|S|!(|H|-|S|-1)!}{|H|!} \left(t\left(S \cup \{i\}\right) - t\left(S\right) \right) \tag{13}$$

In the formula, φ_i is the SHAP value of driving factor i, H is the overall set of all features, S is a subset of H that does not include feature i, S is the feature subset without factor i, and $t(S)$ is the model prediction result corresponding to the feature combination in subset S.

2.8 Spatiotemporal trend analysis methods

To quantify and visualize the evolutionary trends of geographical phenomena over space and time, this study employed Standard Deviational Ellipse (SDE) and centre of gravity analysis. These methods are widely used in geography and economics to reveal the directional trends, dispersion, and core locational shifts of spatial elements over a period.

The SDE summarizes the spatial distribution characteristics of a set of points through several key parameters. The center of the ellipse represents the mean centre or centre of gravity of the distribution. The long and short axes indicate the primary and secondary directions of the data's distribution, reflecting its directionality. The azimuth describes the angle of rotation of the long axis from due North, revealing the main directional trend of expansion. The area of the ellipse reflects the degree of data dispersion; a larger area signifies a more dispersed distribution, while a smaller area indicates greater concentration.

The Centre of Gravity Trajectory Analysis visualizes the overall migration direction and path of the phenomenon by tracking the spatial shifts in the SDE's centre of gravity over different years. By connecting the gravity centres chronologically, the resulting trajectory line clearly illustrates the dynamic evolution of the core area of spatial agglomeration.

In this study, SDE and centre of gravity analysis were applied to the CCD and the CLUI-CLF interaction types to uncover their overall evolutionary patterns and directional trends across the YRD. All analyses were performed using ArcGIS 10.4.1.

3. Results

3.1 Model validation and robustness check

Prior to the main analysis, the credibility and robustness of the analytical framework were rigorously established. The selection of the MAGA-PPM was central to this framework, chosen for its capacity to navigate the complex, nonlinear, and high-dimensional nature of land use data. Since the optimization algorithms seek a minimum while the PPM model seeks to maximize fitness, the objective function was negated for the analysis.

A two-tiered validation process confirmed the model's superiority. First, an internal consistency check using a virtual indicator method demonstrated the model's ability to effectively identify the true global optimum. Second, and more critically, MAGA-PPM was benchmarked against four other widely-used optimization algorithms. The results, summarized in Figure 4, show that MAGA consistently outperformed its counterparts in terms of convergence speed, accuracy, and stability across multiple independent runs. This rigorous validation provides strong confidence in the model's capacity to produce reliable and accurate assessments, forming a solid foundation for the subsequent analysis. Detailed validation data are provided in the Supplementary Information (Tables A1-A4).

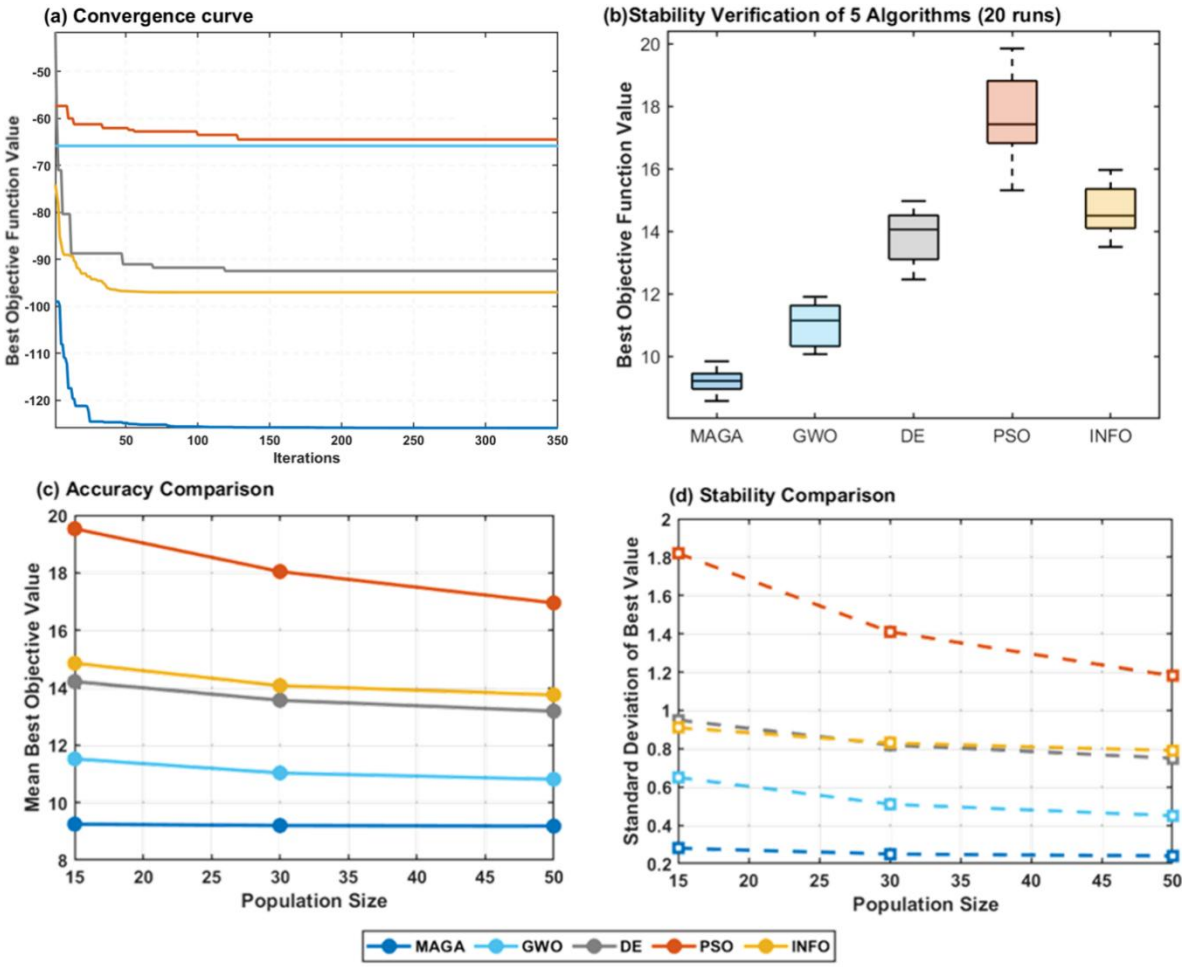


Figure 4. Comprehensive performance verification of five optimization algorithms

3.2 Characterization of the spatiotemporal evolution of the CLUI of YRD

The analysis of CLUI reveals a distinct temporal trend and spatial pattern across the YRD. Temporally, the overall intensification level for the entire region followed an inverted U-shaped curve, increasing from 2000 to a peak in 2005, and subsequently declining steadily until 2023 (Figure 5a). This overarching trend, however, masks considerable local variations. For instance, while most cities mirrored the regional decline post-2005, Shanghai experienced a renewed upward trend after 2015, and Zhoushan showed almost continuous growth throughout the period.

Using ArcGIS 10.4, the spatiotemporal patterns of CLUI in the YRD were visualized (Figure 5b) and classified into five categories based on CLUI values: V (very low intensification), IV (low intensification), III (moderate intensification), II (high intensification), and I (very high intensification). Spatially, the YRD exhibits a clear "periphery-high, interior-low" pattern of intensification (Figure 5b). The highest levels of CLUI are consistently found in the southern and western peripheries, particularly in southern Zhejiang and most of Anhui province, where many cities reached high or very high intensification levels by 2023. In stark contrast, the economically developed core regions, including southern Jiangsu and northern Zhejiang, generally remained at low or very low intensification levels. This spatial divergence became more pronounced over time, with Anhui showing significant improvement while the core areas stagnated, highlighting a growing regional disparity in agricultural land use intensity.

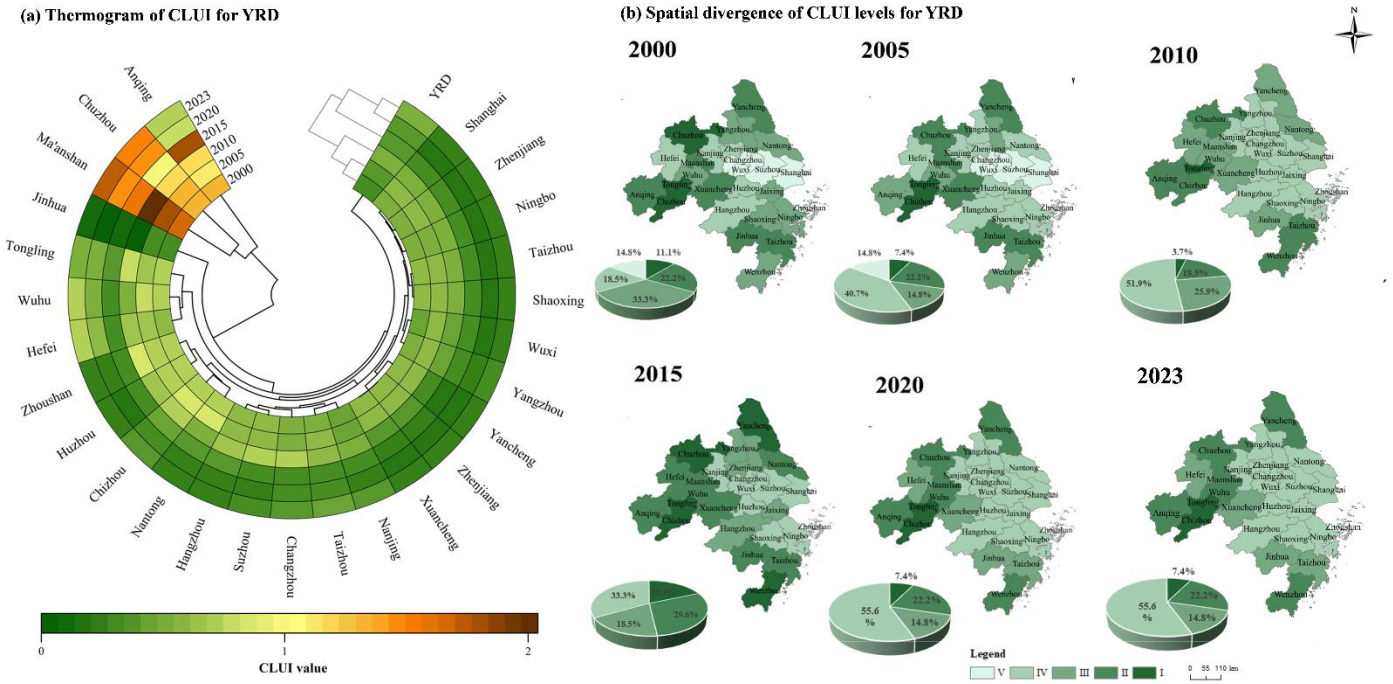


Figure 5. Spatiotemporal differentiation of CLUI in the YRD

3.3 Characterization of the spatiotemporal evolution of the CLF of YRDs

The evolution of CLF in the YRD presents a fluctuating temporal pattern and a stark north-south spatial divide. Temporally, the regional average fragmentation level exhibited a "W-shaped" fluctuation over the study period, with periods of decline followed by subsequent increases (Figure 6a). At the local level, trends were diverse: cities in southern Jiangsu showed a consistent increase in fragmentation after 2010, while some cities in Zhejiang, such as Wenzhou and Jinhua, experienced a general decline.

Spatially, a persistent "south-high, north-low" gradient characterized the distribution of fragmentation (Figure 6b). The highest levels of fragmentation were concentrated in the southern parts of the YRD, with cities in southern Zhejiang (e.g., Wenzhou, Taizhou) and southern Jiangsu reaching "very high" or "extreme" fragmentation. Conversely, the northern cities, particularly Yancheng and Nantong in northern Jiangsu, consistently maintained low fragmentation levels. This distinct spatial pattern suggests that factors related to topography and economic development intensity in the southern regions contribute significantly to a more fragmented landscape.

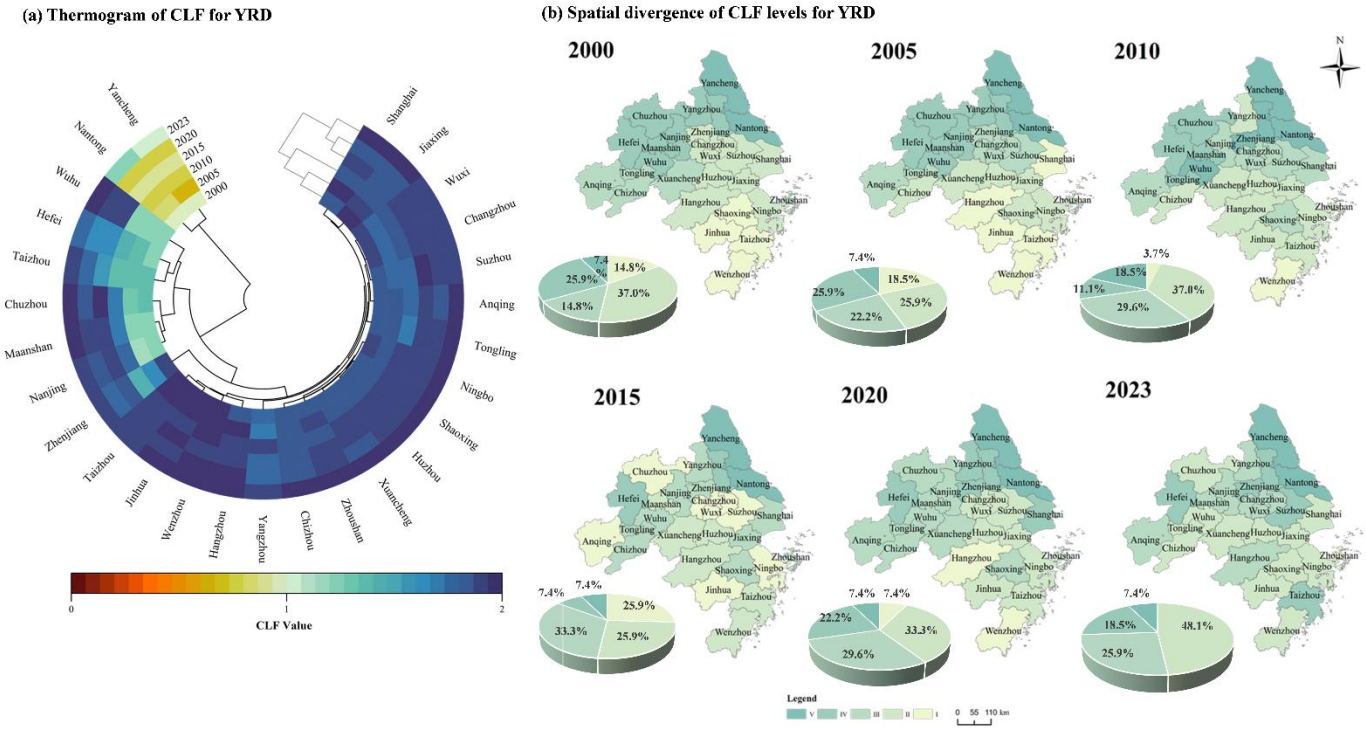


Figure 6. Spatiotemporal differentiation of CLF in the YRD

3.4 Characterization of the spatiotemporal evolution of the CCD of CLUI and CLF

3.4.1 Timing evolution patterns of CCD for CLUI and CLF

The analysis of the Coupling Coordination Degree (CCD) between intensification and fragmentation reveals a general trend of improvement across the YRD (Figure 7a), but with significant and persistent spatial disparities. Overall, the regional CCD level steadily increased, transitioning from a state of mild disorder to elementary coordination. However, this regional average conceals starkly different local trajectories.

Spatially, the distribution of CCD evolved into a complex pattern characterized by four key features (Figure 7c). First, a clear "western high" pattern emerged, with cities in Anhui province consistently showing higher levels of coordination. Second, a "northeastern trapping effect" was observed, where cities like Yancheng and Nantong remained in a state of serious imbalance throughout the 23-year period. Third, central areas of the YRD demonstrated relative equilibrium, gradually improving to a coordinated state. Fourth, the southern region exhibited a "core-periphery imbalance", with cities closer to the economic core showing lower coordination than those on the periphery. The spatiotemporal trend analysis further illustrates this dynamic. The center of gravity of the CCD remained relatively stable near Ningbo, while the standard deviational ellipse (SDE) showed a slight counter-clockwise rotation (Figure 7b). This indicates that the improvements in coordination were most rapid in the northwestern (Anhui) and southeastern (southern Zhejiang) parts of the region, reinforcing the pattern of growing spatial differentiation.

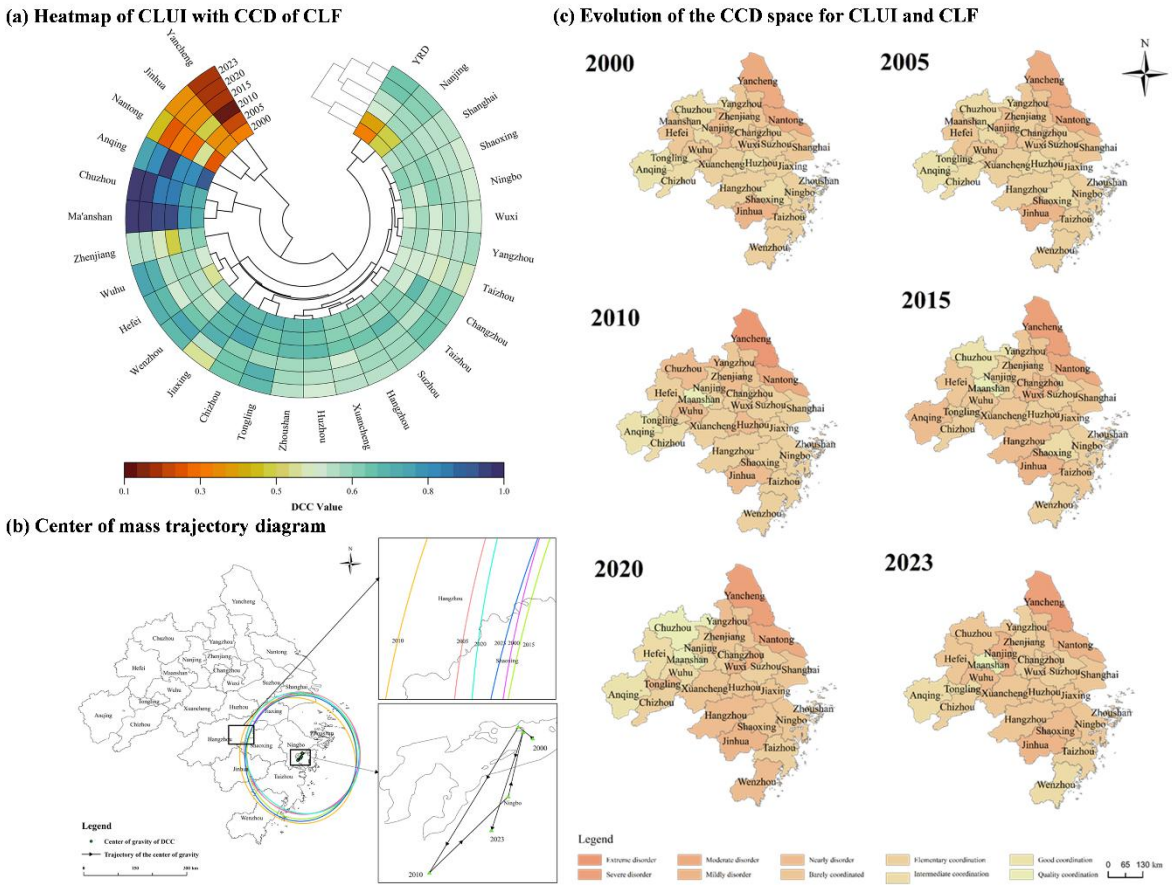


Figure 7. Spatiotemporal differentiation of CCD between CLUI and CLF in the YRD

3.5 Spatiotemporal dynamics of CLUI-CLF interaction types

To translate the interplay between CLUI and CLF into a diagnostic tool for governance, this study classified regions into four interaction types: Advantage (high CLUI, low CLF), Tension (high CLUI, high CLF), Potential (low CLUI, low CLF), and Vulnerable (high CLF, low CLUI).

The spatiotemporal evolution of these types reveals a dramatic regional restructuring over the study period (Figure 8b). While the YRD was initially dominated by Vulnerable and Tension types, a significant transition occurred by 2010. From that point onwards, the spatial pattern largely stabilized into three distinct blocs. The first bloc, encompassing most of Anhui Province, became a model region dominated by the Advantage type, exemplifying effective and intensive land use. The second, primarily covering northern Jiangsu, was characterized by the Potential type, suggesting its contiguous land has yet to achieve high intensification. The third bloc, which includes southern Jiangsu and most of Zhejiang, remained persistently locked in Tension and Vulnerable types, highlighting a sustained conflict between high-intensity urbanization and cultivated land management.

The center of gravity analysis (Figure 8a) confirms this spatial consolidation. A clear migration of the center of gravity from southwest to northeast, coupled with a marked contraction of the SDE area, indicates that these three distinct regional blocs have become more spatially concentrated and their characteristics more pronounced over time.

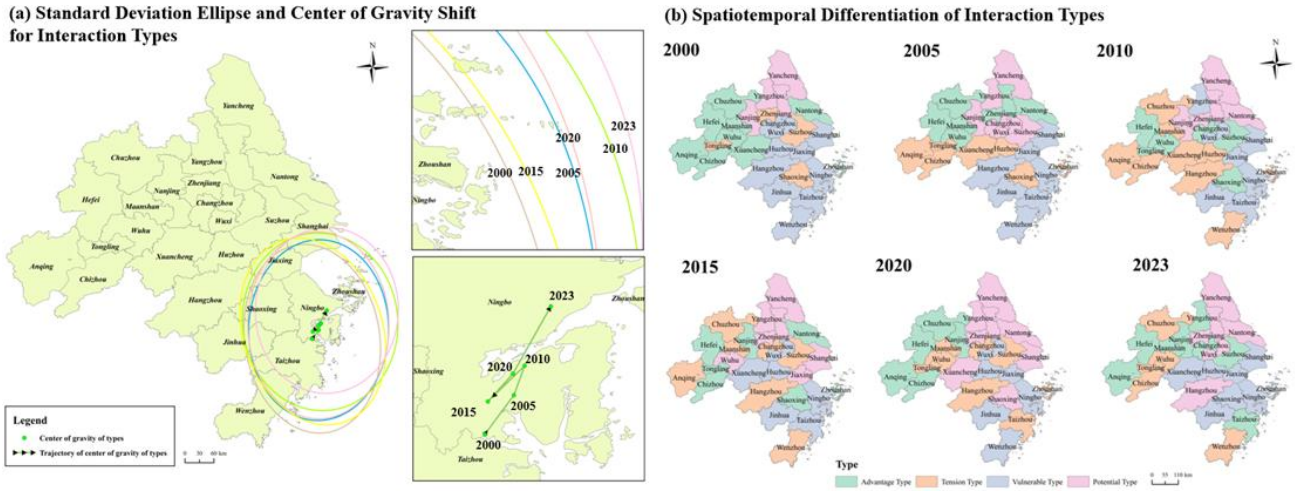


Figure 8. Spatiotemporal Dynamics of CLUI-CLF Interaction Types in the YRD

3.6 Investigation of the CCD drive mechanism

3.6.1 Factor selection

Based on previous research(Fonderflick et al., 2020; Yang et al., 2020), this study selected eight key factors encompassing natural and socio-economic dimensions as the primary drivers influencing the changes in the coordination between CLUI and CLF in the Yangtze River Delta region. Unlike the earlier assessments conducted at the municipal level, the driver analysis was performed at the county scale to obtain more granular spatial insights. The study area included counties and districts within the YRD, and due to missing data for certain years or regions, the final valid sample size for driver analysis was 1,290. Table 8 summarizes the descriptive statistics of these variables.

Table 8. Drivers for CCD and their descriptive statistics

Category	Indicator	Number	Observations	Minimum	Maximum	Mean	Standard deviation
Natural factors	DEM	D1	1290	4.832	646.271	81.686	123.317
	Slope	D2	1290	1.089	22.725	6.645	5.750
	NDVI	D3	1290	0.165	0.891	0.647	0.134
Socioeconomic factors	Road Network Density	D4	1290	0.605	2.449	1.400	0.389
	Industrial Value-Added as Percentage of GDP	D5	1290	0.206	0.626	0.416	0.086
	Population Density	D6	1290	5.220	7.747	6.469	0.509
	GDP Per Capita	D7	1290	0.210	41.420	7.218	6.796
	Urbanization Rate	D8	1290	0.158	0.893	0.619	0.178

3.6.2 Factor detection

All eight driving factors passed the multicollinearity test. The OPGD results reveal significant temporal variations in the explanatory power of various natural and socio-economic factors influencing the CCD between CLUI and CLF in the YRD (Figure 9). Overall, socio-economic factors exhibit stronger explanatory power than natural factors. In particular, the proportion of industrial added value (D5), road density (D4), urbanization rate (D8), and population density (D6) exert notably prominent influences on the CCD. These factors consistently maintained high values across different years, indicating that rapid industrialization and urbanization are core drivers shaping interregional cropland use patterns and the evolution of coordination levels.

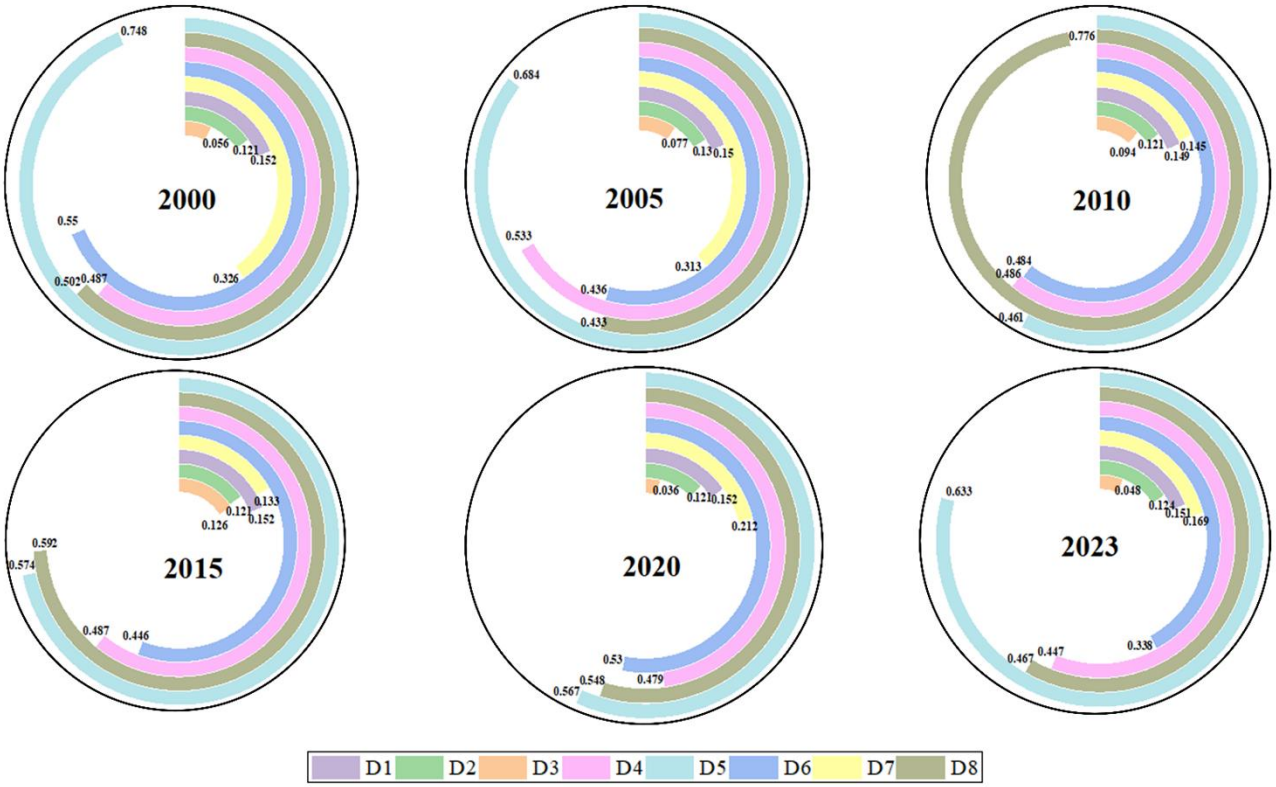


Figure 9. Single-factor detection under the OPGD method

Notably, after 2005, the explanatory power of urbanization rate (D8) and road network density (D4) increased significantly,

reflecting the role of infrastructure development and population agglomeration in promoting land use transitions. In contrast, natural factors consistently showed lower explanatory power with relatively stable fluctuations, suggesting that topographic conditions impose stable constraints on the CLUI–CLF coupling relationship in the YRD. The explanatory power of NDVI (D3) has risen steadily since 2010, implying that vegetation coverage is becoming increasingly important in influencing cropland coordination as regional ecological restoration and cropland improvement proceed.

Building upon the OPGD analysis, this study further incorporates the XGBoost-SHAP model from machine learning methods to provide supplementary validation of the importance of influencing factors (Figure 10). XGBoost-SHAP employs tree model-based feature contribution decomposition to intuitively quantify the marginal impact of each variable on the outcome. Its advantage lies in not only offering comprehensive explanations but also revealing the intensity of individual variables' effects under different scenarios. The analysis demonstrates high consistency between the key influencing factors identified by both XGBoost-SHAP and OPGD methods. For instance, the proportion of industrial added value in X5, X6, and X8 consistently rank highly in importance across different years, with their rankings largely aligned. This finding indicates that the critical drivers identified by OPGD exhibit strong robustness and reliability, while the XGBoost-SHAP results further validate their scientific validity. Meanwhile, slight discrepancies exist in the importance weights of a few factors between the two models, though the overall trends remain consistent, reflecting the multidimensional and complex nature of the influencing mechanisms. In conclusion, the integration of XGBoost-SHAP effectively enhances the rigor of model comparison and result verification, making the conclusions of the influencing factor analysis more persuasive.

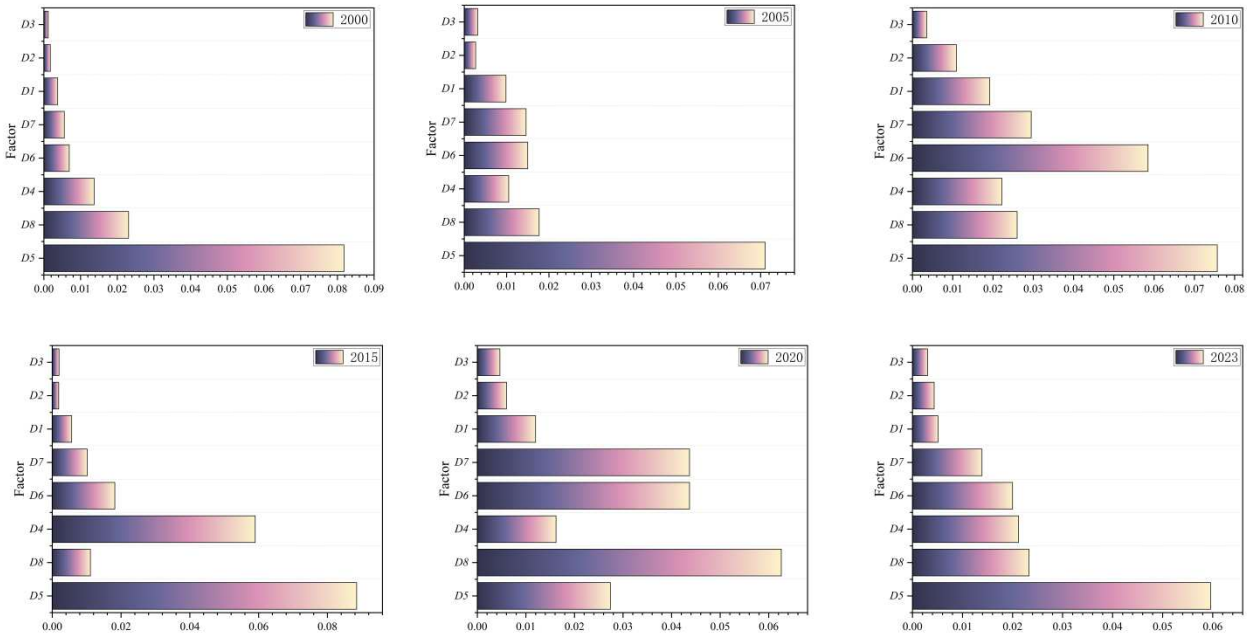


Figure 10. Single-factor detection using the XGBoost-SHAP method

Overall, the formation of CCD in the YRD reflects a shift from natural constraints to socio-economic drivers, with ecological factors gradually gaining importance. This offers valuable insights for optimizing regional land resources and promoting sustainable utilization. The interaction detection results indicate that the pairwise factor combinations significantly enhance the explanatory power of CCD for CLUI and CLF, predominantly through dual-factor enhancement and nonlinear enhancement (Figure 11).

Nearly all factor combinations exhibited higher *q*-values than any single factor. In the early stage, transportation and industrialization demonstrated the strongest synergy, with road network density and industrial added value proportion reaching a q-value of 0.976 in 2000, while population density and GDP per capita achieved 0.936. Between 2005 and 2010, the enhancement effect intensified, with population density and GDP per capita reaching 0.942 in 2005, and industrialization and GDP per capita attaining 0.948 in 2010. Combinations of industrialization with population density and road network density both exceeded 0.92. By 2015, the interaction between transportation and industrialization remained dominant, with road network and industrialization around 0.967, industrialization and GDP per capita at 0.932, and industrialization and population density at 0.922. A new peak appeared in 2020, when urbanization and population density reached 0.976, road network and GDP per capita approximately 0.967, and road network and population density 0.971. By 2023, the synergy between socioeconomic factors and transportation had further strengthened, with roads and industrialization approaching unity and industrialization with GDP per capita around 0.990. Natural and socioeconomic factors also exhibited notable synergistic effects, with terrain–road network and terrain–urbanization

combinations typically exceeding 0.70. In 2020, DEM and urbanization reached approximately 0.85, while slope–road network interactions ranged between 0.47 and 0.70. NDVI combined with road networks or industrialization showed significant increases, rising from lower values to above 0.70 in multiple years and peaking near 0.97. Overall, the interactions among industrialization, urbanization, population, and transportation networks constitute the core driving framework of CCD in the cultivated land system of the Yangtze River Delta, while natural factors primarily play an auxiliary, amplifying role.

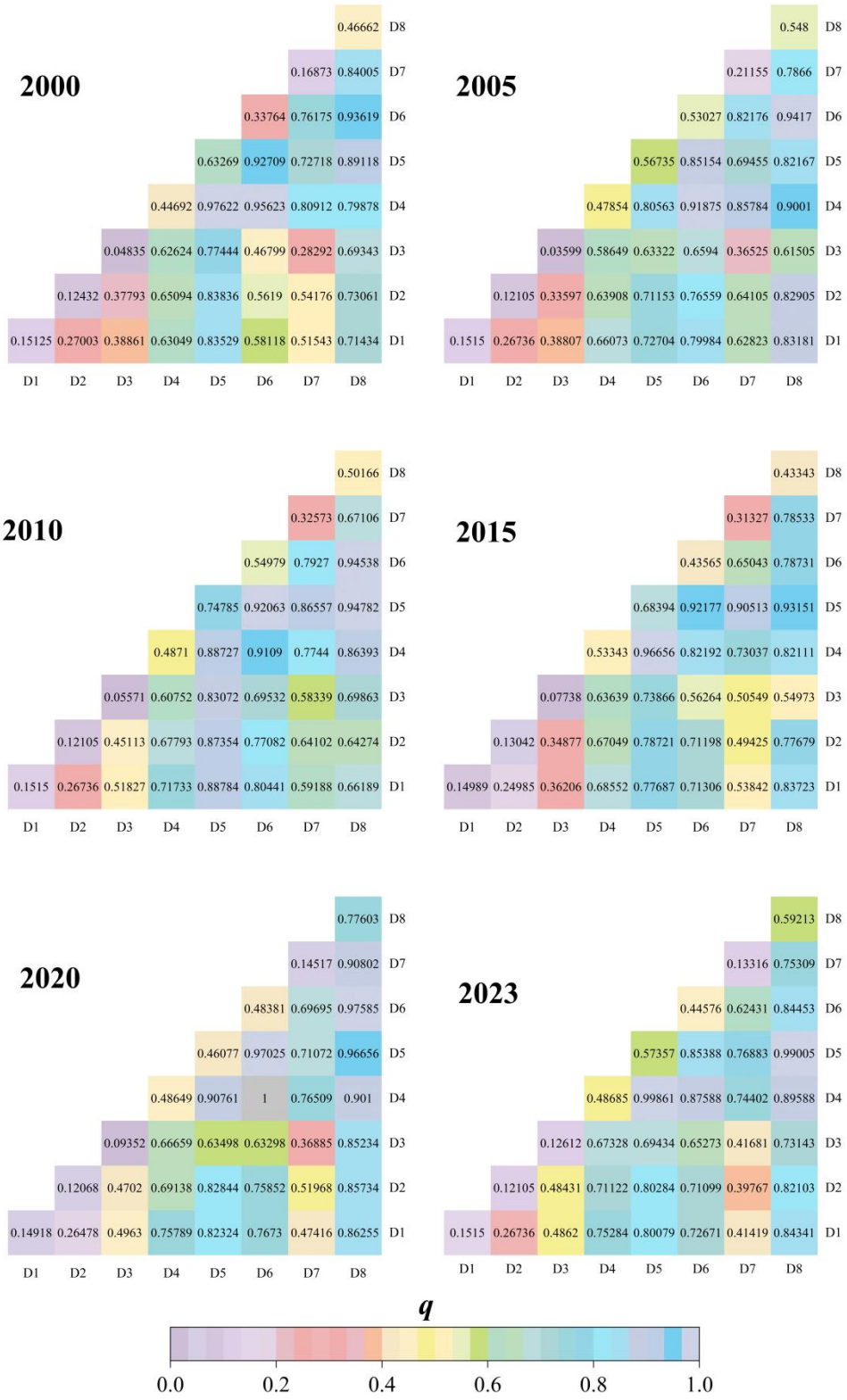


Figure 11. Driver interaction diagram

4. Discussion

4.1 Divergent paths of intensification and fragmentation

The empirical findings from the Yangtze River Delta provide compelling evidence that the relationship between cultivated land intensification and fragmentation is not a simple linear antagonism but a complex, context-dependent dynamic. The analysis reveals

a nuanced reality where the outcome is contingent on the stage and mode of regional development. This is vividly illustrated by the contrasting trajectories within the YRD. In less-urbanized areas like Anhui province, state-led land consolidation policies, coupled with agricultural investment, successfully improved system coordination, leading to the emergence of the Advantage type. This demonstrates a scenario where policy intervention can indeed foster a synergistic relationship.

In stark contrast, the hyper-industrialized core regions, such as southern Jiangsu and parts of Zhejiang, remained trapped in Tension or Vulnerable states. Here, the relentless pressure of urban encroachment and the persistence of smallholder-dominated land tenure created a downward spiral. Rapid urbanization led to the conversion of high-quality, contiguous farmland, leaving behind a more fragmented agricultural landscape. Simultaneously, the rising opportunity cost of farming and labor migration to urban sectors disincentivized on-farm investment, leading to a decline in intensification. This confirms that in certain socioeconomic contexts, fragmentation and de-intensification can become mutually reinforcing problems. The dominance of socioeconomic drivers, robustly identified by both OPGD and XGBoost-SHAP, explains why these regional disparities have solidified over time, challenging any one-size-fits-all policy solution.

4.2 Theoretical and methodological implications for land system science

This study offers significant contributions to land system science by challenging the conventional trade-off narrative. Theoretically, its findings move the discourse beyond a simplistic view of intensification versus fragmentation. By empirically validating a context-dependent, interactive model, the research demonstrates that their relationship can range from synergistic to conflicting. This highlights the critical role of socioeconomic context and governance in shaping land use transitions, enriching systems-thinking approaches to sustainable land management by emphasizing nonlinearity and feedback loops.

Methodologically, this paper pioneers an integrated analytical pipeline that enhances the credibility and depth of land use studies. The application of the MAGA-PPM demonstrates the necessity of robust, nonlinear models for accurately diagnosing complex socio-ecological systems. More importantly, the innovative dual-driver analysis, which synergizes spatial statistics (OPGD) with interpretable machine learning (XGBoost-SHAP), provides a powerful framework for cross-validation. This hybrid approach offers a transferable blueprint for researchers seeking to conduct nuanced, evidence-based assessments of complex land system challenges, thereby increasing the reliability of findings for policy-making.

4.3 A data-driven framework for strategic land governance

The complexity and heterogeneity uncovered by the analysis call for a paradigm shift in governance—from isolated, top-down interventions toward an integrated, adaptive, and spatially differentiated strategy. To this end, a data-driven framework for strategic land governance is proposed, designed to translate analytical insights into actionable principles. At its core is a multidimensional collaborative governance architecture, structured as a hexagon-pillar-cycle model (Figure 12). This represents a holistic blueprint for action.

The outer hexagon defines the core governance domains—Territorial Planning, Property Rights, Business Models, and Technical Support—all operating under the dual objectives of Farmland Protection and Integrated Management. At the heart of the framework lie three crucial pillars: a reformed Property Rights System, innovative Business Models, and advanced Technical Support. These pillars are specifically designed to address the predominantly socioeconomic drivers identified by the OPGD and XGBoost-SHAP analysis. The framework's dynamism is captured by two interconnected feedback cycles: a governance cycle, where territorial planning guides policy design to curb negative trends, and an efficiency cycle, where institutional and technical innovations empower stakeholders. This strategic design moves beyond simple physical consolidation to a systemic approach that orchestrates institutional, market, and technological forces toward sustainable land use.

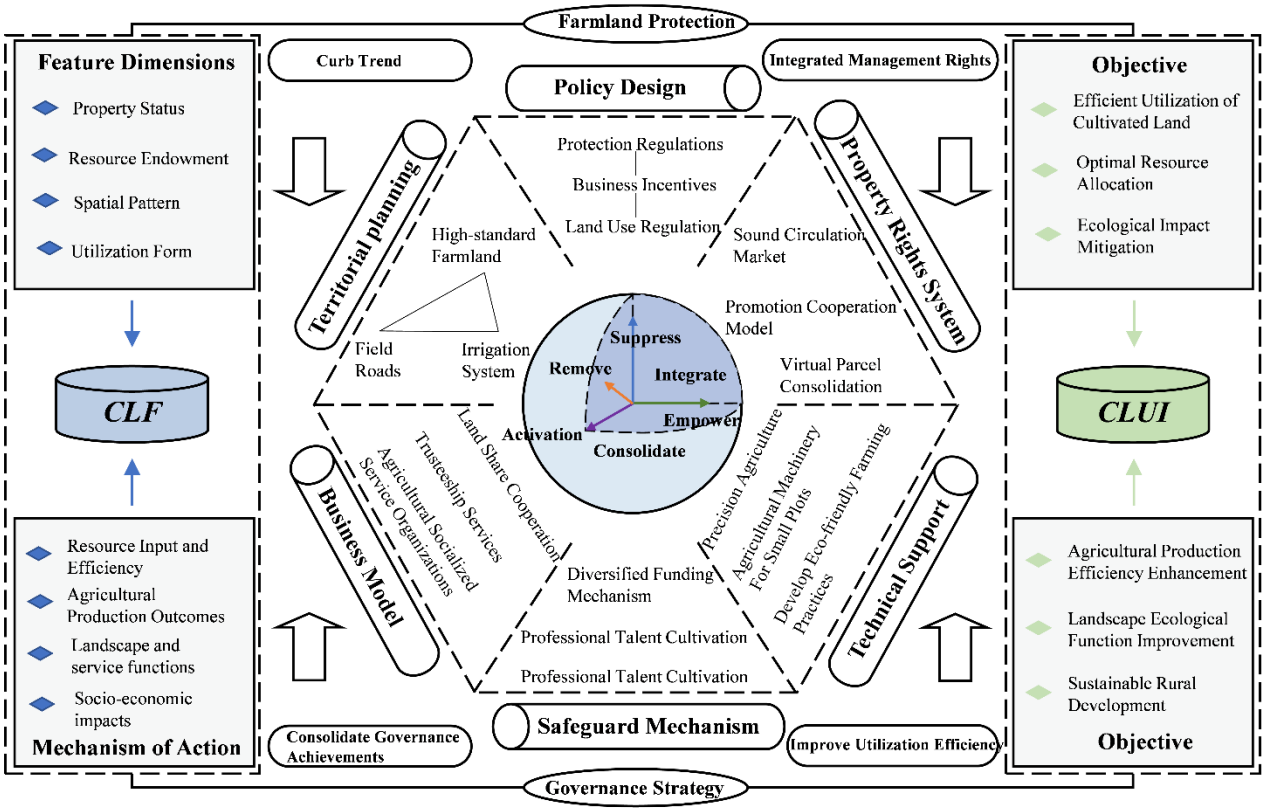


Figure 12. A multidimensional collaborative governance framework for CLF

4.4 Operationalizing governance with a policy toolkit

While the strategic framework provides the guiding philosophy, a crucial gap often exists between broad strategy and concrete intervention. To bridge this know-do gap, a tactical layer is essential: the Four-Quadrant Governance Toolkit (Figure 13). This toolkit serves as a practical decision-support system, directly linking the regional diagnosis (the four interaction types) to targeted policy prescriptions. It operationalizes the principle of spatially differentiated governance by providing a clear, evidence-based rationale for selecting the right tool for the right problem.

For example, in Tension type regions (e.g., southern Jiangsu), characterized by high intensification but high fragmentation, the primary conflict is operational efficiency. Here, the toolkit recommends service-based consolidation through institutional innovations like virtual plot integration via cooperatives and the promotion of agricultural machinery sharing platforms, explicitly avoiding disruptive physical land parcelling. Conversely, for Vulnerable type areas (e.g., Anhui in the early 2000s), where both intensification and contiguity are low, the focus shifts to foundational consolidation. The prescribed tools include investing in high-standard farmland construction, clarifying land tenure to incentivize long-term investment, and providing basic technical training. For Potential type regions (e.g., northern Jiangsu), the toolkit prioritizes intensification-boosting measures, such as introducing high-yield varieties and precision agriculture technologies. Finally, for the ideal Advantage type regions, the focus is on sustainability enhancement, promoting ecological farming practices to ensure long-term resilience. This toolkit transforms complex analysis into an adaptable instrument for evidence-based policy design.

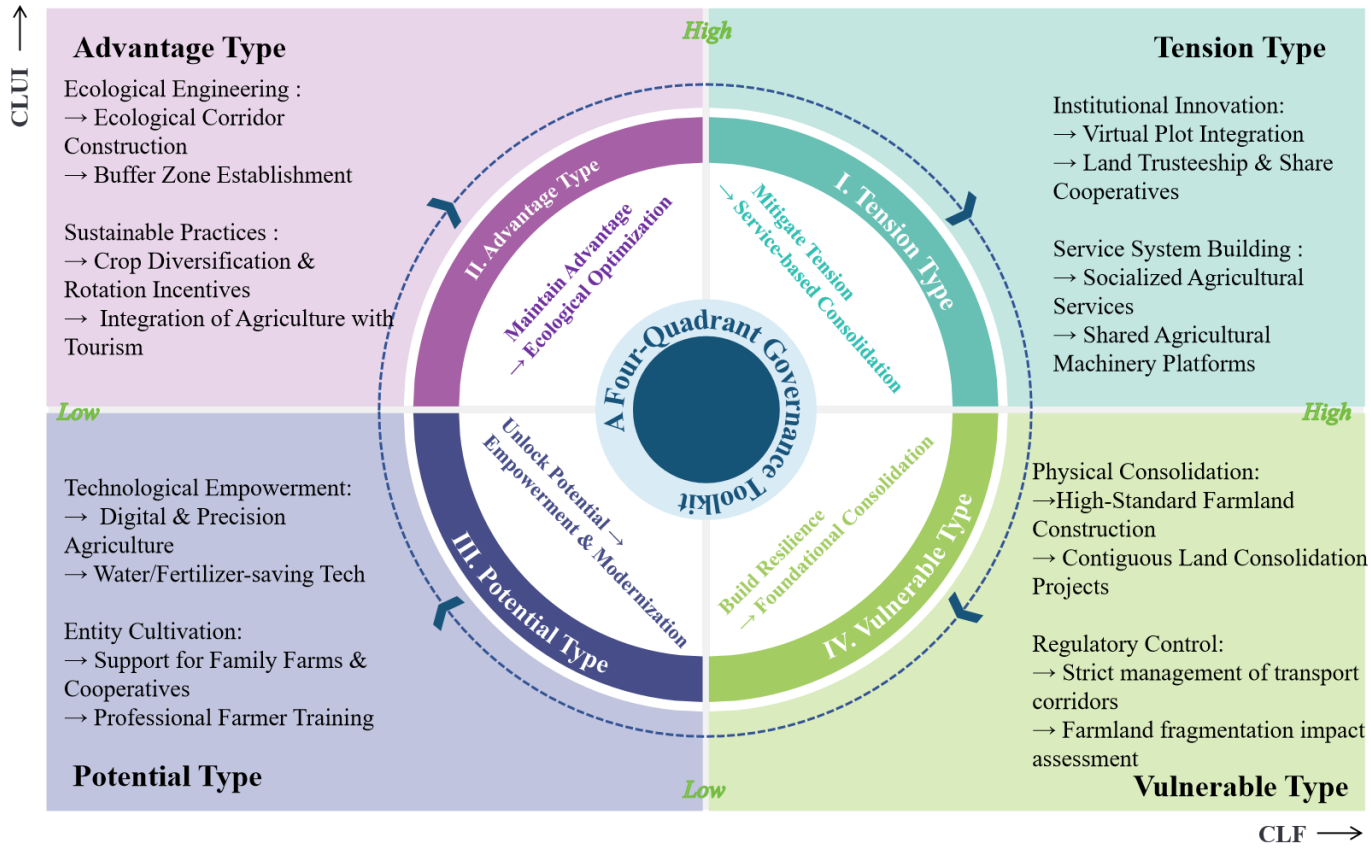


Figure 13. A Four-Quadrant Governance Toolkit for Differentiated Policy Intervention

5. Conclusions

This study establishes a comprehensive evaluation framework for CLUI and CLF in the YRD, introducing the MAGA-PPM method for integrated spatiotemporal analysis and benchmarking its performance against seven alternative algorithms. The CCD model and the OPGD are employed to examine the coupling relationship and driving mechanisms. Additionally, qualitative methods are used to explore the dynamic evolution of these mechanisms in depth. Four key findings emerge.

First, CLUI shows an overall fluctuating trend—initially increasing and then decreasing—forming a spatial pattern characterized by higher values in peripheral areas and lower values in central regions. CLUI is notably higher in Anhui and southern Zhejiang, and lower in northern Jiangsu and northern Zhejiang. Second, CLF exhibits a "W-shaped" temporal fluctuation, with a spatial distribution of higher fragmentation in southern areas and lower fragmentation in the north. Specifically, CLF is more pronounced in southern Jiangsu, southern Zhejiang, and southern Anhui. Third, the CCD between CLUI and CLF has generally improved, indicating progress toward initial coordination. However, spatial disparities remain: the western YRD shows higher CCD values, the northeastern region has long remained imbalanced, the central region maintains relative stability, and the southern region displays a pattern of lower CCD values in the interior but higher values along the periphery. Fourth, from the perspective of the formation mechanism, both the proportion of industrial added value in GDP and road network density have been identified as core influencing factors through the dual analytical methods of OPGD and XGBoost-SHAP.

Despite its contributions, the study has several limitations that point to directions for future research. First, the reliance on macro-level statistical data and remote sensing imagery limits representation of micro-level drivers, such as farmers' decision-making and grassroots policy enforcement. Incorporating survey-based microdata and indicators for policy intensity and farmer responses would enhance the framework. Second, the prefecture-level analysis may obscure finer spatial heterogeneity; future work should adopt higher-resolution data at the county or township scale. Third, the regional focus restricts generalizability. Comparative analyses with other major agricultural regions, such as the Pearl River Delta or the Chengdu–Chongqing Urban Agglomeration, could validate and broaden the applicability of these findings.

Declarations

Competing interests

The authors declare no competing interests.

Ethical approval

This study did not involve any human participants, human data, or animal experiments. Therefore, ethical approval was not required.

Informed consent

This study did not involve any human participants, human data, or animal experiments. Therefore, ethical approval was not required, and informed consent was not applicable.

Data availability

The raw data supporting the conclusions of this article will be made available by the corresponding author upon reasonable request.

Funding

This work was supported by the National Natural Science Foundation of China (No. 72273048), the Major Program of National Fund of Philosophy and Social Science of China (No. 18ZDA054), the National Social Science Foundation of China (No. 22FJYB041), the Ministry of education of Humanities and Social Science project (No. 22YJC790846, No. HZKY20220349), and the Fundamental Research Funds for the Central Universities of China (No. 2662025GGPY003).

References

Chen Y-D, Wang H-Y, Zhou J-M, Xing L, Zhu B-S, Zhao Y-C, Chen X-Q (2013) Minimum Data Set for Assessi ng Soil Quality in Farmland of Northeast China. *Pedosphere* 23(5):564–576. https://doi.org/10.1016/S1002-0 160(13)60050-8

Fonderflick J, Besnard A, Chardès M-C, Lanuzel L, Thill C, Pointereau P (2020) Impacts of agricultural intensific ation on arable plants in extensive mixed crop-livestock systems. *Agriculture, Ecosystems & Environment* 2 90:106778. https://doi.org/10.1016/j.agee.2019.106778

Fu X, Wang M, Zhang D, Chen F, Peng X, Wang L, Tan SK (2025) An XGBoost-SHAP framework for identifyi ng key drivers of urban flooding and developing targeted mitigation strategies. *Ecological Indicators* 175:1 13579. https://doi.org/10.1016/j.ecolind.2025.113579

Han M, Chen G (2018) Global arable land transfers embodied in Mainland China's foreign trade. *Land Use Polic y* 70:521–534. https://doi.org/10.1016/j.landusepol.2017.07.022

Hou K, Li X, Wang J jing, Zhang J (2016) An analysis of the impact on land use and ecological vulnerability o f the policy of returning farmland to forest in Yan'an, China. *Environmental Science and Pollution Resear ch* 23(5):4670–4680. https://doi.org/10.1007/s11356-015-5679-9

Hou Y, Zhang K, Zhu Y, Liu W (2021) Spatial and temporal differentiation and influencing factors of environme ntal governance performance in the Yangtze River Delta, China. *Science of the Total Environment* 801:149 699. https://doi.org/10.1016/j.scitotenv.2021.149699

Hultgren A, Carleton T, Delgado M, Gergel DR, Greenstone M, Houser T, Hsiang S, Jina A, Kopp RE, Malevich SB, McCusker KE, Mayer T, Nath I, Rising J, Rode A, Yuan J (2025) Impacts of climate change on gl obal agriculture accounting for adaptation. *Nature* 642(8068):644–652. https://doi.org/10.1038/s41586-025-09 085-w

Johansen PH, Ejrnæs R, Kronvang B, Olsen JV, Præstholm S, Schou JS, Johansen SK (2020) Exploring the inter disciplinary potential of the Agenda2030—Interactions between five Danish societal demands for sustainabl e land use. *Land Use Policy* 94:104501. https://doi.org/10.1016/j.landusepol.2020.104501

Li L, Fan Z, Feng W, Yuxin C, Keyu Q (2022) Coupling coordination degree spatial analysis and driving factor between socio-economic and eco-environment in northern China. *Ecological Indicators* 135:108555. https:// doi.org/10.1016/j.ecolind.2022.108555

Li Y, Li Y, Zhou Y, Shi Y, Zhu X (2012) Investigation of a coupling model of coordination between urbanization and the environment. *Journal of Environmental Management* 98:127–133. https://doi.org/10.1016/j.jenvman. 2011.12.025

Liu J, Jin X, Lin J, Liang X, Zhang X, Zhou Y (2024) Identification and characteristic analysis of semi-natural h abitats in China's economically developed areas: New insights to inform cultivated land system ecological conservation. *Journal of Environmental Management* 351:119804. https://doi.org/10.1016/j.jenvman.2023.1198 04

Liu Y, Wang H, Guan X, Meng Y, Xu H (2025) Urban Flood Depth Prediction and Visualization Based on the

XGBoost-SHAP Model. *Water Resources Management* 39(3):1353–1375. https://doi.org/10.1007/s11269-024-04020-6

Looga J, Jürgenson E, Sikk K, Matveev E, Maasikamäe S (2018) Land fragmentation and other determinants of agricultural farm productivity: The case of Estonia. *Land Use Policy* 79:285–292. https://doi.org/10.1016/j.landusepol.2018.08.021

Loperfido N (2018) Skewness-based projection pursuit: A computational approach. *Computational Statistics & Data Analysis* 120:42–57. https://doi.org/10.1016/j.csda.2017.11.001

Lyu X, Peng W, Niu S, Qu Y, Xin Z (2022) Evaluation of sustainable intensification of cultivated land use according to farming households' livelihood types. *Ecological Indicators* 138:108848. https://doi.org/10.1016/j.ecolind.2022.108848

Ma L, Long H, Tu S, Zhang Y, Zheng Y (2020) Farmland transition in China and its policy implications. *Land Use Policy* 92:104470. https://doi.org/10.1016/j.landusepol.2020.104470

Mahanty B, Lhamo P, Sahoo NK (2023) Inconsistency of PCA-based water quality index – Does it reflect the quality? *Science of the Total Environment* 866:161353. https://doi.org/10.1016/j.scitotenv.2022.161353

Oldekop JA, Rasmussen LV, Agrawal A, Bebbington AJ, Meyfroidt P, Bengston DN, Blackman A, Brooks S, Davidson-Hunt I, Davies P, Dinsi SC, Fontana LB, Gumucio T, Kumar C, Kumar K, Moran D, Mwampamba TH, Nasi R, Nilsson M, Pinedo-Vasquez MA, Rhemtulla JM, Sutherland WJ, Watkins C, Wilson SJ (2020) Forest-linked livelihoods in a globalized world. *Nature Plants* 6(12):1400–1407. https://doi.org/10.1038/s41477-020-00814-9

Prabhakar SVRK (2021) A succinct review and analysis of drivers and impacts of agricultural land transformations in Asia. *Land Use Policy* 102:105238. https://doi.org/10.1016/j.landusepol.2020.105238

Su Y, Liu Y, Zhou Y, Liu J (2024) Research on the Coupling and Coordination of Land Ecological Security and High-Quality Agricultural Development in the Han River Basin. *Land* 13(10):1666. https://doi.org/10.3390/land13101666

Tao Y, Fan G, Liu F, Luo J, Zou R, Huang Y, Cao Y, Long J, Xu K-Q (2025) Precision prediction of trihalomethanes in drinking water and three-stage regulation mechanism: A study based on XGBoost-SHAP framework. *Journal Of Environmental Chemical Engineering* 13(6):119316. https://doi.org/10.1016/j.jece.2025.119316

Teffera ZL, Li J, Debsu TM, Menegesha BY (2018) Assessing land use and land cover dynamics using composites of spectral indices and principal component analysis: A case study in middle Awash subbasin, Ethiopia. *Applied Geography* 96:109–129. https://doi.org/10.1016/j.apgeog.2018.05.015

von Jeetze PJ, Weindl I, Johnson JA, Borrelli P, Panagos P, Molina Bacca EJ, Karstens K, Humpenöder F, Dietrich JP, Minoli S, Müller C, Lotze-Campen H, Popp A (2023) Projected landscape-scale repercussions of global action for climate and biodiversity protection. *Nature Communications* 14(1):2515. https://doi.org/10.1038/s41467-023-38043-1

Wang F, Zhang W, Nian X, Roberts AP, Zhao X, Shang Y, Ge C, Dong Y (2020) Magnetic evidence for Yellow River sediment in the late Holocene deposit of the Yangtze River Delta, China. *Marine Geology* 427:106274. https://doi.org/10.1016/j.margeo.2020.106274

Wang H, Qin F, Xu C, Li B, Guo L, Wang Z (2021) Evaluating the suitability of urban development land with a Geodetector. *Ecological Indicators* 123:107339. https://doi.org/10.1016/j.ecolind.2021.107339

Wang Z, Li X, Zhao H (2022) Identifying spatial influence of urban elements on road-deposited sediment and the associated phosphorus by coupling Geodetector and Bayesian Networks. *Journal of Environmental Management* 315:115170. https://doi.org/10.1016/j.jenvman.2022.115170

Xu Y, McNamara P, Wu Y, Dong Y (2013) An econometric analysis of changes in arable land utilization using multinomial logit model in Pinggu district, Beijing, China. *Journal of Environmental Management* 128:324–334. https://doi.org/10.1016/j.jenvman.2013.05.020

Yang C, Zeng W, Yang X (2020) Coupling coordination evaluation and sustainable development pattern of geo-ecological environment and urbanization in Chongqing municipality, China. *Sustainable Cities and Society* 61:102271. https://doi.org/10.1016/j.scs.2020.102271

Yang J, Wang H, Roberts DJ, Du H-N, Yu X-F, Zhu N-Z, Meng X-Z (2020) Persistence of antibiotic resistance

genes from river water to tap water in the Yangtze River Delta. *Science of the Total Environment* 742:140592. https://doi.org/10.1016/j.scitotenv.2020.140592

Ye S, Ren S, Song C, Du Z, Wang K, Du B, Cheng F, Zhu D (2024) Spatial pattern of cultivated land fragmentation in mainland China: Characteristics, dominant factors, and countermeasures. *Land Use Policy* 139:107070. https://doi.org/10.1016/j.landusepol.2024.107070

Yu Q, Hu Q, van Vliet J, Verburg PH, Wu W (2018) GlobeLand30 shows little cropland area loss but greater fragmentation in China. *International Journal of Applied Earth Observation and Geoinformation* 66:37–45. https://doi.org/10.1016/j.jag.2017.11.002

Yu Q, Xiang M, Sun Z, Wu W (2021) The complexity of measuring cropland use intensity: An empirical study. *Agricultural Systems* 192:103180. https://doi.org/10.1016/j.agsy.2021.103180

Zhang K, Jin Y, Li D, Wang S, Liu W (2024) Spatiotemporal variation and evolutionary analysis of the coupling coordination between urban social-economic development and ecological environments in the Yangtze River Delta cities. *Sustainable Cities and Society* 111:105561. https://doi.org/10.1016/j.scs.2024.105561

Zhao R, Zhan L, Yao M, Yang L (2020) A geographically weighted regression model augmented by Geodetector analysis and principal component analysis for the spatial distribution of PM2.5. *Sustainable Cities and Society* 56:102106. https://doi.org/10.1016/j.scs.2020.102106

Zhou X, Liu G, Wu Y, Wang G, Xu Y, Liu T (2025) XGBoost-SHAP based multi-scale evaluation and nonlinear response of river health in Guangdong Province. *Ecological Indicators* 178:114138. https://doi.org/10.1016/j.ecolind.2025.114138

Zhu X, Xiao G, Wang S (2022) Suitability evaluation of potential arable land in the Mediterranean region. *Journal of Environmental Management* 313:115011. https://doi.org/10.1016/j.jenvman.2022.115011

Zuo Z, Guo H, Cheng J, Li Y (2021) How to achieve new progress in ecological civilization construction? – Based on cloud model and coupling coordination degree model. *Ecological Indicators* 127:107789. https://doi.org/10.1016/j.ecolind.2021.107789

Appendix

A. Internal Validation Data for the MAGA-PPM Model

This section provides the detailed data used for the internal validation of the MAGA-PPM. The validation was performed for both the CLUI and CLF models, using the "virtual indicator' method as described in the main manuscript.

A1. Data for CLUI Evaluation Validation

Table A1. Validation indicators for CLUI evaluations

Year	Validation indicators							
	U_1	U_2	U_3	U_4	U_5	U'_1	U'_2	U'_3
2010	0.442	0.558	0.632	0.728	0.684	1	0.442	0.368
2011	0.452	0.549	0.631	0.722	0.664	1	0.451	0.369
2012	0.451	0.489	0.628	0.72	0.662	1	0.511	0.372
2013	0.412	0.475	0.624	0.688	0.649	1	0.525	0.376
2014	0.425	0.448	0.622	0.685	0.632	1	0.552	0.378
2015	0.429	0.465	0.587	0.675	0.612	1	0.535	0.413
2016	0.488	0.482	0.582	0.662	0.581	1	0.518	0.418
2017	0.426	0.381	0.584	0.651	0.582	1	0.619	0.416
2018	0.485	0.37	0.583	0.642	0.548	1	0.63	0.417
2019	0.535	0.354	0.582	0.632	0.551	1	0.646	0.418
2020	0.534	0.353	0.583	0.629	0.558	1	0.647	0.417
2021	0.539	0.352	0.565	0.617	0.601	1	0.648	0.435
2022	0.512	0.358	0.532	0.613	0.557	1	0.642	0.468
2023	0.511	0.355	0.536	0.611	0.552	1	0.645	0.464

576 **Table A2.** Optimal projection score vector for the validation metrics

Project	Optimal projection direction vectors							
	u_1	u_2	u_3	u_4	u_5	u'_1	u'_2	u'_3
Scenario 1	0.563	0.332	0.485	-0.428	0.629			
Scenario 2	0.591	0.325	0.471	-0.425	0.612	-0.0028		
Scenario 3	0.457	0.322	0.488	-0.438	0.577	0.0032	-0.319	0.477

577

578 **Table A3.** Validation indicators for CLF evaluation

Year	Validation indicators							
	J_1	J_2	J_3	J_4	J_5	J'_1	J'_2	J'_3
2010	0.588	0.458	0.596	0.328	0.643	1	0.542	0.404
2011	0.596	0.452	0.588	0.335	0.655	1	0.548	0.412
2012	0.642	0.455	0.583	0.339	0.642	1	0.545	0.417
2013	0.625	0.468	0.581	0.342	0.648	1	0.532	0.419
2014	0.629	0.465	0.578	0.348	0.649	1	0.535	0.422
2015	0.638	0.469	0.573	0.354	0.645	1	0.531	0.427
2016	0.629	0.488	0.57	0.359	0.632	1	0.512	0.43
2017	0.633	0.492	0.575	0.355	0.628	1	0.508	0.425
2018	0.618	0.512	0.588	0.358	0.633	1	0.488	0.412
2019	0.613	0.515	0.569	0.385	0.624	1	0.485	0.431
2020	0.605	0.508	0.566	0.411	0.626	1	0.492	0.434
2021	0.598	0.511	0.581	0.416	0.629	1	0.489	0.419
2022	0.583	0.529	0.583	0.422	0.624	1	0.471	0.417
2023	0.581	0.536	0.588	0.424	0.622	1	0.464	0.412

579

580 **Table A4.** Optimal projection direction vectors

Project	Optimal projection direction vectors							
	J_1	J_2	J_3	J_4	J_5	J'_1	J'_2	J'_3
Scenario 1	0.612	0.425	0.388	0.512	-0.387			
Scenario 2	0.578	0.423	0.395	0.495	-0.389	-0.0037		
Scenario 3	0.586	0.433	0.412	0.502	-0.374	0.0027	-0.428	0.417

581