

Article

Spatial Expansion and Driving Mechanisms of the Yangtze River Delta, Based on RF-RFECV Feature Selection and Night-Time Light Remote Sensing Data

Dandan Shao ¹, KyungJin Zoh ^{2,*} and Huiyuan Liu ³

¹ Environmental Planning Institute, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Republic of Korea; 2022-36592@snu.ac.kr

² Urban Landscape Planning and Design LAB (ULPD), Interdisciplinary Program in Landscape Architecture, Graduate School of Environmental Studies, Seoul National University, Seoul 08826, Republic of Korea

³ Graduate School of Environmental Studies, Seoul National University, Seoul 08826, Republic of Korea; liuhuiyuan@snu.ac.kr

* Correspondence: kjzoh@snu.ac.kr

Highlights

What are the main findings?

- A harmonized 2010–2022 CNLI time series derived from fused DMSP-OLS and NPP-VIIRS data reveals rising YRD urbanization with provincial convergence, while substantial city-level gaps persist.
- Built-up areas exhibit “core agglomeration–corridor guidance–intensification”: the built-up centroid remains near Southern Jiangsu, the dominant expansion axis shifts along the southeast coastal belt, and expansion decelerates as land-use efficiency improves.

What are the implications of the main findings?

- Urbanization governance should move from land-led outward growth to efficiency-oriented strategies, aligning transport corridors with station-area development and public-service provision to mitigate high-density externalities.
- Driver attribution (RF-RFECV + SHAP) indicates a transition toward productivity and green-innovation-led urbanization, implying that openness benefits depend on local absorptive capacity and institutional support for high-quality transformation.

Abstract

Rapid urbanization has promoted socioeconomic growth but has exacerbated spatial-structure imbalances. This study investigates 41 prefecture-level cities in the Yangtze River Delta (YRD) from 2010 to 2022. Using nighttime light data, we compute the Comprehensive Nighttime Light Index (CNLI) to track urbanization dynamics and delineate built-up areas. Furthermore, we apply random-forest recursive feature elimination with cross-validation (RF-RFECV) and a Shapley additive explanations (SHAP)-based interpretation framework to quantify the spatiotemporal evolution of urbanization drivers. The results indicate that urbanization in the YRD increased steadily overall during the study period. Shanghai maintained its core leadership, Jiangsu and Zhejiang advanced steadily, and Anhui rapidly caught up driven by regional integration policies. Although regional disparities generally converged, persistent absolute gaps in small and medium-sized cities and inland areas remain a prominent challenge to balanced development. Spatially, urbanization exhibits a gradient differentiation of “higher in the east and lower in the west, and higher along rivers and coasts than inland.” The regional spatial structure gradually shifted from an early



Academic Editor: Yuji Murayama

Received: 30 December 2025

Revised: 9 March 2026

Accepted: 14 March 2026

Published: 30 March 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

“pole-core-belt” pattern to a polycentric and networked urban agglomeration system, with metropolitan areas and economic belts serving as important carriers for promoting spatial balance. Furthermore, built-up areas exhibit a trajectory of “core agglomeration, corridor-oriented expansion, and intensive transition.” The shrinking coverage of the standard deviational ellipse and a slowdown in expansion rates suggest a shift from extensive outward sprawl to more concentrated development. Regarding driving mechanisms, YRD urbanization has evolved from early-stage factor-scale expansion to a later-stage efficiency- and innovation-driven trajectory. While population density remained the dominant driver, early-stage reliance on transport infrastructure and fiscal decentralization was largely replaced by the strengthening effects of per capita output and green innovation. Overall, these findings provide empirical evidence for optimizing spatial patterns and designing differentiated policies for high-quality urbanization in the YRD.

Keywords: Yangtze River Delta; urbanization; night-time lights; RF-RFECV; SHAP-ML; built-up area expansion; driving mechanisms

1. Introduction

Urbanization refers to the socio-economic transformation in which production and livelihood systems shift from predominantly agricultural activities to industrial and service-oriented urban forms. This process is typically accompanied by profound changes in lifestyles, production modes, and social structures [1]. In China, the urbanization rate had exceeded 60% by 2019. However, rapid urban expansion has produced a range of ecological and socio-spatial problems—loss of arable land, air and water pollution, urban heat islands, and serious intra-urban developmental imbalances [2]. Consequently, investigating the spatiotemporal evolution of urbanization is essential: it reveals regional differences in urbanization trajectories and spatial distributions, clarifies expansion patterns and interregional disparities, and informs more efficient land-use and resource allocation. Such analysis is crucial for promoting sustainable urbanization, supporting rural development, and advancing integrated urban–rural development [3]. Urban agglomerations are the main carriers of China’s new-type urbanization and key units for participating in international competition. The spatial expansion of urban agglomerations can trigger internal regional changes such as population migration, industrial specialization, and the agglomeration and diffusion of capital and other factors. Investigating the patterns of spatial expansion in urban agglomerations is therefore of profound significance for sound regional development and industrial upgrading. Within China’s urbanization process, the Yangtze River Delta (YRD) urban agglomeration holds particular empirical and theoretical significance. As one of the country’s most economically advanced and highly urbanized regions, the YRD—comprising 41 prefecture-level cities in Shanghai, Jiangsu, Zhejiang, and Anhui—is a central engine of China’s socio-economic development. The region is characterized by high economic output, dense industrial agglomerations, and concentrated population, while simultaneously reflecting prominent urban–rural disparities and environmental pressures associated with rapid urban growth [4]. Therefore, analyzing the spatiotemporal characteristics of urbanization in the YRD provides important theoretical insight into regional heterogeneity and evolution within China’s broader urbanization pattern. Such findings not only supply scientific evidence for regional development policy but also offer transferable lessons for other rapidly urbanizing regions.

A rigorous measurement of urbanization levels is a central topic in urban geography and regional economics. To date, a substantial body of research has examined major

and relatively mature urban agglomerations in China—such as the YRD, the Pearl River Delta, and the Beijing–Tianjin–Hebei region—across multiple spatial scales, providing an important foundation for understanding regional urbanization processes. However, existing studies either investigate early-stage dynamics at the provincial scale [5,6] or quantify several discrete time points for specific provinces at the county scale [7,8], which makes it difficult to systematically reveal continuous, region-wide evolutionary trajectories. This limitation stems partly from the lag of socio-economic statistics, inconsistencies in statistical definitions, and discontinuities in county- and city-level datasets, and partly from the constraints of conventional methods in capturing spatiotemporal dynamics and disentangling complex mechanisms.

To address these challenges, approaches to urbanization measurement have evolved from single statistical indicators (e.g., the share of urban population) [9], to multi-indicator composite evaluation systems [10], and further toward the integration of remote sensing and geospatial technologies [11]. Among these, nighttime light (NTL) data have become a key technical avenue for quantifying regional urbanization levels and monitoring urban spatial expansion, because they can objectively and continuously reflect the intensity of human activities and the extent of spatial development, thereby compensating for deficiencies in traditional data sources [12]. Notably, the Comprehensive Nighttime Light Index (CNLI) derived from such remote sensing data exhibits notable advantages over traditional statistical indicators in characterizing urbanization levels. First, its high objectivity and longitudinal consistency effectively circumvent data distortions caused by administrative division adjustments or manual statistical errors. Second, by transcending administrative boundaries, the CNLI provides pixel-level high spatial resolution, enabling the precise delineation of the spatial morphology and agglomeration characteristics of urbanization. Finally, it demonstrates a robust comprehensive representative capacity; the CNLI not only reflects the spatial concentration of population but also endogenously integrates socioeconomic attributes such as infrastructure development, economic vitality, and energy consumption [13]. A critical step in extracting built-up areas from NTL data is the determination of an appropriate light-intensity threshold. The main approaches include the empirical threshold method, change-point (abrupt-change) detection, the reference comparison method, and high-resolution imagery-based methods [14–17]. The empirical threshold method involves substantial subjectivity. Change-point detection often fails to account for heterogeneity in regional development processes; moreover, remote-sensing imagery covering large extents can be difficult to obtain and computationally demanding to process. High-resolution imagery-based methods that integrate multi-source remote-sensing and geographic data and apply supervised classification or machine-learning models can, to some extent, mitigate light spillover and improve spatial detail; however, they typically impose higher requirements on training samples, feature engineering, and inter-temporal consistency, and may introduce new challenges such as limited model transferability and insufficient interpretability. By contrast, the reference comparison method identifies an optimal threshold by comparing the built-up area extracted under candidate thresholds with officially reported built-up area statistics and selecting the threshold that minimizes the discrepancy. This approach is straightforward to implement, achieves relatively high accuracy, and has therefore been widely adopted. Therefore, given the need for an objective and feasible thresholding strategy for large-scale longitudinal analysis in the Yangtze River Delta (YRD), this study adopts the reference comparison method to determine NTL thresholds and extract urban built-up areas.

To address the issues of cross-sensor inconsistency and the lack of inter-annual comparability in long-term urbanization measurements, this study employs the cross-sensor calibrated annual DMSP-OLS-like nighttime light dataset for China (digital number, DN,

range: 0–63) published by [4]. From the annual time-series products, nighttime light images for the 2010–2022 study period were extracted. Based on raster data with a unified projection and spatial resolution, we further performed study area clipping, invalid pixel removal, and outlier control. An inter-annual continuity test was also conducted to mitigate the risks of noise and structural breaks. On this basis, the Comprehensive Nighttime Light Index (CNLI) was calculated using regional light intensity I and the spatial extent of lit areas S to achieve a spatially explicit measurement of urbanization intensity. Concurrently, using statistical yearbook panel data, an Urbanization Development Level (UDL) index comprising four dimensions and 21 indicators was constructed via the entropy method for consistency testing and cross-validation with the CNLI.

In exploring the driving mechanisms of urban spatial expansion, prior studies have primarily relied on multiple linear regression, geographically weighted regression, and spatial econometric models. However, the expansion of urban agglomerations is a complex and nonlinear process shaped by multidimensional drivers across economic, demographic, industrial, and ecological domains. When high-dimensional predictors are considered, conventional statistical models often suffer from multicollinearity and have limited ability to represent nonlinear interactions. To overcome this, methodological bottleneck, this study introduces an innovative hybrid machine learning approach to identify driving mechanisms: an analytical framework coupling random forest-based recursive feature elimination with cross-validation (RF-RFECV) and the Shapley additive explanations (SHAP) model. In the phase of feature selection and dimensionality reduction, the selection of RF-RFECV over other mainstream methods is justified by its distinct comparative advantages and applicability. First, relative to least absolute shrinkage and selection operator (LASSO), which relies on linear functional assumptions, the tree-based RF-RFECV can capture nonlinear relationships and interaction effects among candidate drivers. Second, compared with Boruta, which is designed to retain all relevant features and may, therefore, preserve redundant predictors, RF-RFECV aims to identify a more compact feature subset, facilitating parsimonious model construction. Third, unlike permutation importance used as a standalone post-hoc diagnostic, RF-RFECV performs iterative feature elimination under cross-validation, enabling data-driven selection of a feature set that yields robust predictive performance. Overall, this procedure helps reduce redundancy among multi-source indicators and improves model stability in the presence of correlated predictors. To enhance interpretability, SHAP is employed to provide both local and global explanations grounded in cooperative game theory. In addition to ranking the overall importance of key drivers, SHAP characterizes the direction and potential nonlinear forms of their effects (e.g., threshold-like responses). Using CNLI as the dependent variable, the proposed framework is applied to compare driving mechanisms across two periods (2010–2015 and 2016–2022).

Guided by this framework, the study pursues three objectives: (1) to construct the CNLI for the 2010–2022 period and conduct correlation and consistency tests with the UDL, thereby evaluating the reliability and sensitivity to anomalous years of the long-term measurements; (2) to reveal the spatiotemporal evolution patterns of urban built-up areas in the Yangtze River Delta (YRD), including the shifts in their center of gravity and directional characteristics; and (3) to identify the key driving factors of the CNLI based on the RF-RFECV, and utilize the SHAP model to delineate their contributions, directionality, and stage-specific differences, ultimately providing an empirical basis for high-quality urbanization and spatial governance in the YRD urban agglomeration (Figure 1).

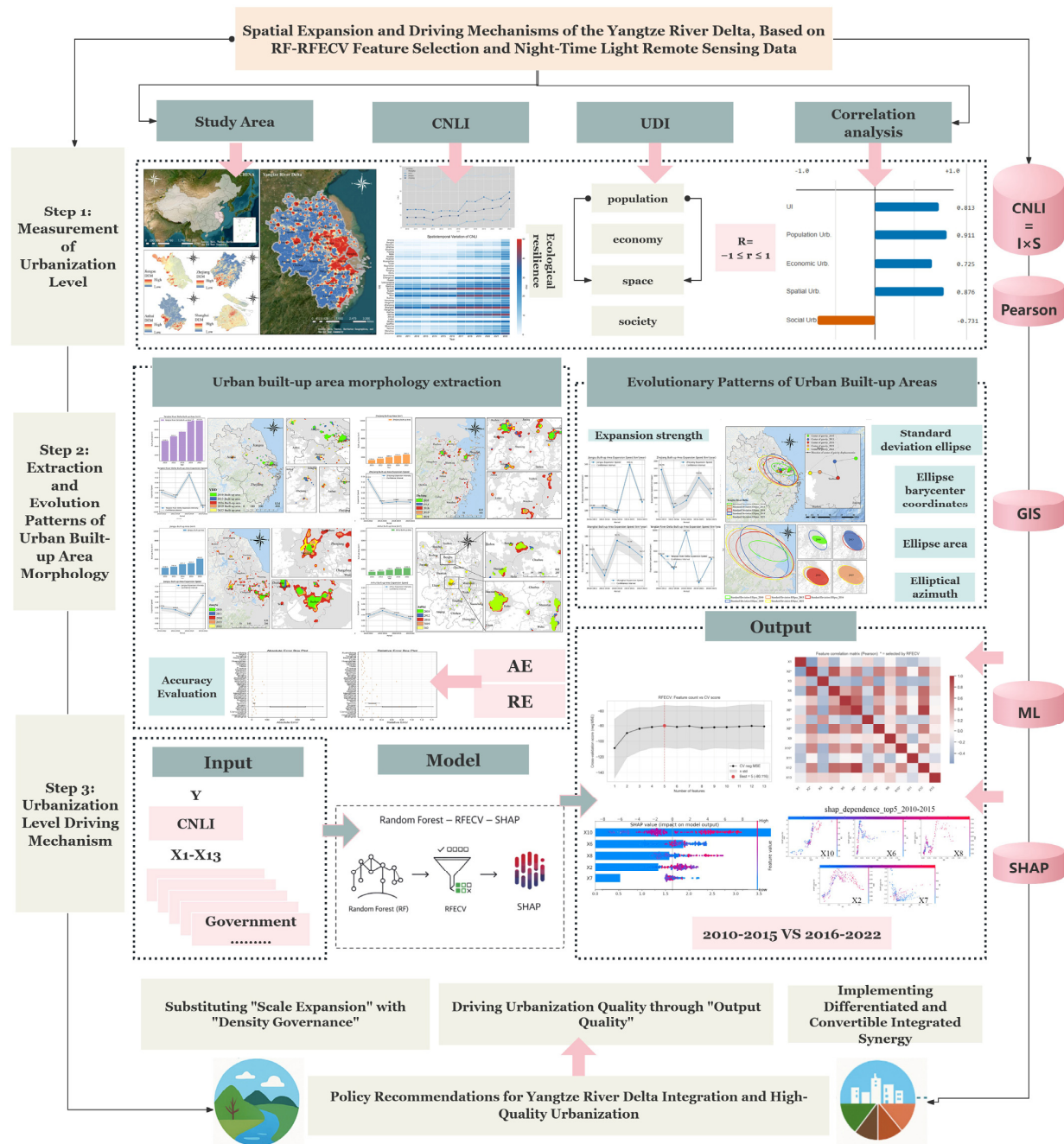


Figure 1. Research framework.

2. Materials and Methods

2.1. Study Area

The YRD is an alluvial plain located near the estuary of the Yangtze River, with Shanghai serving as its core. According to the Outline of the Yangtze River Delta Regional Integrated Development Plan issued in December 2019, the YRD comprises Shanghai Municipality and the three provinces of Jiangsu, Zhejiang, and Anhui, covering 41 prefecture-level cities, including Shanghai, Nanjing, Suzhou, Hangzhou, and Hefei [18]. The region spans approximately 358,000 km² and is home to more than 200 million permanent residents. It represents one of China's most highly urbanized areas, characterized by the highest degree of urban concentration and among the greatest city densities nationwide. The YRD also constitutes a key intersection between the Belt and Road Initiative and the Yangtze River Economic Belt, occupying a pivotal strategic position in China's national

modernization agenda and opening-up framework. It serves as a leading region of the Yangtze River Economic Belt, a major engine of China's socio-economic development, and an important platform for international competition (Figure 2).

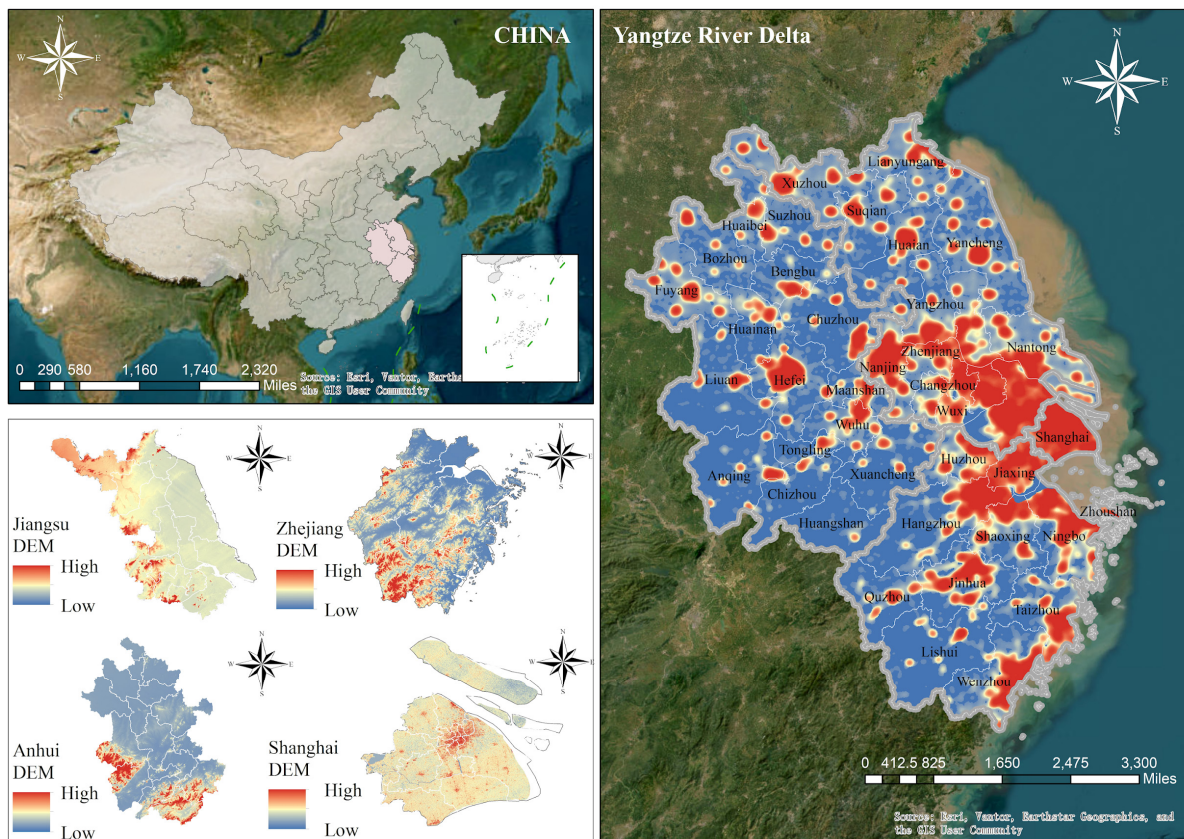


Figure 2. Study area.

2.2. Data Sources

2.2.1. Remote Sensing Data

The nighttime light data utilized in this study are derived from the cross-sensor calibrated annual ‘DMSP-OLS-like’ nighttime light dataset for China, published by Associate Professor Shi Kaifang’s team [19]. The description of ‘data calibration and fusion’ in the original manuscript refers to the construction process this dataset underwent prior to its release. To ensure the reliability of the long-term time series, the dataset creators have already completed rigorous cross-sensor harmonization at the product level. The core steps included: (1) extracting features using an Auto-encoder model; (2) establishing cross-sensor mapping relationships to convert the absolute radiance of (S)NPP/VIIRS into a discrete DMSP-like DN scale (value range: 0–63); and (3) eliminating spurious abrupt changes caused by sensor discrepancies.

Given that this dataset already possesses a high degree of cross-sensor consistency, this study does not perform secondary cross-sensor regression calibration on the raw DMSP and VIIRS data. Instead, the data processing efforts in this study primarily focused on quality control at the study area scale, with the specific steps as follows:

Step 1: Data extraction and spatial harmonization. Data for the study area from 2010 to 2022 were extracted from the annual sequential products, reprojected to a unified coordinate system, and resampled to an identical spatial resolution.

Step 2: Outlier control. Extremely high anomalous pixels generated by incidental factors (e.g., gas flares, auroras) within the study area were removed to ensure that the DN values were strictly distributed within the valid range of 0–63.

Step 3: Temporal continuity check. A time-series logical consistency filtering method was applied to ensure that, assuming no significant urban decline occurred, the DN value of a given pixel maintained a non-decreasing trend over the time series. This procedure provides a reliable foundation for the subsequent construction of the Comprehensive Nighttime Light Index (CNLI) and the analysis of spatiotemporal evolution.

2.2.2. Statistical Data

This study systematically compiled statistical yearbook data for all prefecture-level cities in the YRD from 2010 to 2022, covering multiple dimensions including demographic, economic, social, and spatial attributes [20]. The data were obtained from the China City Statistical Yearbook (2010–2022), the Jiangsu Statistical Yearbook (2010–2022), the Anhui Statistical Yearbook (2010–2022), the Zhejiang Statistical Yearbook (2010–2022), and the Shanghai Statistical Yearbook (2010–2022). These statistics not only provide a basis for validating and cross-analyzing nighttime light data, but also support key procedures such as constructing a composite urbanization assessment indicator system, conducting correlation analyses, and evaluating the accuracy of built-up area extraction.

2.2.3. Ancillary Data

Vector administrative boundary data were obtained from the 1:4,000,000 National Fundamental Geographic Information dataset of China.

2.3. Methods

2.3.1. Urbanization Level Estimation

In this study, nighttime light (NTL) remote-sensing data were used as the core data source. Based on the Compound Nighttime Light Index (CNLI), we quantified the spatiotemporal dynamics of urbanization at the prefecture-level city scale in the YRD from 2010 to 2022. To ensure the scientific validity and applicability of CNLI for quantitative urbanization assessment, the Urbanization Development Index (UDI) was used as a benchmark, and Pearson correlation analysis was employed to systematically examine the inter-annual and inter-city consistency and to validate the measurement performance of the two indices [21–23]. Specifically, we first computed the annual UDI values and its four sub-indices for each of the 41 prefecture-level cities in the YRD during 2010–2022. We then evaluated the overall goodness-of-fit between CNLI and UDI using Pearson correlation coefficients, and further assessed the correlation strength between CNLI and each UDI dimension—population urbanization, economic urbanization, spatial urbanization, and social urbanization. This procedure enables not only a quantitative assessment of the degree of agreement between CNLI and UDI (and its sub-indices), but also a rigorous examination of the suitability and robustness of NTL data for multi-dimensional urbanization quantification.

1. Calculation of urbanization level (CNLI) based on night-time light remote sensing data

CNLI directly reflects the intensity of night-time lights in urban areas, and changes in night-time light are typically closely related to the urbanization process. Therefore, this study adopts the CNLI index proposed by Zhuo (2003) [24] to measure the urbanization level of the Yangtze River Delta. The specific calculation process consists of the following three steps [25].

Step 1: Extraction of effective lit pixels and threshold setting

To mitigate background noise and ensure inter-period consistency, this study adopts a zero-threshold approach to extract effective lit pixels. Within the administrative boundary

of a prefecture-level city c in year t , let $D_{p,c,t}$ denote the nighttime light brightness (DN value) of pixel p . The set of effective lit pixels, denoted as $\Omega_{c,t}$, is defined as follows:

$$\Omega_{c,t} = \{p \mid D_{p,c,t} > 0\},$$

Specifically, only pixels with a brightness value strictly greater than 0 are identified as effective lit pixels, while the remainder are classified as unlit background and excluded from the statistical calculations. This approach precludes the biases introduced by subjective threshold selection, maintains a uniform identification criterion throughout the entire study period, and effectively enhances the inter-annual comparability of the data.

Step 2: Calculation of total nighttime light and average intensity

After determining the set of effective lit pixels $\Omega_{c,t}$, let $N_{c,t} = |\Omega_{c,t}|$ denote the total number of effective pixels. Accordingly, the sum of urban nighttime light $S_{c,t}$ and the average nighttime light intensity $I_{c,t}$ for a prefecture-level city c in year t are calculated as follows:

$$S_{c,t} = \sum_{p \in \Omega_{c,t}} D_{p,c,t}.$$

$$I_{c,t} = \frac{1}{N_{c,t}} \sum_{p \in \Omega_{c,t}} D_{p,c,t}.$$

where $S_{c,t}$ represents the total expansion volume of the urban scale, and $I_{c,t}$ represents the average agglomeration intensity of nighttime lights within the city.

Step 3: Calculation of the CNLI

Finally, the final Comprehensive Nighttime Light Index (CNLI) is derived by multiplying the average nighttime light intensity by the sum of nighttime light at the city scale:

$$CNLI_{c,t} = I_{c,t} \times S_{c,t},$$

This index comprehensively accounts for both the breadth (scale) and depth (intensity) of urban development. It is utilized to scientifically characterize the urbanization levels and their spatiotemporal evolution patterns across prefecture-level cities from 2010 to 2022.

2. Urbanization level estimation based on statistical yearbook panel data (UDI)

With the continuous expansion of the concept of urbanization, evaluation standards have evolved from early single-indicator schemes to comprehensive assessment frameworks spanning multiple domains and dimensions. Urbanization is not merely a quantitative shift of population from rural to urban areas; rather, it entails a series of structural transformations, including industrial upgrading and enhanced efficiency of factor allocation, the reconfiguration of land-use patterns and spatial organization, and improvements in public service provision and social welfare [26]. Consequently, relying solely on a single indicator—such as the share of urban population or the size of built-up areas—fails to capture the integrated implications of “quality improvement” and “coordinated development” in urbanization. Against this backdrop, drawing on the existing literature and considering the specific development characteristics of the YRD, this study follows the principles of comprehensiveness, representativeness, objectivity, and data availability in indicator selection. We conceptualize urbanization through four criterion dimensions: population urbanization, economic urbanization, spatial urbanization, and social urbanization [27,28], thereby systematically reflecting multiple urbanization effects, including people-centered development, economic growth, social equity, and cultural advancement. Specifically, population urbanization emphasizes population agglomeration, employment-structure transition, and the process of citizenization, and thus represents the fundamental manifestation of urbanization. Economic urbanization captures the level of industrial development, the evolution of economic structure, and the capacity for factor agglomeration, serving as

the core driving force and material basis for sustained urbanization. Spatial urbanization refers to the expansion of built-up land, the restructuring of spatial form, and the optimization of infrastructure configuration, representing the direct projection of urbanization on the Earth's surface. Social urbanization focuses on public service provision, social security coverage, and improvements in quality of life, thereby reflecting the goals of people-centered new-type urbanization and the extent to which urbanization benefits are shared [29]. These four dimensions are mutually coupled and constrained, jointly shaping the scale, pace, and quality of urbanization, and providing a comprehensive explanation for the urbanization characteristics and disparities of the YRD under conditions of high-intensity development, frequent factor mobility, and increasing demand for equalized public services.

This study constructs an evaluation indicator system comprising 21 refined indicators to measure the urbanization level of the Yangtze River Delta (YRD) urban agglomeration. Regarding the calculation of indicator weights, to circumvent the artificial biases introduced by subjective weighting, this study employs the Entropy Weight Method (EWM) for objective weight assignment. By determining the weights based on the degree of dispersion of the indicator data, this method scientifically reflects the information content and relative importance of each indicator within the comprehensive evaluation. For further details, please refer to Table A1.

2.3.2. Consistency Validation Between the Urbanization Development Index (UDI) Based on Statistical Yearbook Panel Data and the CNLI

To verify the applicability and scientific validity of the nighttime light-driven CNLI in the quantitative evaluation of urbanization, this study conducts statistical consistency tests using the UDI and its four sub-indices as benchmarks. Specifically, this study performs correlation comparisons at the annual cross-sectional level: for each year t , using the CNLI and UDI of 41 prefecture-level cities as samples, the Pearson correlation coefficient is calculated, followed by a two-tailed significance test (significance level $\alpha = 0.05$). Simultaneously, the Fisher r -to- z transformation is employed to construct a 95% confidence interval for each correlation coefficient, which is used to report the uncertainty range of the correlation estimates. In addition to the Pearson correlation, to test the robustness against outliers, the Spearman rank correlation is additionally reported as a reference.

Considering that nighttime light data in certain years may be affected by noise or extreme values, leading to anomalous fluctuations in city-scale indicators, this study first screens the inter-annual change rate of the CNLI for each city. Extreme observations located at both ends of the full-sample distribution are flagged as anomaly candidates and verified in conjunction with the spatial distribution of the nighttime light rasters. Subsequently, the confirmed anomalous observations are Winsorized by truncating them to the 1–99% percentile range. In the sensitivity analysis, the CNLI-UDI correlations and subsequent model results under two scenarios (“untreated vs. Winsorized”) are compared to ensure that the conclusions do not rely on a few anomalous years or extreme city samples.

2.3.3. Methods for Analyzing the Expansion Pattern of Urban Built-Up Areas

1. Urban built-up area extraction and accuracy evaluation

Built-up area extraction uses remote sensing imagery and geographic information systems (GIS) to identify and delineate the built environment of a city or region from multi-source geospatial data, including artificial constructions such as buildings, roads, plazas and various infrastructures [30]. In this study, night-time light remote sensing data are used as the primary source, and a bisection model is applied to automate the extraction of built-up areas in the Yangtze River Delta. The bisection model has the advantages of few parameters, fast convergence and strong adaptability, and can effectively overcome the subjectivity and

inefficiency of traditional manual interpretation and empirical thresholding methods. In practice, the target built-up area parameter is first set according to prefecture-level built-up area statistics published in statistical yearbooks. Then, in combination with the DN value distribution characteristics of the night-time light imagery, the brightness threshold is dynamically adjusted to classify pixels into “built-up” and “non-built-up” categories. By iteratively computing the error between the extraction result and the statistical built-up area and minimizing the absolute error, the optimal extraction threshold is determined and the optimal spatial distribution of built-up areas is obtained [31].

To comprehensively evaluate the accuracy and reliability of built-up area extraction, this study introduces two indicators: Absolute Error (*AE*) and Relative Error (*RE*). Their calculation formulas are as follows:

$$AE = Area_{statistics} - Area_{light}$$

$$RE = \frac{Area_{light} - Area_{statistics}}{Area_{statistics}} \times 100\%$$

Arealign denotes the built-up area derived from night-time light data, and *Areastatistics* denotes the built-up area reported in statistical yearbooks. By comparing *AE* and *RE* across different cities and years, the applicability and error distribution of the bisection-model extraction method for large-scale built-up area identification can be systematically assessed, thereby providing a solid data foundation and methodological support for subsequent spatiotemporal studies of urbanization.

2. Expansion Intensity

Expansion intensity measures the relative growth rate of built-up area per unit time, suitable for reflecting the efficiency and intensity of urban expansion. It is commonly used to analyze the relative rate of land expansion in cities of different sizes [32]. This indicator eliminates the influence of city size, facilitating cross-city comparisons. The calculation formula is:

$$EI = (UA_t - UA_0) / (UA_0 \times \Delta t)$$

where:

UA_t : End-of-period built-up area.

UA_0 : Beginning-of-period built-up area.

Δt : Duration of the study period (years).

EI represents the average annual relative expansion rate of the built-up area during the study period compared with its initial (beginning-of-period) size, reflecting the intensity and efficiency of built-up area expansion.

3. Standard deviation ellipse

The Standard Deviation Ellipse (SDE) is a widely used spatial analysis method for characterizing the central tendency and distributional morphology of geographic features. The SDE is an ellipse centered on the mean coordinate of the analyzed features, with the lengths of the major and minor axes determined by the standard deviations of the feature distribution along the principal directions. It, therefore, summarizes both the directional trend and the spatial dispersion of the geographic distribution.

The length of the principal axis (x-axis) of the standard deviation ellipse:

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

where σ_x denotes the length of the major axis of the standard deviational ellipse in the x-direction, x_i is the x-coordinate of the i -th sample point, $\bar{x}$ is the mean of the x-coordinates, and n is the total number of sample points.

The length of the secondary axis (y-axis) of the standard deviation ellipse:

$$\sigma_y = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n}}$$

where σ_y denotes the length of the minor axis of the standard deviational ellipse in the y-direction, y_i is the y-coordinate of the i -th sample point, $\bar{y}$ is the mean of the y-coordinates, and n is the total number of sample points.

The standard deviational ellipse summarizes the central tendency and spatial dispersion of a feature distribution. Characteristics of the ellipse—such as the ratio of major to minor axis, orientation angle, and area—can be used to analyze spatial attributes of the distribution. Specifically, the ellipse centroid represents the center of gravity of built-up area distribution for the study region; the ellipse area reflects the degree of spatial concentration versus dispersion; and the orientation (azimuth) indicates the principal direction of urban expansion. The formulas for calculating the centroid, area, and azimuth of an ellipse are as follows:

Ellipse barycenter coordinates (x, y):

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

where $\bar{x}$ and $\bar{y}$ are the x- and y-coordinates of the ellipse's centroid, respectively; x_i and y_i are the x- and y-coordinates of the i -th sample point; and n is the total number of sample points.

Ellipse area (A):

$$A = \pi \times a \times b$$

where A is the area of the ellipse, and a and b are the semi-major and semi-minor axes (radii) of the ellipse, respectively.

Elliptical azimuth (θ):

$$\theta = \frac{1}{2} \text{atan2} \left(2 \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}), \sum_{i=1}^n (x_i - \bar{x})^2 - \sum_{i=1}^n (y_i - \bar{y})^2 \right).$$

2.3.4. Analysis Methods for Driving Mechanisms of Urbanization Level

1. Indicator System and Variable Definitions for Urbanization Drivers

Building on prior studies on the determinants of urbanization and considering the development characteristics of the Yangtze River Delta (YRD) urban agglomeration, this study constructs an indicator system of urbanization drivers organized into four dimensions: administrative force, market force, external (openness) force, and endogenous force. In addition to exogenous drivers such as government investment and intervention, market development and factor allocation, and external economic linkages, we further characterize the endogenous foundations and quality constraints of urbanization from four aspects—population, society, innovation, and environment. Specifically, population captures the scale effects of agglomeration and factor allocation as well as potential congestion costs, thereby directly reflecting the conditions for urban expansion and carrying capacity. Society represents the supporting role of public service provision and human-capital accumulation in facilitating in-migration, residential stability, and urban attractiveness. Innovation reflects the impetus of industrial upgrading and technological progress for trans-

forming urban growth modes, which is particularly consistent with the YRD's stage-wise shift from factor-driven growth toward efficiency- and green-transition-oriented development. The environment captures ecological carrying capacity and livability constraints, constituting a key boundary condition for achieving quality- and efficiency-enhancing, sustainable urbanization in high-density areas.

Based on this framework, 13 indicators (X1–X13) are selected as explanatory variables, and the CNLI for each prefecture-level city is used as the dependent variable. An interpretable machine-learning framework is then employed to examine the urbanization driving mechanisms in the YRD. Data sources and calculation methods for all indicators are provided in Table 1.

Table 1. Indicator system of urbanization influencing factors for the Yangtze River Delta (YRD) urban agglomeration, China.

Type	Primary Dimension	Secondary Dimension	Indicator	Variable
Dependent variable	Urbanization level	CNLI	/	Y
Independent variables	Administrative force	Government investment and intervention	Policy intervention intensity	X1
		Fiscal structure	Degree of fiscal decentralization	X2
		Infrastructure development	Infrastructure level	X3
	Market force	Economic development	GDP per capita	X4
		Industrial structure	Industrial structure upgrading	X5
		Market size	Total retail sales of consumer goods	X6
		Transport conditions	Transport accessibility	X7
	External force	External economic linkages	Degree of openness	X8
		External economic linkages	Foreign direct investment (FDI) level	X9
	Endogenous force	Demographic factors	Population density	X10
		Social development	Education expenditure level	X11
		Innovation capacity	Total number of authorized green patents	X12
		Environmental quality	Green coverage ratio in built-up areas	X13

2. Machine learning-driven mechanism identification framework (RF-RFECV + SHAP)

Data preparation (samples, variables, and preprocessing):

A panel dataset $\{(x_i, y_i)\}_{i=1}^n$ is constructed using “city-year” as the sample unit. Here, y_i denotes the urbanization level indicator CNLI, and $x_i \in \mathbb{R}^p$ represents the vector of candidate driving factors. This study predetermines $p = 13$ potential influencing factors, and the same procedure is executed separately on two-stage sub-samples (2010–2015 and 2016–2022) to compare inter-stage differences.

Random Forest base learner and parameter settings:

This study employs Random Forest (RF) regression as the base learner to capture non-linear relationships and potential interactions. The RF prediction function is expressed as:

$$\hat{f}(x) = \frac{1}{B} \sum_{b=1}^B T_b(x)$$

where B denotes the total number of trees, and $T_b(\cdot)$ represents the b -th regression tree. The model parameters are set as follows: the number of trees is $B = 500$; `max_features` is set to “sqrt” to control the number of candidate features evaluated during the splitting of an individual tree, thereby mitigating overfitting and enhancing generalization stability.

Apart from the aforementioned key parameters, the remaining hyperparameters adopt the default settings of the scikit-learn Random Forest regressor.

RFECV feature selection process (RF-RFECV):

To screen a robust and interpretable subset of variables while ensuring predictive performance, this study employs RFECV to perform recursive feature elimination within a cross-validation framework. The procedure is as follows: (1) Train the Random Forest model using the current feature subset; (2) rank the variables based on the feature importance provided by the model, and eliminate the feature with the least contribution; (3) conduct cross-validation evaluations on feature subsets of varying sizes, and select the number of features and the corresponding feature set that yield the optimal validation error.

The specific implementation settings are as follows: a 10-fold KFold is adopted for cross-validation; RFECV eliminates one feature per iteration (step = 1); the scoring metric used is `neg_mean_squared_error` (where a smaller absolute value indicates a lower mean squared error and better model performance); and the minimum number of features to retain is set to 1 (`min_features_to_select` = 1). The optimal feature set is denoted as S^* , and its selection principle can be expressed as:

$$S^* = \arg \min SCV\text{Err}(S)$$

where $CV\text{Err}(S)$ represents the mean squared error under 10-fold cross-validation.

Validation strategy:

This study utilizes the 10-fold cross-validation error within the RFECV as the optimization criterion for feature subsets, thereby conducting a consistent evaluation of the model's generalization error across different feature dimensions. To ensure the comparability of cross-period analysis, the two stages independently screen and obtain their respective optimal feature subsets and models using the same candidate variable set (13 items) and identical algorithm parameters.

SHAP value calculation method and interpretation logic:

After obtaining the optimal feature subset S^* , this study further employs SHAP to interpret the Random Forest model, aiming to quantify the magnitude, direction, and heterogeneity of the contributions of key features to the CNLI prediction. The model output for sample i can be written in an additive decomposition form:

$$\hat{y}_i = \phi_0 + \sum_{j \in S^*} \phi_{ij}$$

where ϕ_0 is the baseline output (typically the expected value of the predictions on the training set), and ϕ_{ij} represents the marginal contribution (Shapley value) of feature j to the predicted value for sample i . At the reporting and interpretation level, this study adopts three statistics to characterize the global effects of the features:

`mean_abs_shap`: $\text{mean}(|\phi_{ij}|)$, which measures the average absolute contribution of a feature to the prediction and is used for global importance ranking.

`mean_shap`: $\text{mean}(\phi_{ij})$, which reflects the overall direction of the feature's impact on the prediction.

`std_abs_shap`: $\text{std}(|\phi_{ij}|)$, which measures the heterogeneity of the impact intensity.

3. Results

3.1. Spatiotemporal Evolution Characteristics of Urbanization Level in the Yangtze River Delta

The Pearson correlation results are presented in Table 2. Using pooled city-year observations for the 41 prefecture-level cities in the Yangtze River Delta over 2010–2022 ($n = 533$), the composite nighttime light-based urbanization index (CNLI) is strongly and positively correlated with the Urbanization Development Index (UDI), with $r = 0.813$

($p < 0.01$). This suggests substantial overall agreement between the two measures across the pooled sample. Further analyses show that CNLI is significantly correlated with the UDI subindices for population urbanization ($r = 0.911$), economic urbanization ($r = 0.725$), spatial urbanization ($r = 0.876$), and social urbanization ($r = 0.645$) (all $p < 0.01$). These results are broadly consistent with previous evidence on the multidimensional nature of urbanization [33]. Notably, the strongest correlations are observed for population and spatial urbanization, supporting the suitability of CNLI for capturing regional urbanization levels and associated spatial heterogeneity.

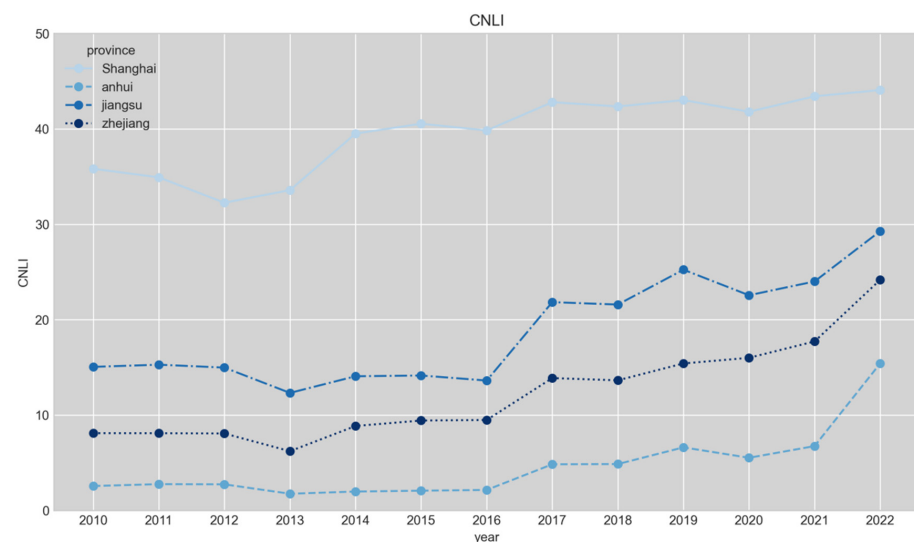
Table 2. Pearson correlation analysis results of CNLI and UDI for the Yangtze River Delta.

Index (vs. CNLI)	r	p-Value
UDI	0.813	<0.01
Population Urbanization	0.911	<0.01
Economic Urbanization	0.725	<0.01
Spatial Urbanization	0.876	<0.01
Social Urbanization	0.645	<0.01

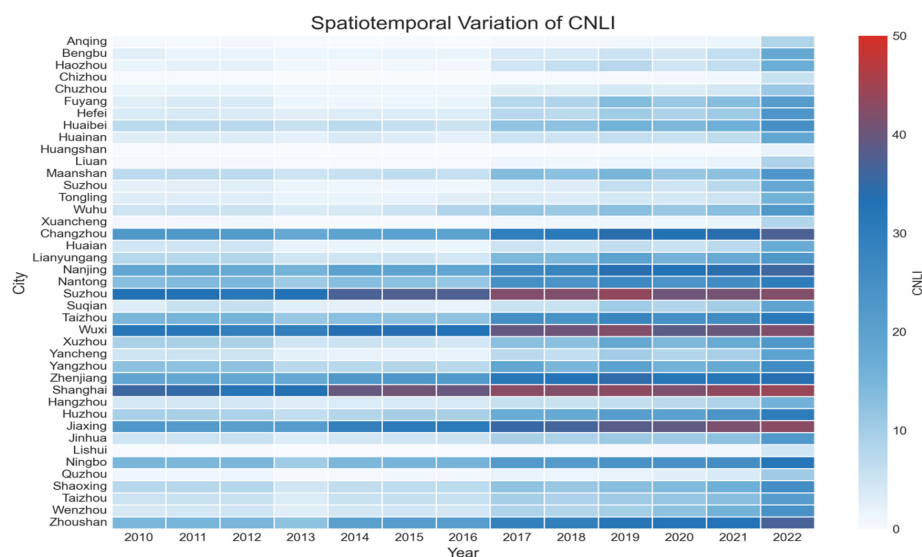
Temporally, the urbanization level of the YRD exhibited a sustained upward trend from 2010 to 2022, with the three provinces and one municipality following differentiated paths characterized as “high-level leadership, steady advancement, and late-comer catch-up [28].” Shanghai consistently ranked highest in urbanization: its CNLI increased steadily from 35.82 in 2010 to 44.07 in 2022, demonstrating its role as the regional core. The growth rate in later years tended to stabilize, reflecting features of high-quality, intensive development. Jiangsu followed closely, rising from 15.05 to 29.28 over the period, showing steady improvement with episodic acceleration. Zhejiang started from a middle–high level (8.11) and experienced marked acceleration after 2017—coinciding with regional integration policies—reaching 24.18 in 2022 and displaying strong catching-up momentum. Anhui began at the lowest level (2.57 in 2010) but, benefiting from late-comer advantages and regional integration dividends, achieved leapfrogging growth after 2017, rising to 15.41 in 2022; the absolute gap with Shanghai, Jiangsu and Zhejiang narrowed substantially. This dynamic pattern indicates that, since 2017, a clearer “leading head—late-comer catching up” dynamic has emerged and intra-regional disparities have tended to converge, reflecting the significant effect of YRD integration policies in promoting coordinated quality improvement [34].

At the prefecture level, cities follow markedly heterogeneous trajectories. Leading cities such as Shanghai, Suzhou, Wuxi, Nanjing, Hangzhou, Ningbo, and Jiaxing have long exhibited CNLI values well above the regional average—exceeding 40 in 2022—and continue to exert strong driving effects on the region. Simultaneously, late-starter cities represented by Hefei, Wuhu, and Ma’anshan in Anhui have shown significant acceleration under regional coordination policies and have risen into the second tier of regional development. Nevertheless, despite widespread increases in urbanization and acceleration phases observed in 2016–2019 and after 2020, a number of small and medium cities in Northern and Southern Anhui, Northern Jiangsu and Western Zhejiang (e.g., Anqing, Huangshan, Chizhou, Lishui) remain constrained by a low baseline; their absolute levels continue to lag substantially behind the leading cities, indicating persistence of the “Matthew effect” [35] (Figure 3A). This city-level differentiation pattern is closely associated with the transition of regional integration into a stage of national top-level design and institutionalized advancement following the promulgation of the Outline of the Regional Integrated Development Plan of the Yangtze River Delta in 2019. First, the Outline promoted the interconnection of cross-provincial infrastructure and the expansion of commuting circles,

which reduced spatial friction costs. Consequently, high-end employment opportunities, research and development (R&D) platforms, and producer services became more prone to agglomerating in core and strong-node cities characterized by greater market thickness and more comprehensive industrial chains. This dynamic further amplified the advantages of factor absorption and resource allocation in Shanghai and the leading cities of Southern Jiangsu and Northern Zhejiang during the 2019–2022 period. Second, the advancement of industrial synergy and division of labor systems frequently manifests as a hierarchical structure of “core R&D/management versus peripheral manufacturing/supporting.” Certain recipient nodes (e.g., Hefei and Wuhu) managed to accelerate their catch-up process and enter the second tier by leveraging improvements in their network positions. Conversely, peripheral cities lacking the capacity for industrial transfer and adequate public service provision risk being locked into low-value-added segments, thereby perpetuating relative disparities. Third, compounded by the shock of the COVID-19 pandemic and the digital transformation post-2019, the rising preference of high-end production factors for superior institutional environments and high-quality public services has further reinforced the cumulative advantage of “the strong getting stronger”.

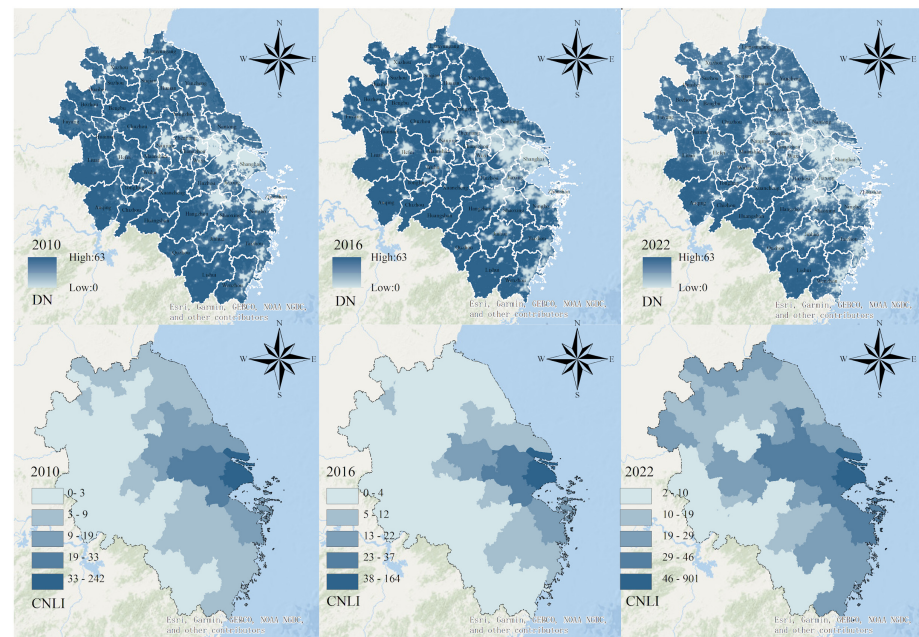


(A)



(B)

Figure 3. Cont.



(C)

Figure 3. (A) Line chart of temporal changes in the CNLI index for Anhui Province, Jiangsu Province, Shanghai Municipality, and Zhejiang Province (2010–2022); (B) heatmap of the CNLI index for 41 prefecture-level cities in the Yangtze River Delta (2010–2022); (C) spatial distribution maps of DN and CNLI indices for the Yangtze River Delta (2010–2022).

At the provincial scale, Anhui has generally entered a rapid urbanization track but exhibits pronounced internal heterogeneity. Southern Jiangsu cities continue high-level growth, yet a pronounced north–south gradient remains (e.g., Suzhou 42.36 vs. Suqian 19.44 in 2022). Zhejiang’s overall growth is robust, with internal coastal–mountain divergence showing signs of convergence. Shanghai continues to radiate high-quality, fine-grained development to its surroundings (Figure 3B).

Spatially, the urbanization level of the YRD from 2010 to 2022 exhibits a characteristic “higher in the east, lower in the west; higher in the south, lower in the north” pattern (Figure 3C). Shanghai functions as the regional core, with CNLI values consistently well above other cities. Jiangsu Province maintains a long-term high urbanization level, forming a continuous high-value belt along the Yangtze River and in Southern Jiangsu. Zhejiang has experienced rapid improvements from a mid-to-high level, progressively narrowing the gradient gap with Jiangsu. Although Anhui recorded a leap in CNLI from 2.57 to 15.41, its absolute gap with Shanghai, Jiangsu, and Zhejiang remains pronounced, reflecting persistent intra-regional imbalances.

At the prefecture scale, the spatial differentiation generally follows an “east-high, west-low; river-and-coast higher than inland; prominent core, and decreasing gradient” pattern [4]. High-value zones are highly concentrated in Shanghai, the Southern Jiangsu core (Suzhou, Wuxi, Changzhou, Nanjing, Nantong), and the Northern Zhejiang coastal–riverine core (Jiaxing, Ningbo, Zhoushan, Huzhou). In contrast, Northern Jiangsu, Northern Anhui, Western Zhejiang, and Southern Anhui constitute clear urbanization lowlands. Temporal analysis indicates an ongoing evolution of regional spatial structure. In 2010, high-value areas were primarily localized within core cities; by 2022, high-value zones had substantially expanded outward along major transport corridors and economic belts—such as the Shanghai–Nanjing and Hangzhou–Ningbo axes and the riverside high-speed rail lines—forming multiple high-value clusters. The spatial morphology has thus shifted from an early “core–belt” configuration toward a more mature, “networked–clustered” urban-

agglomeration structure. This spatial evolution is characterized by a simultaneous process of polarization and uplift. On the one hand, high-urbanization cities continue to strengthen their capacity to concentrate high-quality resources, which at certain stages intensifies spatial polarization. On the other hand, driven by regional integration, traditionally low-value areas have experienced substantive increases in urbanization levels, contributing to overall upward convergence. Intercity spatial correlations have strengthened, and the spillover effects of high-value cities on their surroundings has become increasingly evident, promoting the transformation of the YRD toward a more balanced and tightly connected networked urban system [36]. Compared with other Chinese urban agglomerations such as the Beijing-Tianjin-Hebei and the Pearl River Delta, the YRD's "multi-pole driving—overall upgrading" development pattern has been more effective in reducing internal gradients and optimizing spatial structure, and is progressively forming a world-class, networked urban-agglomeration spatial configuration.

3.2. Characteristics of Urban Built-Up Area Expansion in the Yangtze River Delta

3.2.1. Extraction Results of Urban Built-Up Area Morphology in the Yangtze River Delta

This study extracted the built-up area for the 41 prefecture-level cities in the Yangtze River Delta for 2010–2022 and computed both absolute error (AE) and relative error (RE) for each extraction. In terms of relative error, most cities cluster around a RE of approximately 0.2, indicating relatively small proportional deviations between extracted and reported values and, therefore, high extraction accuracy. Conventionally, RE below 0.1 is considered highly accurate, while RE within 0.2 is regarded as acceptable. A few cities (e.g., Lianyungang, Ma'anshan, and Luan) exhibit larger relative errors exceeding 0.2; this is attributable to interference in DMSP/OLS-based extraction introduced by large contiguous water bodies (e.g., Taihu Lake, the East China Sea), which affect the radiance distribution and thus the extraction accuracy [37]. With respect to absolute error, thresholds of 100 km² and 200 km² are commonly used to indicate small and acceptable errors, respectively. Most cities in this study show absolute errors on the order of 100 km². Overall, the extracted dataset demonstrates satisfactory accuracy and reliability for subsequent spatiotemporal analyses.

3.2.2. Expansion Rate of Urban Built-Up Area in the Yangtze River Delta

From 2010 to 2022, inter-provincial differences in urban spatial expansion rates within the YRD were pronounced (Figure 4) Jiangsu and Zhejiang exhibited the highest expansion rates—8.418% and 8.194%, respectively—substantially exceeding those of Anhui (5.259%) and Shanghai (4.736%) (Table 3). This spatial gradient of expansion is closely related to economic base, industrial structure, policy orientation, and natural conditions. Jiangsu and Zhejiang possess strong economic foundations and diversified industry structures; the agglomeration of advanced manufacturing and high-tech industries has continuously attracted population and resources, thereby raising urban land demand and accelerating spatial expansion. Shanghai, by contrast, operates as a global city with an industrial focus on high-end finance, technological innovation and services; its urban development model is comparatively compact and high-density. Policy emphasis in Shanghai on optimizing core-area functions and resource allocation, and on restraining uncontrolled urban sprawl, contributes to its relatively lower expansion rate. This pattern resembles the compact-growth trajectories observed in other global metropolitan regions that, after a period of suburbanization, have implemented planning and land-use controls to limit sprawl and promote higher-density redevelopment (e.g., New York metropolitan area) [38]. Anhui records the lowest expansion rate, largely due to a weaker economic base and more restrictive natural terrain. Anhui's spatial growth has been driven mainly by the reception of industrial transfers from the Shanghai–Jiangsu–Zhejiang core, following a relatively

conservative growth pattern concentrated in industrial parks and new-development zones; limited plain area further constrains uncontrolled urban expansion. At the regional scale, the YRD's urban expansion rate peaked during 2016–2019 (annual rate ≈ 0.1313). Since 2019, however, multiple structural shocks have altered the macro environment: national policy clarified a “housing for living, not for speculation” stance and introduced the “three red lines” to tighten land and credit supply; economic growth deceleration and persistent global trade frictions weakened market demand; and the COVID-19 outbreak in 2020 directly halted many construction projects via disruptions to construction activity, supply chains and municipal finances. Under the combined effect of these three pressures, the expansion rate plunged during 2019–2022 to near stagnation (≈ 0.0013), marking a phase-specific exhaustion of the region's previous urbanization expansion momentum.

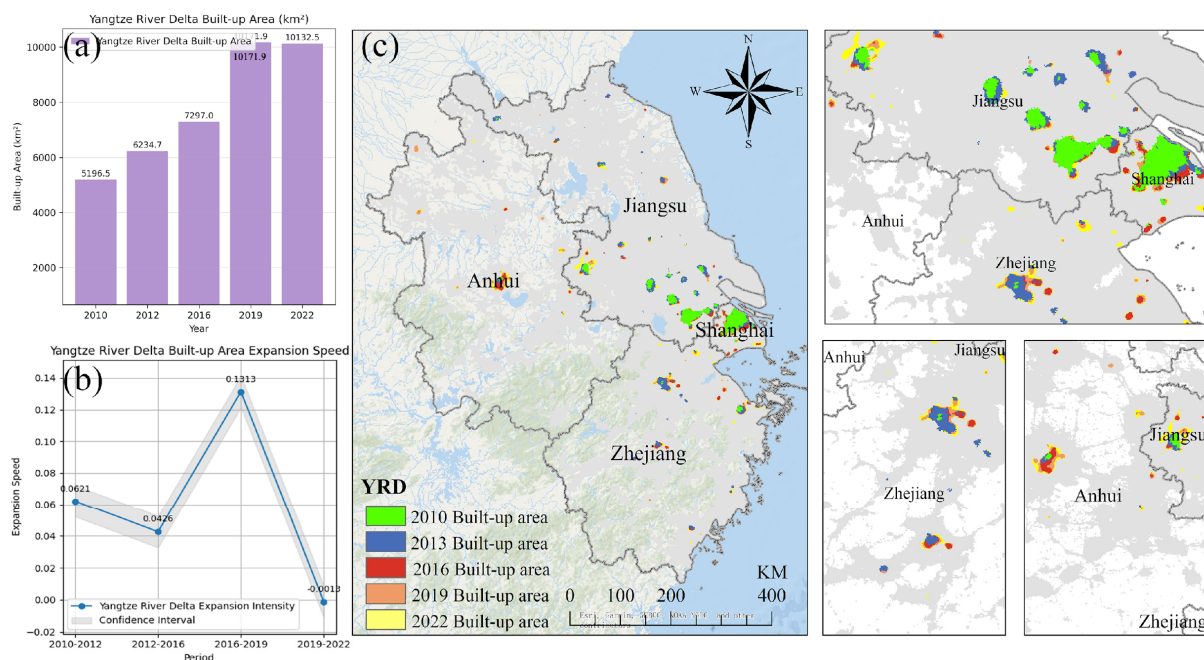


Figure 4. Spatiotemporal evolution of built-up areas in the Yangtze River Delta (2010–2022): (a) total built-up area (km²) from 2010 to 2022; (b) built-up area expansion speed across periods; (c) maps of built-up areas in 2010, 2013, 2016, 2019, and 2022.

Table 3. Expansion rate of urban built-up areas in the Yangtze River Delta (2010–2022).

	2010–2012	2012–2016	2016–2019	2019–2022	2010–2022
Jiangsu	0.057	0.051	0.030	0.106	0.084
Zhejiang	0.130	0.042	0.046	0.045	0.082
Anhui	0.035	0.056	0.040	0.026	0.053
Shanghai	0.045	0.011	0.068	0.032	0.047
YRD	0.062	0.043	0.131	0.001	0.086

At the province/city level, Jiangsu's expansion rate rose markedly to 0.1059 during 2019–2022, indicating renewed spatial growth potential driven by a new wave of industrial clustering and innovation-led development (Figure 5). Zhejiang exhibited its highest expansion rate in 2010–2012 (0.1300), after which the rate declined progressively to 0.0417 in 2012–2016, reflecting the restraining effects of industrial upgrading and compact-development policies on land expansion (Figure 6). Anhui's expansion rates remained relatively low and stable across 2010–2022, with interval values ranging from 0.0258 to

0.0556, consistent with a “late-comer—industrial reception—steady expansion” pattern (Figure 7). Shanghai’s expansion rate dropped sharply to 0.0112 in 2012–2016, recovered somewhat to 0.0684 in 2016–2019, but remained below its 2010–2012 level, corroborating its compact, high-quality spatial development characteristics (Figure 8).

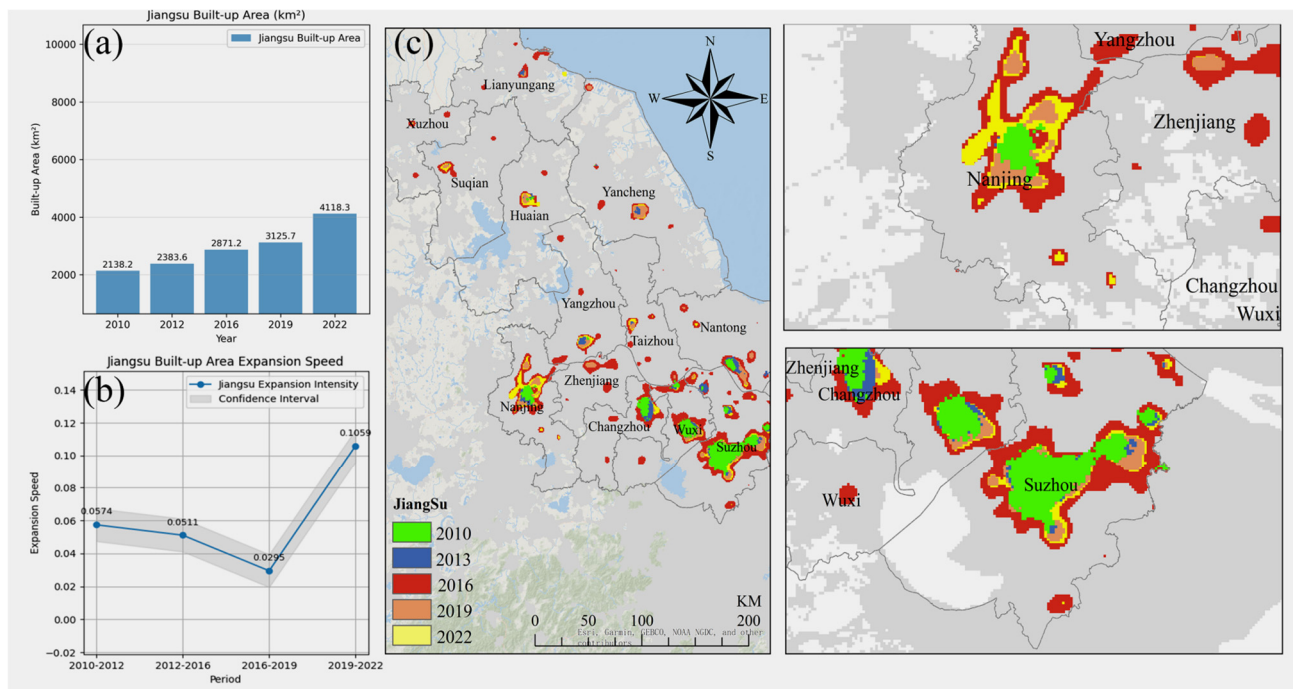


Figure 5. Spatiotemporal evolution of built-up areas in the Jiangsu urban agglomeration (2010–2022): (a) total built-up area; (b) built-up area expansion speed; (c) spatial distribution of built-up areas in 2010, 2013, 2016, 2019, and 2022.

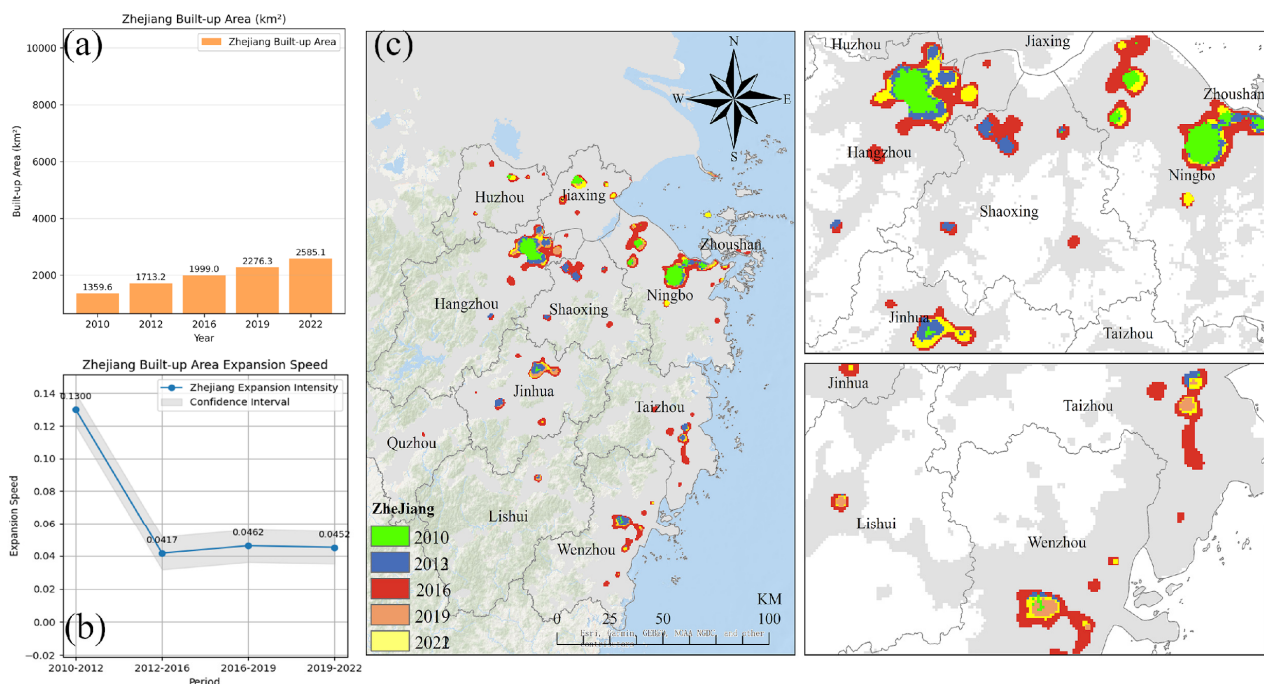


Figure 6. Spatiotemporal evolution of built-up areas in the Zhejiang urban agglomeration (2010–2022): (a) total built-up area; (b) built-up area expansion speed; (c) spatial distribution of built-up areas in 2010, 2013, 2016, 2019, and 2022.

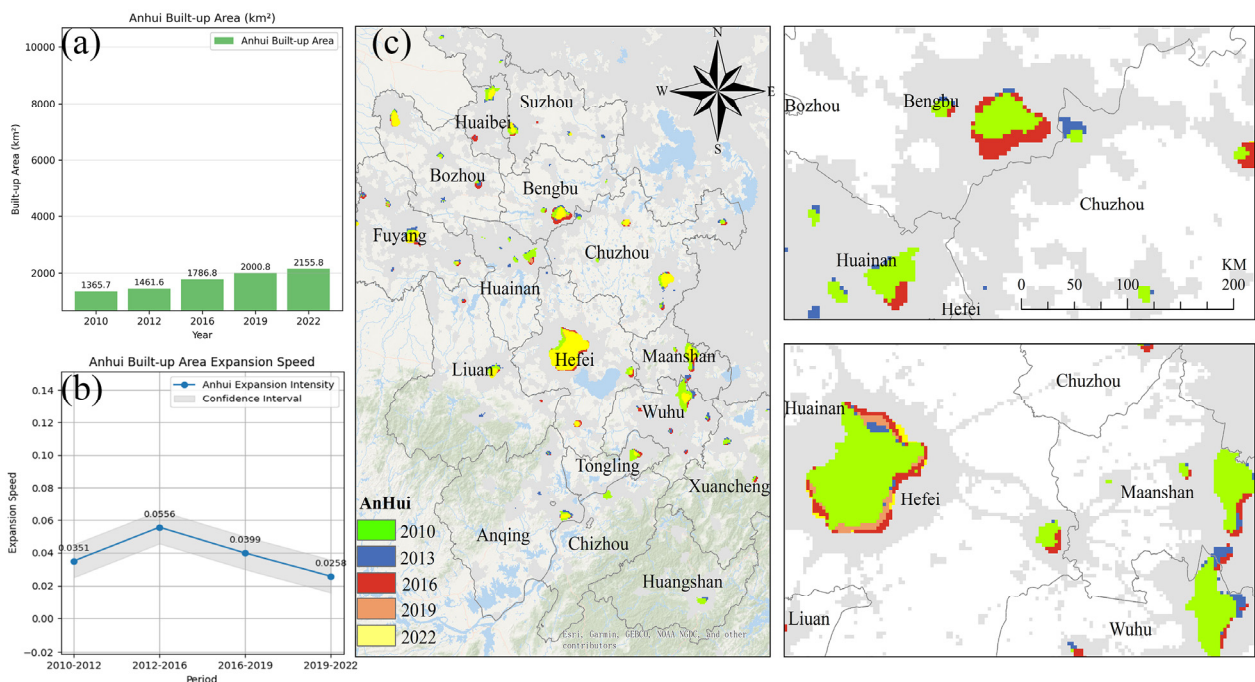


Figure 7. Spatial-temporal characteristics of built-up areas in Anhui Province. (a) Built-up area changes from 2010 to 2022. (b) Expansion speed (intensity) of built-up areas for different periods, with the confidence interval. (c) Spatial distribution of built-up areas and their expansion in Anhui Province, including the province-wide map and two zoomed-in views of representative urban agglomerations.

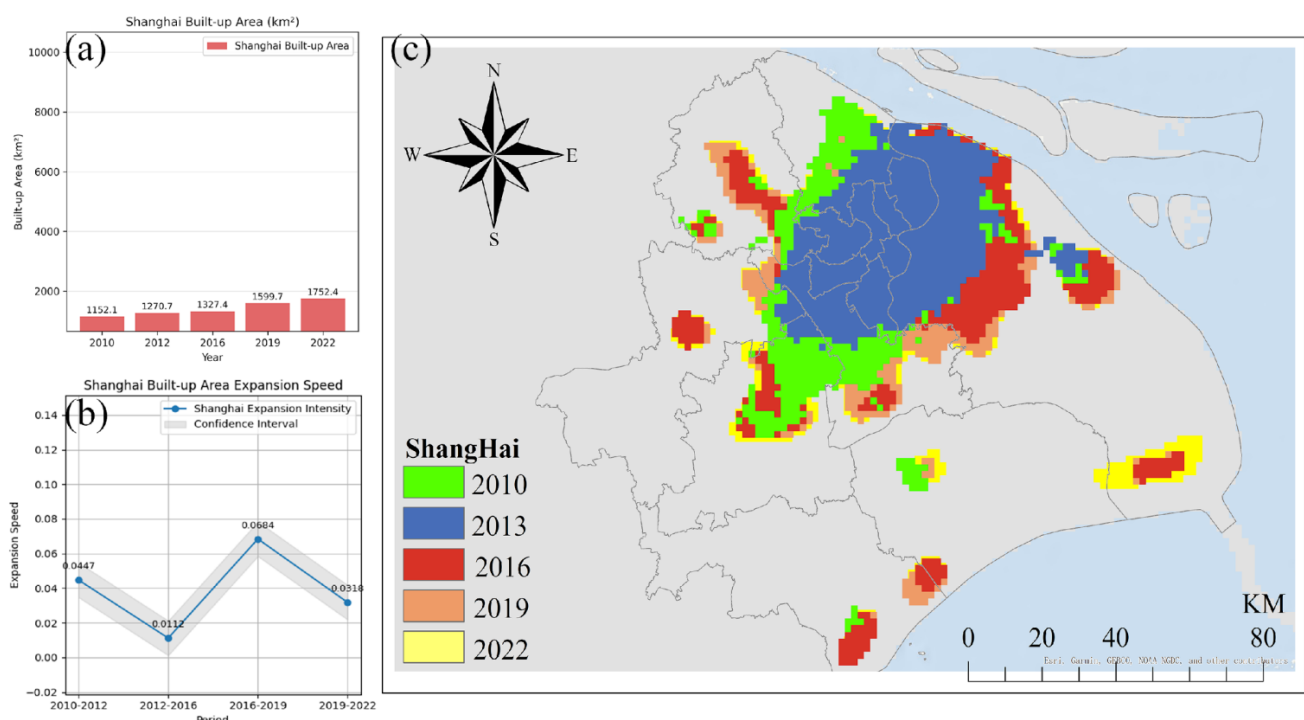


Figure 8. Spatiotemporal evolution of built-up areas in Shanghai (2010–2022): (a) total built-up area; (b) built-up area expansion speed; (c) spatial distribution of built-up areas in 2010, 2013, 2016, 2019, and 2022.

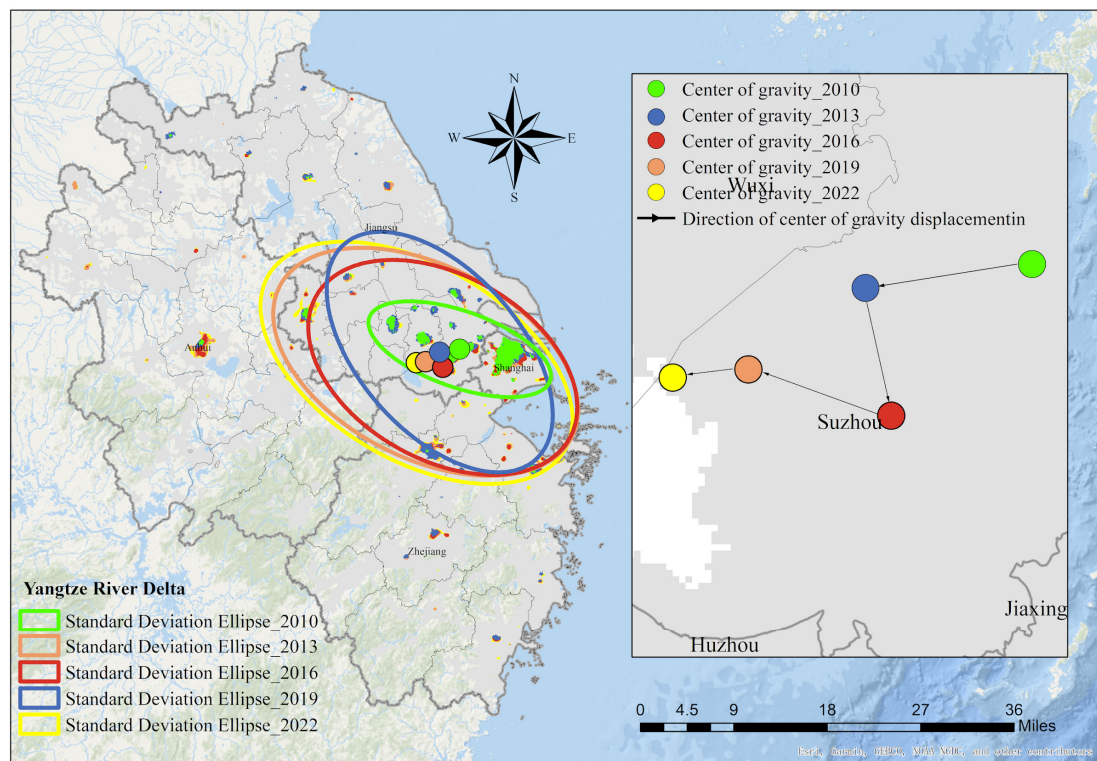
Stage-wise comparisons indicate that during 2010–2016 Zhejiang and Anhui exhibited higher rates of urban spatial expansion than in 2016–2022, whereas Jiangsu and Shanghai showed accelerated expansion in the later period. In 2010–2016 Zhejiang's growth

was driven primarily by industrial upgrading and new-town development: cities such as Hangzhou and Ningbo accelerated land development for information technology and high-tech industrial parks, with particularly pronounced expansion of new urban districts [39]. Concurrently, Anhui actively absorbed industrial transfers; industrial parks in Hefei and Wuhu attracted substantial population and investment, driving rapid spatial growth. During 2016–2022, spatial expansion accelerated in Jiangsu and Shanghai. Under the impetus of emerging technologies and advanced manufacturing, cities in Southern and Central Jiangsu intensified high-density new-district and industrial-park development, attracting capital and talent and promoting predominantly intensive spatial expansion [40]. Shanghai, through innovation-driven strategies, the construction of Lingang New Area and new zones in the free-trade area, and innovations in land-use policy, advanced high-quality urban spatial extension consistent with the post-industrial metropolitan pattern of “functional upgrading—high-density development—limited outward expansion.” Compared with other major Chinese urban agglomerations such as the Beijing-Tianjin-Hebei and the Pearl River Delta regions, the Yangtze River Delta exhibits a more balanced pace of spatial expansion and a more pronounced tendency toward spatial concentration.

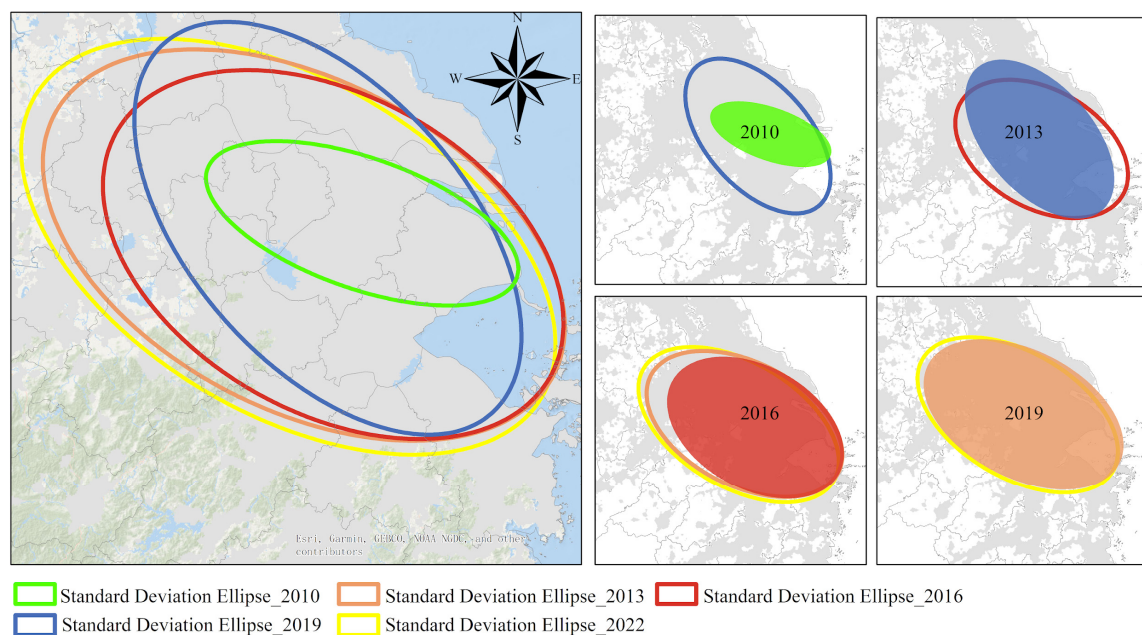
3.2.3. Spatial Morphology Evolution Trend of Urban Built-Up Areas in the Yangtze River Delta

Using ArcGIS 10.8, this study extracted centroids of urban land patches in the Yangtze River Delta for 2010–2022 and computed the annual built-up centroid coordinates, standard deviational ellipse centroids, and ellipse coverage areas to reveal centroid migration trajectories and aggregation characteristics of built-up areas. Through the analysis of spatial data, this section aims to elucidate the mechanisms underlying regional spatial evolution, namely “core agglomeration, corridor-led development, and intensive transformation.”

First, the spatial distribution characteristics of the built-up area centroid reveal a pronounced “core agglomeration” effect. Results indicate that during 2010–2022, the spatial centroid of built-up areas in the Yangtze River Delta (YRD) urban agglomeration consistently remained in the vicinity of Suzhou. This spatial phenomenon not only reflects the dominant role of the Southern Jiangsu region in regional urbanization but also fundamentally demonstrates the “core agglomeration” mechanism, which is characterized by the continuous convergence of population, capital, and industrial factors toward high-tier cities (Figure 9a). Suzhou’s favorable location—at the junction of Jiangsu, Zhejiang, and Shanghai—provides superior accessibility and transport connectivity relative to surrounding areas. Concurrently, the Southern Jiangsu cluster (e.g., Suzhou, Wuxi, Changzhou), together with Shanghai and Hangzhou, constitutes the regional core and forms a highly interconnected economic corridor that has driven substantial increases in regional urbanization [41]. By contrast, Northern Zhejiang cities such as Huzhou and Jiaxing have placed greater emphasis on ecological protection and service-sector development, subjecting land development to stricter constraints and thus resulting in relatively slower spatial expansion. Differences in industrial structure have further amplified the urbanization gap between Southern Jiangsu and Northern Zhejiang.



(a)



(b)

Figure 9. Migration of the centroid of urban built-up areas in the Yangtze River Delta. (a) spatial distribution of the center of gravity (2010, 2013, 2016, 2019, and 2022) and its displacement direction, with standard deviation ellipses; (b) overall standard deviation ellipses for 2010–2022.

Second, the orientation of the standard deviational ellipse (SDE) and the centroid trajectory jointly corroborate the “corridor-led development” effect of transportation and industrial belts on spatial expansion. The SDE analysis reveals that during 2010–2022, the spatial expansion of built-up areas in the Yangtze River Delta (YRD) generally exhibited a northwest–southeast orientation. The built-up area centroid shifted significantly toward

the southwest (approximately 40 km) between 2010 and 2016. Subsequently, the magnitude of this shift gradually converged during 2016–2022, indicating a stabilization in the center of urban expansion. The core economic belts of the YRD (e.g., Shanghai-Suzhou-Hangzhou and Ningbo-Zhoushan) have agglomerated high-tier industries, including manufacturing, technology, and port logistics. These industrial belts highly overlap spatially with high-speed rail and expressway networks, thereby forming robust development corridors. Driven by this “corridor-led” effect, urban expansion is no longer characterized by disorderly sprawl; rather, it undergoes dynamic adjustments strictly aligned with high-value-added industrial layouts and major transportation arteries. This trend exhibits high similarity to the spatial evolution of world-class urban agglomerations. The spatial collaboration within urban networks and the interconnectivity of transportation corridors provide the institutional and infrastructural guarantees for urban spatial agglomeration and diffusion.

Third, the contraction of the standard deviational ellipse (SDE) coverage area intuitively characterizes the “intensive transformation (intensification)” of regional land-use patterns. Data indicate that the SDE coverage area decreased substantially from $7.646 \times 10^8 \text{ m}^2$ in 2010 to $3.515 \times 10^8 \text{ m}^2$ in 2022, with an average annual decline of approximately 37.6 km^2 (Figure 9b). This significant reduction in area and the deceleration of the expansion rate do not imply developmental stagnation; rather, they serve as empirical evidence for a fundamental shift in the regional spatial expansion mode—transitioning entirely from early extensive outward expansion to an “intensive transformation” centered on enhancing land-use efficiency. The improvement of infrastructure and industrial systems in core cities has prompted the high-density agglomeration of the economy and population within limited spaces. Simultaneously, peripheral regions are subject to strict ecological protection policies, leading to a relative decline in development demand. This bidirectional interaction jointly drives the “intensive and connotative” development of built-up areas.

Finally, regarding the azimuth angle of the SDE, it gradually increased from 67.27° to 70.61° during 2010–2022, indicating that the primary direction of urban expansion in the YRD gradually shifted from north-by-east to the southeast. This directional deflection further supplements the explanation of the dynamic changes in the “corridor-led” effect: in recent years, as high-value-added industries agglomerate in southeastern coastal cities such as Shanghai, Hangzhou, and Ningbo, the construction of regional integration demonstration zones, free trade zones, and port economic circles has accelerated the extension of urban expansion toward the southeast. This suggests that future spatial planning must not only conform to the trends of economic restructuring and industrial upgrading but also focus on functional enhancement and spatial synergy in emerging growth poles such as the southeastern coast.

3.3. Analysis of the Driving Mechanism of Urbanization

3.3.1. Identifying the Driving Factors of Phased Urbanization

To characterize and compare the similarities and differences in the mechanisms influencing urbanization in the Yangtze River Delta (YRD) urban agglomeration across two stages, this study implemented a consistent feature selection and modeling process under unified candidate variables and parameter settings to ensure cross-period comparability. The Random Forest-Recursive Feature Elimination with Cross-Validation (RF-RFECV) method was employed to select a robust subset of variables from 13 potential influencing factors. The parameters of the Random Forest regressor were set as follows: $n\_estimators = 500$, $max\_depth = 6$, $min\_samples\_leaf = 5$, $min\_samples\_split = 10$, $n\_jobs = -1$, with a fixed random seed of $random\_state = 42$. A 10-fold KFold cross-validation was utilized ($shuffle = True$, $random\_state = 42$). The RFECV was configured with $step = 1$,

`min_features_to_select = 1`, and `scoring = "neg_mean_squared_error"`, determining the optimal number of features based on the highest average score from the 10-fold cross-validation. Given that the number of candidate features was 13 and `step = 1`, the RFECV evaluated 13 configurations ranging from 1 to 13 features, with each configuration undergoing a 10-fold cross-validation. Here, `neg_mean_squared_error` equals $-\text{MSE}$; a higher score indicates a smaller mean squared error and better model fit.

After obtaining the optimal feature subset, the SHAP method was further applied to interpret the importance and impact direction of key features. The TreeExplainer, specifically designed for tree-based models, was used to calculate the SHAP values for each sample, reporting the following three statistical indicators: (1) `mean_abs_shap`: the mean absolute SHAP value, reflecting global importance, where a larger value indicates a stronger overall impact of the feature; (2) `mean_shap`: the mean SHAP value, reflecting the overall impact direction (values greater than 0 indicate an overall positive push on the predicted value, while values less than 0 indicate an overall negative pull); and (3) `std_abs_shap`: the standard deviation of absolute SHAP values, measuring the heterogeneity of the impact intensity, where a larger value indicates more pronounced differences in the feature's impact across different samples.

Given that the sample consists of city-year panel data, we further assessed robustness of the conclusions, in addition to the random 10-fold KFold, this study further supplemented two more stringent validation strategies: (1) Leave-One-Year-Out (LOYO) validation: iteratively leaving out one year as the validation set and using the remaining years as the training set to test the temporal extrapolation capability for unseen years; and (2) GroupKFold by city: conducting cross-validation grouped by city to ensure that different years of the same city do not appear simultaneously in both the training and validation sets, thereby mitigating potential information leakage under the panel structure. The robustness test results indicate that core features were stably selected under all three validation strategies, demonstrating the strong robustness of the key influencing factors.

1. The driving mechanism of urbanization in 2010–2015

During Stage I, the model was trained and cross-validated on the 2010–2015 sample. Starting from 13 candidate predictors, RF-RFECV iteratively reduced the feature set and selected five optimal predictors; the cross-validation score was -80.116 (Figure 10a). The selected variables were: X2 (fiscal decentralization), X6 (total retail sales of consumer goods), X7 (transport accessibility), X8 (degree of openness), and X10 (population density) (Figure 10b). Global SHAP statistics indicate the feature importance ranking for 2010–2015 as: $X10 > X6 > X8 > X2 > X7$ (Figure 10c). Overall, early-stage urbanization was jointly driven by five factor categories summarized as “density–market–openness–institution–corridor”: population agglomeration X10 served as the primary threshold and the strongest explanatory factor; consumption vitality X6 provided robust domestic demand support; openness X8 brought external resources and opportunities; fiscal decentralization X2 improved public-goods provision and resource allocation efficiency; and transport accessibility (X7) was selected but had relatively lower importance.

Population density X10 was the primary driving factor of the urbanization level during 2010–2015. In a portion of the samples, an increase in density significantly “pushed up” the predicted values, reflecting the matching, sharing, and learning effects brought about by agglomeration economies. However, when the density exceeded a certain threshold, the SHAP contribution turned negative, and a “high-density penalty” emerged (Table 4 and Figure 10d). This aligns with the mechanism of agglomeration diseconomies emphasized in geography and urban economics, where rising marginal costs—such as congestion costs, escalating land and housing prices, environmental externalities, and the overstretching of public services—begin to offset or even reverse the agglomeration

dividends, resulting in a “pulling down” effect on the predictions. Consequently, the slightly negative average direction of X10 does not imply that “density is overall harmful”; rather, it indicates that a substantial proportion of cities in the sample may have crossed the “optimal density” interval. The marginal return of urbanization quality to density has shifted from positive to negative, necessitating a transition from “scaling up” to “density governance and carrying capacity enhancement [42]”.

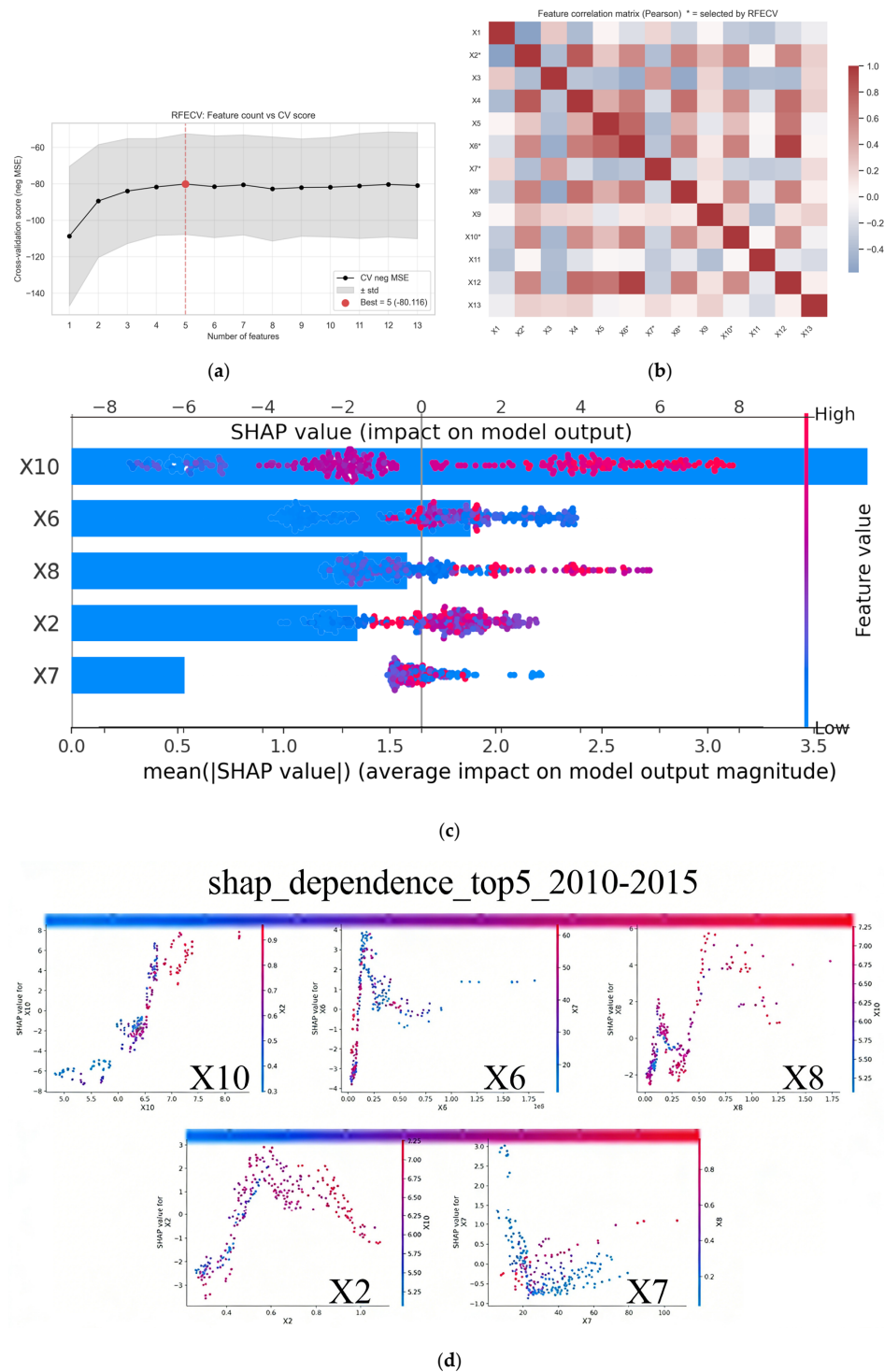


Figure 10. (a) Feature selection via RFECV: effect of the number of features on model performance (2010–2015). (b) Pearson correlation heatmap for Stage I (2010–2015) (* denotes features selected by RFECV). (c) SHAP summary plot for Stage I (2010–2015): feature importance and directional effects. (d) SHAP dependence plots for the top five drivers in Stage I (2010–2015).

Table 4. SHAP summary metrics for the top five core features (2010–2015).

Feature	mean_abs_shap	mean_shap	std_abs_shap
X10	3.753	−0.171	2.085
X6	1.880	0.118	1.221
X8	1.582	−0.006	1.220
X2	1.347	0.124	0.830
X7	0.532	−0.063	0.511

Consumption scale X6 (0.118) showed a stable and significant positive contribution (Table 4 and Figure 10d), consistent with the YRD’s service-oriented and consumption-upgrade dynamics: between 2010 and 2015, commercial services, cultural and tourism activities, conventions/exhibitions and digital consumption expanded rapidly—led by Shanghai and supported by Nanjing, Hangzhou, Ningbo, and Suzhou—driving employment absorption and functional upgrading and acting as a key endogenous engine of sustained urbanization.

The mean SHAP direction for openness X8 (−0.006) was close to zero (Figure 10d), reflecting a typical “double-edged sword” effect that aligns with spatial heterogeneity and cyclical fluctuations of the YRD’s export-oriented economic differentiation [43]. In port/processing-trade agglomerations, openness significantly raised urbanization levels; conversely, in localities lacking absorptive capacity in supporting industries or services, openness-driven flows failed to materialize into local development and were occasionally negative under external demand shocks, producing an overall mean near zero with high dispersion.

Fiscal decentralization X2 (0.124) had an overall positive effect (Figure 10d), indicating that decentralization and local governance incentives exerted a pronounced quality-improvement effect during this period. Transport accessibility X7 (−0.063) showed a weak and slightly negative average effect (Figure 10d), consistent with the YRD reality of “well-developed corridors but remaining structural deficiencies”: during this stage, marginal returns to traditional corridor expansion declined, inducing cross-city commuting spillovers and low-density diffusion, which exerted a modest negative impact on the composite urbanization score.

2. The driving mechanism of urbanization in 2016–2022

In Stage II (2016–2022), using the same RF-RFECV workflow as in Stage I, the model selected five key features: X4 (per-capita GDP), X6 (total retail sales of consumer goods), X8 (degree of openness), X10 (population density), and X12 (number of green patent grants) (Figure 11a,b). Global SHAP statistics rank feature importance as $X10 > X4 > X6 > X12 > X8$ (Figure 11c).

Population density X10 remains the primary threshold variable, exhibiting the largest |SHAP| and high heterogeneity, which indicates substantial variation in its impact on urbanization quality across cities (Table 5 and Figure 11d). Although the mean SHAP for X10 is negative, it has notably converged toward zero relative to Stage I, suggesting that the adverse marginal effects of high density (e.g., congestion, housing, and environmental costs) have been partially mitigated by policies and interventions such as integrated transit development, urban renewal/retrofit, expansion of public services, and station-city integration.

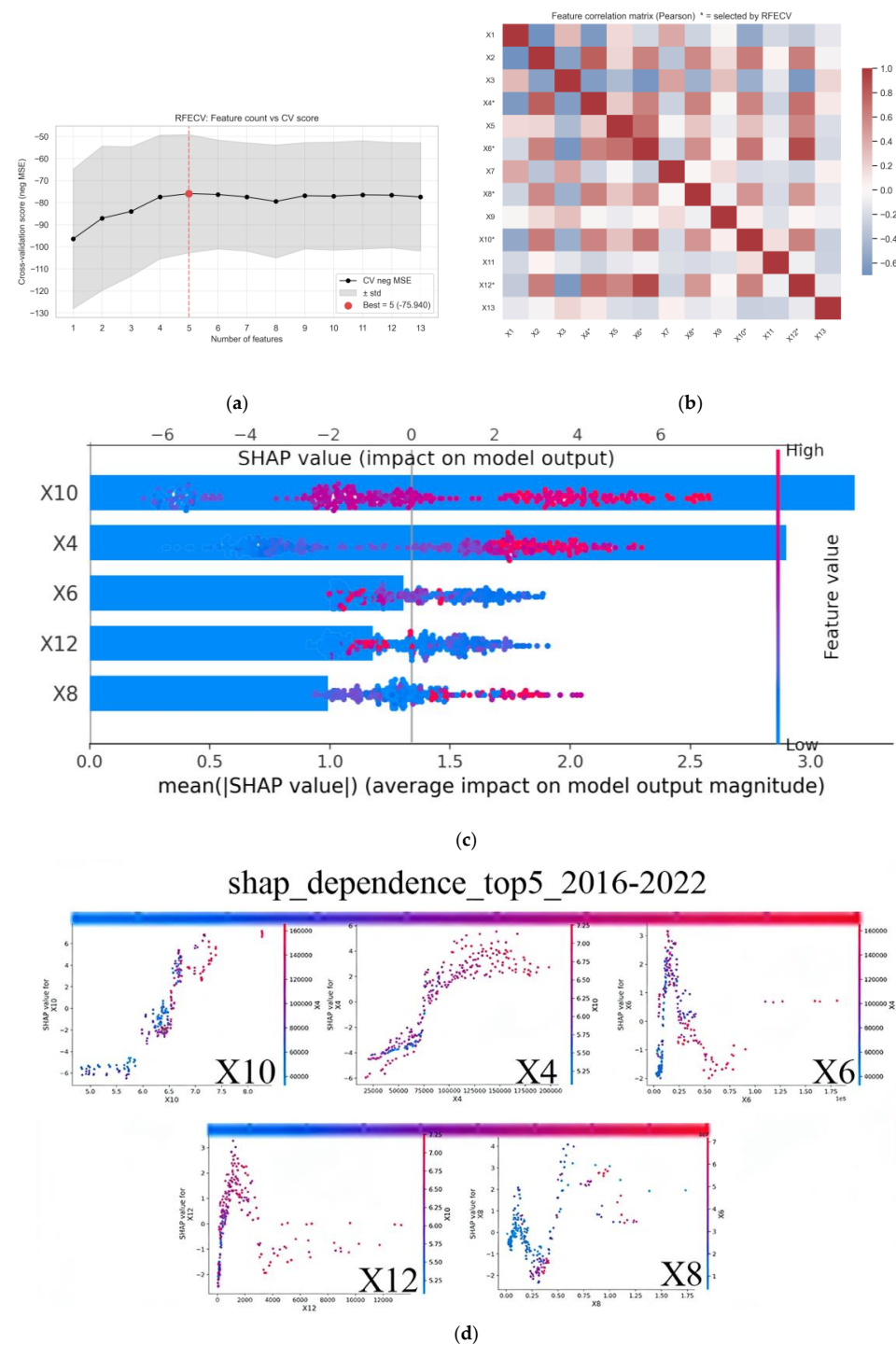


Figure 11. (a) Feature selection via RFECV: effect of the number of features on model performance (2016–2022). (b) Pearson correlation heatmap for Stage II (2016–2022) (* denotes features selected by RFECV). (c) SHAP summary plot for Stage II (2016–2022): feature importance and directional effects. (d) SHAP dependence plots for the top five drivers in Stage II (2016–2022).

Per-capita GDP (X4) rises to the second position in global importance, but its mean direction is slightly negative (mean_shap ≈ -0.073) (Figure 11d). This implies that, under rising factor costs and high-density constraints, increases in per-capita output do not automatically translate into synchronous improvements in urbanization quality. Without commensurate enhancements in public services, spatial organization, and living-environment quality, higher output can be accompanied by increased living costs and crowding externalities, producing a modest downward effect on composite urbanization scores [44].

Table 5. SHAP summary metrics for the top five core features (2016–2022).

Feature	mean_abs_shap	mean_shap	std_abs_shap
X10	3.185	−0.063	2.024
X4	2.898	−0.073	1.239
X6	1.305	0.083	0.695
X12	1.179	−0.004	0.735
X8	0.991	0.034	0.878

Consumption scale X6 (0.083) continues to provide a stable and clearly positive impetus and remains one of the most robust endogenous drivers across periods (Figure 11d). However, its global strength is lower than in Stage I, indicating a decline in marginal elasticity. This trend is consistent with factors such as online consumption substitution, shrinking daily activity circles, changes in commuting radii, and polycentric development that dilute traditional commercial-flow agglomeration effects.

Green patent grants X12 enter the mid-importance tier with an average SHAP value close to zero (−0.004), shown in Figure 11d, exhibiting a diffusion-stage pattern of “strong effects in leading cities, early-stage emergence across the broader region.” In cities with mature innovation platforms and complete industrial ecosystems, green technologies and digital transformation increasingly contribute to urbanization quality; yet in most cities, these effects remain nascent or suffer from incomplete transmission chains, yielding a flattened sample mean.

The global importance of openness X8 is comparatively weak; its mean effect is slightly positive but highly heterogeneous (Figure 11d). This finding indicates that the benefits of openness are increasingly contingent on institutionalized forms of openness and on local absorptive and supporting capacities.

3.3.2. Relative Impact Intensity and Intercorrelation Test of Driving Factors

1. Variable Correlation Structure

To examine the intercorrelations among driving factors and identify potential coupling structures, this study calculated the Spearman correlation coefficient matrices for X1–X13 using the samples from 2010–2015 and 2016–2022, respectively, and extracted highly correlated variable pairs for comparison. The results indicate that distinct “modular linkages” existed in both stages, with core elements such as fiscal decentralization, economic development, market vitality, degree of opening-up, and green innovation exhibiting significant positive covariation. In 2010–2015, the degree of fiscal decentralization X2 showed a strong positive correlation with per capita GDP (X4); per capita GDP (X4) was highly correlated with the degree of opening-up X8; and the correlation coefficient between total retail sales of consumer goods X6 and the total number of green patent grants X12 reached up to $\rho = 0.912$ (Figure 12a). In 2016–2022, the aforementioned linkage relationships remained robust. Furthermore, the correlations between population density X10 and green innovation X12, market vitality X6, and fiscal decentralization X2 were significantly strengthened, as shown in Figure 12b, indicating a more prominent “agglomeration-innovation-market” coupling in the second stage. Meanwhile, the degree of policy intervention X1 exhibited a significant negative correlation with key variables such as fiscal decentralization X2, per capita GDP X4, and green innovation X12 in both stages, suggesting a structural directional divergence between policy intervention and the elements of marketization, openness, and innovation.

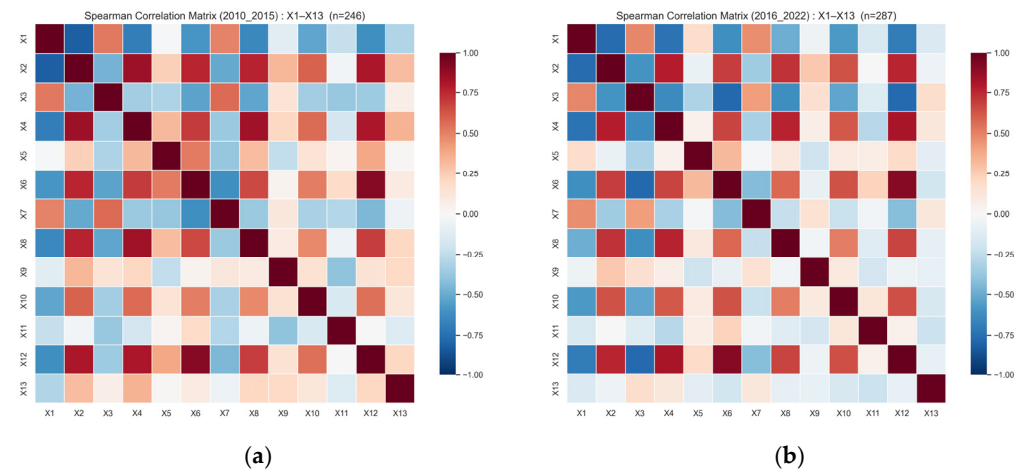


Figure 12. (a) Spearman correlation matrix of explanatory variables (X1–X13), 2010–2015; (b) Spearman correlation matrix of explanatory variables (X1–X13), 2016–2022.

2. Partial Correlation Analysis

To more rigorously examine the relatively independent associations between key driving factors and the urbanization level (Y), this study conducted partial correlation tests on the top five key features selected by RF-SHAP within the staged samples. Partial correlations were calculated using the “residual correlation” method; that is, linear regressions were performed for Y and the variable to be tested, respectively, given a set of control variables. The residuals were then extracted, and the correlation coefficient between the two sets of residuals was calculated. To prevent the conclusions from being influenced by a single control scheme, this study established two sets of control variables: Control Set A included per capita GDP (X4) and population density X10, while Control Set B added the degree of opening-up X8 and the total number of green patent grants X12 to Set A. When the variable being tested belonged to the control set, it was excluded from the control set before the calculation.

The results show that in the 2010–2015 stage, population density X10 exhibited a stable and significant positive partial correlation with the urbanization level (A: $r = 0.331$, $p = 1.09 \times 10^{-7}$; B: $r = 0.347$, $p = 2.81 \times 10^{-8}$). Traffic convenience (X7) showed a stable and significant negative partial correlation with the urbanization level (A: $r = -0.301$, $p = 1.70 \times 10^{-6}$; B: $r = -0.321$, $p = 3.21 \times 10^{-7}$). In contrast, the partial correlation coefficients for the degree of fiscal decentralization X2 and the degree of opening-up X8 were small under both control sets and failed to pass the significance test, indicating that their relationships with the urbanization level are more likely derived from their covariation with other development elements. The partial correlation of total retail sales of consumer goods X6 was relatively sensitive to the control scheme: it was almost zero and insignificant under Control Set A, but turned into a significant positive correlation under the expanded Control Set B. This suggests that the coupling of market vitality with elements such as openness and innovation in the first stage might affect its apparent association.

In the 2016–2022 stage, both population density X10 and per capita GDP (X4) maintained significant positive partial correlations under both control sets, and the correlation intensity further strengthened under the expanded Control Set B (X10: A $r = 0.374$, $p = 6.25 \times 10^{-11}$; B $r = 0.438$, $p = 9.33 \times 10^{-15}$; X4: A $r = 0.244$, $p = 3.02 \times 10^{-5}$; B $r = 0.335$, $p = 7.00 \times 10^{-9}$). This indicates a more robust independent association between the urbanization level and the “agglomeration-development” elements in the second stage. The total number of green patent grants (X12) exhibited a significant negative partial correlation under both control sets (A $r = -0.299$, $p = 2.81 \times 10^{-7}$; B $r = -0.300$, $p = 2.69 \times 10^{-7}$),

showing an independent inverse relationship with the urbanization level after controlling for development level, agglomeration, and openness factors. The partial correlation of the degree of opening-up X8 was insignificant under both control sets, while total retail sales of consumer goods X6 was significantly negative under Control Set A and insignificant (approaching zero) under Control Set B. This demonstrates that its independent association is similarly sensitive to the control of openness and innovation elements.

4. Discussion

4.1. Stage-Wise Evolution and Comparative Differences in the Driving Mechanisms

A comparative synthesis of the two periods indicates that the urbanization dynamics in the YRD exhibit a typical pattern of “threshold constraints–structural reconfiguration–spatial differentiation.” The main findings can be summarized as follows. (1) The density threshold persists across both periods. Feature selection results for both stages consistently identify population density X10 as the primary threshold variable, suggesting that population agglomeration remains a fundamental constraint and a key trigger for advancing urbanization in the YRD; however, its potential negative externalities need to be mitigated and converted through transportation provision, public service equalization, and refined spatial governance. (2) The driving mechanism shifts from “expansion” to “quality upgrading.” During 2010–2015, an expansion-oriented chain—“density–market development–external openness–fiscal decentralization–transport accessibility”—was observed, highlighting the dominant role of Shanghai-centered port–trade linkages and local fiscal incentives in the scale-expansion phase. In 2016–2022, the mechanism evolved into “density–per capita output–consumption–green innovation–external openness,” indicating a transition from scale-driven growth to a quality-enhancing trajectory powered by industrial upgrading, innovation, and green transformation. (3) Key variables display stage-specific differentiation. Social consumption X6 continues to support urbanization in both periods, yet its marginal effect declines in more mature areas. The explanatory power of per capita GDP (X4) rises markedly in the later period, implying that shifts in industrial structure and factor prices exert a more pronounced influence on living costs and spatial organization. Green patents X12 begin to take effect, though the overall pattern remains one of “earlier manifestation in core cities.” (4) Spatial heterogeneity conditions the effectiveness of openness and innovation. The effects of population density vary substantially by city type: in core metropolitan areas, the adverse effects of high density are partially alleviated through rail transit expansion and inner-city redevelopment, whereas some satellite and resource-based cities face a capacity bottleneck associated with insufficient density. Moreover, the positive effect of external openness X8 is highly contingent on local industrial absorption capacity and pathways toward high-quality openness, resulting in pronounced intra-regional heterogeneity in effect magnitudes.

4.2. Policy Implications for High-Quality Urbanization in the Yangtze River Delta

(1) Substituting “Scale Expansion” with “Density Governance”: Enhancing Carrying Efficiency and Livability under Threshold Constraints. As demonstrated by the aforementioned non-linear model, although population density acts as the strongest explanatory factor, it exhibits significant threshold constraints and a “high-density penalty,” confirming the existence of a “scale-density paradox.” Once urban density exceeds a specific threshold, “agglomeration diseconomies”—such as congestion costs, environmental externalities, and the over-extraction of public services—will gradually offset the demographic dividends of scale. Therefore, urbanization policies should shift from “expanding scale” to “density governance,” focusing on improving the carrying efficiency per unit of density under given constraints. For high-density core areas such as Shanghai, Suzhou, and Hangzhou,

a dynamic carrying capacity evaluation system should be established, centered on commuting time, public service load (e.g., school quotas and hospital beds), ecological capacity, and housing affordability, to accurately identify “high-density and high-efficiency” versus “high-density and high-pressure” spaces. On the one hand, utilizing Transit-Oriented Development (TOD) and station-city integration as key levers, public services like education and healthcare should be spatially reallocated in tandem with population and industry to mitigate congestion externalities. On the other hand, urban renewal of existing stocks should replace outward expansion, phasing out inefficient land use to enhance the composite output efficiency of urban spaces. For receiving cities such as Nantong, Yancheng, and Wuhu, policy priorities should focus on increasing employment density and residential attractiveness, strictly avoiding inefficient land use and local debt risks caused by blind low-density expansion. Concurrently, through the construction of cross-city commuting circles, these cities should deeply embed themselves into the functional networks of core metropolitan areas to enhance the stability of population and industrial relocation.

(2) Driving Urbanization Quality through “Output Quality”: Promoting the Coupling and Transformation of “Output-Innovation-Green” at the Metropolitan Area Level. Empirical results indicate that the driving trajectory of urbanization has evolved from an early “expansionary chain” to a “quality-enhancing network” supported by multiple factors. Notably, if innovation indicators, such as green patent authorizations, are detached from local industrial absorption, they can easily translate into high environmental compliance costs and trigger a “crowding-out effect.” Consequently, policy design must not simply equate “innovation input” with quality improvement; rather, it must construct a coupling mechanism of “industrial upgrading-consumption scenarios-green technology diffusion” taking the metropolitan area as the fundamental unit. Regarding the construction of a cross-regional collaborative transformation network, cross-city green industry demonstration belts should be established relying on core nodes like Shanghai, Suzhou, and Hangzhou. Public services, such as green technology pilot platforms, carbon accounting, and standard certification should be forward-looking deployed along the Shanghai-Nanjing and Shanghai-Hangzhou corridors to reduce the institutional costs of transitioning green technologies from “patents” to “industrialization.” In implementing “chain-style relocation” and skill matching, receiving cities should target segments with scalable diffusion and upward mobility potential in the value chain when deploying industrial supporting facilities, avoiding the “low-end lock-in” of simple capacity transfers. Simultaneously, the matching degree between labor skill supply and industrial upgrading must be strengthened to ensure that the “enhancement of output quality” effectively translates into continuous improvements in local urbanization quality.

(3) Implementing Differentiated and Convertible Integrated Synergy: Shifting Openness and Innovation from “Hub Agglomeration” to “Regional Diffusion”. The study finds that after stripping away systemic coupling effects, the independent correlation of opening-up and innovation elements weakens, highlighting the “network embeddedness” characteristic of modern urbanization drivers. Openness lacking local market anchoring is highly prone to triggering a “siphon effect” of factor loss. Therefore, the advancement of YRD integration should not merely stop at “improving openness indicators,” but should focus on convertible mechanisms to form a regional diffusion pathway of “rule supply-carrying capacity-public goods synergy.” Hub cities (e.g., Shanghai) should strengthen “rule spillovers” by focusing on institutional openness and high-end factor allocation, forming a replicable and exportable standard system for cross-border services, sci-tech innovation services, and green compliance. Through regional collaboration mechanisms, “hub capabilities” can be transformed into “regional dividends.” Node cities, on the other hand, should focus on “carrying capacity construction.” Most prefecture-level cities should

prioritize addressing local relocation shortcomings, forming a project-based promotion list around industrial supporting facilities, park governance, business environments, and innovation services. Enhancing local “stickiness” for high-quality foreign investment and producer services will break the dilemma of factors “flowing through but failing to settle,” fundamentally preventing the risk of core-periphery polarization.

5. Conclusions

This study systematically analyzed the spatiotemporal evolution of urbanization and the expansion dynamics of built-up areas in the YRD for 2010–2022. The principal findings are summarized as follows.

(1) Temporal dynamics and inter-provincial divergence. Overall urbanization in the YRD increased continuously during 2010–2022. Shanghai maintained a leading high level, Jiangsu and Zhejiang advanced steadily, and Anhui achieved a rapid catch-up under regional integration policies, significantly narrowing its gap with the lead provinces. Nevertheless, a “Matthew effect” persists at the city scale: despite accelerated growth in some small and medium cities and interior areas, absolute gaps remain substantial, posing an ongoing challenge to regional balanced development.

(2) Spatial pattern and multi-centering. The spatial pattern exhibits an “east-high, west-low; river/coast higher than inland” gradient. With deepening regional integration, the spatial structure has evolved from an early core–belt distribution toward a multi-centric, networked urban-agglomeration system. Metropolitan circles and economic belts have become the principal carriers of urbanization, promoting a more balanced regional spatial development.

(3) Built-up area evolution: agglomeration, corridor guidance, and intensification. From 2010 to 2022, the spatial evolution of built-up areas in the YRD displayed a clear trend of “core agglomeration, corridor guidance, and intensive transformation.” Spatially, the centroid of built-up area remained concentrated in the Southern Jiangsu core (near Suzhou), while the principal expansion axis dynamically shifted toward the southeast coastal economic belt, reflecting the strong steering role of high-value industry location and transport corridors. In terms of development mode, the marked contraction of the standard-deviation ellipse area and the deceleration of expansion jointly confirm a shift from extensive peripheral growth to an “intensive, land-use-efficiency centered” model. These changes underscore the central role of industrial upgrading and economic restructuring and demonstrate the effectiveness of ecological regulation and land-use policy.

(4) Driving mechanisms: from scale to quality and innovation. Combined RF-RFECV and SHAP analysis indicates that the early phase was characterized by extensive expansion, where population density served as the primary threshold, synergized by consumption and openness, and supported by fiscal decentralization and transportation infrastructure. In the later phase, this evolved into an intensive enhancement driven by per capita output and green innovation, while population density maintained its paramount position. Throughout this transition, the foundational role of consumption remained robust, albeit with a diminishing marginal effect, whereas the impact of opening-up became increasingly contingent upon the institutional environment and local industrial absorption capacity.

This study identifies driving mechanisms using panel and remote-sensing data for 2010–2022 combined with interpretable machine-learning methods, but several limitations remain. The selected time window and variable set may not fully capture heterogeneous short-term shocks (e.g., the COVID-19 pandemic); causal inference requires quasi-experimental or natural-experiment designs for further policy validation. Moreover, the micro-mechanisms of endogenous dynamics and institutional change in small and medium cities warrant further investigation through fieldwork and case studies. Future

research should extend the temporal span, incorporate finer-grained flow and institutional data, and evaluate the implementation pathways and real-time effects of the proposed measures to provide more granular decision support for high-quality regional integration in the YRD.

Author Contributions: Conceptualization, D.S. and K.Z.; Methodology, D.S. and K.Z.; Software, D.S.; Validation, D.S.; Formal analysis, D.S.; Investigation, D.S.; Resources, D.S.; Data curation, D.S.; Writing—original draft, D.S.; Writing—review & editing, D.S. and K.Z.; Visualization, D.S. and K.Z.; Supervision, D.S., K.Z. and H.L.; Project administration, D.S., K.Z. and H.L.; Funding acquisition, D.S. and K.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available on request from the author.

Acknowledgments: The authors would like to thank colleagues on the Working in the Interdisciplinary Program in Landscape Architecture for their support and comments that helped shape this paper. All responsibility for the final version though lies with the authors.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Table A1. Urbanization level indicator measurement system (UDI).

Indicator Category	Indicator Layer	Indicator Meaning	Indicator Code	Unit	Calculation Method	Type	Weight
Population urbanization	Human Resources Structure	Working-age population participation rate [45]	U1_1	(%)	Economically active population ratio	+	0.022
	Employment Methods	Employment concentration in manufacturing [46]	U1_2	(%)	Proportion of employees in the secondary industry	+	0.019
		Employment concentration in modern service industry [47]	U1_3	(%)	Proportion of employees in the tertiary industry	+	0.017
	Human capital accumulation	Higher education population penetration rate [48]	U1_4	(%)	Number of students in regular undergraduate and junior college programs (persons)/total population at the end of the year (persons)	+	0.095
Economic urbanization	Productivity level	Total labor productivity [49]	U2_1	Yuan	GDP per capita	+	0.031
		Urban residents' consumption upgrade coefficient [20]	U2_2	(%)	Engel coefficient of urban residents	+	0.024
	Urban-rural coordination	Urban-rural income balance index [50]	U2_3	(%)	Ratio of per capita disposable income of urban and rural residents	+	0.024
		Consumption Convergence Index [51]	U2_4	(%)	Ratio of per capita consumption level of urban and rural residents	+	0.008
	Industry support	Industrial structure upgrading index [28]	U2_5	/	1 × the proportion of the primary industry + 2 × the proportion of the secondary industry + 3 × the proportion of the tertiary industry	+	0.013
Spatial urbanization	Urban land	Scale of urban construction land [52]	U3_1	km ²	/	+	0.108
		Urbanization rate [53]	U3_2	(%)	/	+	0.011
	Living Environment	Urban residential land size [54]	U3_3	km ²	Residential area	+	0.103
		Urban road facilities level [55]	U3_4	m ²	Urban road area per capita	+	0.002
		Urban housing market activity [55]	U3_5	10 km ²	Residential sales area	+	0.013
		Supply scale of affordable housing in urban areas [55]	U3_6	m ²	Affordable housing supply area	+	0.091

Table A1. Cont.

Indicator Category	Indicator Layer	Indicator Meaning	Indicator Code	Unit	Calculation Method	Type	Weight
Social urbanization	Healthcare	Level of allocation of medical and health resources [56]	U4_1	/	Number of medical institution beds per thousand people	+	0.054
	Cultural atmosphere	Level of public cultural service supply [57]	U4_2	/	Number of books in public libraries per 100 people	+	0.014
	Public Education	Coverage level of basic education facilities [58]	U4_3	/	Number of primary and secondary schools per 10,000 people	+	0.037
	Transportation Facilities	Public transport service capacity [59]	U4_4	/	Buses per 10,000 people	+	0.230
	Employment Environment	Stable level of urban employment [59]	U4_5	(%)	Urban registered employment rate	+	0.026
	Communication level	Level of information development [60]	U4_6	/	Number of mobile phone users at the end of the year	+	0.056

References

1. Chaolin, G.U.; Liya, W.U.; Cook, I. Progress in research on Chinese urbanization. *Front. Archit. Res.* **2012**, *1*, 101–149. [\[CrossRef\]](#)
2. Wong, C.; Zheng, W. When East meets West: The prospects of Chinese urbanisation research. *Trans. Plan. Urban Res.* **2022**, *1*, 7–16. [\[CrossRef\]](#)
3. Chen, M.; Liu, W.; Tao, X. Evolution and assessment on China's urbanization 1960–2010: Under-urbanization or over-urbanization? *Habitat Int.* **2013**, *38*, 25–33. [\[CrossRef\]](#)
4. Gu, C.; Hu, L.; Zhang, X.; Wang, X.; Guo, J. Climate change and urbanization in the Yangtze River Delta. *Habitat Int.* **2011**, *35*, 544–552. [\[CrossRef\]](#)
5. Hangli, Z.; Xinnan, A. Research on the coordinated development between land urbanization and population urbanization in Shaanxi Province, China. *Sci. Rep.* **2024**, *14*, 7977. [\[CrossRef\]](#)
6. Sui, Y.; Hu, J.; Zhang, N.; Ma, F. Exploring the dynamic equilibrium relationship between urbanization and ecological environment—A case study of Shandong Province, China. *Ecol. Indic.* **2024**, *158*, 111456. [\[CrossRef\]](#)
7. Jin, Z.; Wang, C.; Jiao, X.; Yu, S.; Yang, C.; Xie, F.; Miao, Y. Spatiotemporal pattern and influencing factors of urbanization quality in county areas of Shandong Province, China. *Ecol. Indic.* **2024**, *163*, 112132. [\[CrossRef\]](#)
8. Hu, S.; Wang, Y. Spatial pattern of rural authenticity and its relation to urbanization: Insights from Henan Province, China. *Habitat Int.* **2025**, *156*, 103291. [\[CrossRef\]](#)
9. Shangguan, Z. Exploring the impact of population urbanization on the green economy development: A case study of 30 provincial-level administrative regions in China. *Phys. Chem. Earth Parts A/B/C* **2024**, *136*, 103727. [\[CrossRef\]](#)
10. Zhang, Q.; Kong, Q.; Zhang, M.; Huang, H. New-type urbanization and ecological well-being performance: A coupling coordination analysis in the middle reaches of the Yangtze River urban agglomerations, China. *Ecol. Indic.* **2024**, *159*, 111678. [\[CrossRef\]](#)
11. Singh, S.; Shukla, A.; Jain, K. Assessing the urbanization-induced impact on environmental parameters of a city from a remote-sensing perspective. *Remote Sens. Appl. Soc. Environ.* **2024**, *34*, 101169. [\[CrossRef\]](#)
12. Shakibul Islam, K.; Wu, Q.; Islam, M.R.; Abdullah, H.M. Unveiling the Drivers of Unplanned Urbanization: A High-Resolution Night Light Development Index Approach for Assessing Regional Inequality and Urban Growth in Dhaka. *Remote Sens.* **2025**, *17*, 1397. [\[CrossRef\]](#)
13. Li, Y.; Ye, H.; Gao, X.; Sun, D.; Li, Z.; Zhang, N.; Leng, X.; Meng, D.; Zheng, J. Spatiotemporal patterns of urbanization in the three most developed urban agglomerations in China based on continuous nighttime light data (2000–2018). *Remote Sens.* **2022**, *13*, 2245. [\[CrossRef\]](#)
14. Zhou, Y.; Tu, M.; Wang, S.; Liu, W. A novel approach for identifying urban built-up area boundaries using high-resolution remote-sensing data based on the scale effect. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 135. [\[CrossRef\]](#)
15. Kaplan, G. Evaluating the roles of green and built-up areas in reducing a surface urban heat island using remote sensing data. *Urban. Izziv* **2019**, *30*, 105–112. [\[CrossRef\]](#)
16. Faisal, K.; Shaker, A.; Habbani, S. Modelling the relationship between the gross domestic product and built-up area using remote sensing and GIS data: A case study of seven major cities in Canada. *ISPRS Int. J. Geo-Inf.* **2016**, *5*, 23. [\[CrossRef\]](#)
17. Weng, Q. Modeling urban growth effects on surface runoff with the integration of remote sensing and GIS. *Environ. Manag.* **2001**, *28*, 737–748. [\[CrossRef\]](#)
18. Haas, J.; Ban, Y. Urban growth and environmental impacts in jing-jin-ji, the yangtze, river delta and the pearl river delta. *Int. J. Appl. Earth Obs. Geoinf.* **2014**, *30*, 42–55. [\[CrossRef\]](#)

19. Wu, Y.; Shi, K.; Chen, Z.; Liu, S.; Chang, Z. Developing improved time-series DMSP-OLS-like data (1992–2019) in China by integrating DMSP-OLS and SNPP-VIIRS. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 4407714. [[CrossRef](#)]
20. Zheng, W.; Walsh, P.P. Economic growth, urbanization and energy consumption—A provincial level analysis of China. *Energy Econ.* **2019**, *80*, 153–162. [[CrossRef](#)]
21. Xie, Y.; Weng, Q.; Fu, P. Temporal variations of artificial nighttime lights and their implications for urbanization in the conterminous United States, 2013–2017. *Remote Sens. Environ.* **2019**, *225*, 160–174. [[CrossRef](#)]
22. Ma, T.; Zhou, C.; Pei, T.; Haynie, S.; Fan, J. Quantitative estimation of urbanization dynamics using time series of DMSP/OLS nighttime light data: A comparative case study from China's cities. *Remote Sens. Environ.* **2012**, *124*, 99–107. [[CrossRef](#)]
23. Ma, T.; Zhou, Y.; Zhou, C.; Haynie, S.; Pei, T.; Xu, T. Night-time light derived estimation of spatio-temporal characteristics of urbanization dynamics using DMSP/OLS satellite data. *Remote Sens. Environ.* **2015**, *158*, 453–464. [[CrossRef](#)]
24. Zhuo, L.; Shi, P.; Chen, J. Application of compound night light index derived from DMSP/OLS data to urbanization analysis in China in the 1990s. *ACTA Geogr. Sin. Ed.* **2003**, *58*, 893902.
25. Zhang, X.; Fan, H.; Liu, F.; Lv, T.; Sun, L.; Li, Z.; Shang, W.; Xu, G. Coupling coordination between the ecological environment and urbanization in the middle reaches of the Yangtze River urban agglomeration. *Urban Clim.* **2023**, *52*, 101698. [[CrossRef](#)]
26. Zhang, K.H.; Song, S. Rural–urban migration and urbanization in China: Evidence from time-series and cross-section analyses. *China Econ. Rev.* **2003**, *14*, 386–400. [[CrossRef](#)]
27. Zhao, Z.; Pan, Y.; Zhu, J.; Wu, J.; Zhu, R. The impact of urbanization on the delivery of public service–related SDGs in China. *Sustain. Cities Soc.* **2022**, *80*, 103776. [[CrossRef](#)]
28. Xu, H.; Jiao, M. City size, industrial structure and urbanization quality—A case study of the Yangtze River Delta urban agglomeration in China. *Land Use Policy* **2021**, *111*, 105735. [[CrossRef](#)]
29. Guan, X.; Wei, H.; Lu, S.; Dai, Q.; Su, H. Assessment on the urbanization strategy in China: Achievements, challenges and reflections. *Habitat Int.* **2018**, *71*, 97–109. [[CrossRef](#)]
30. Basu, T.; Das, A.; Das, K.; Pereira, P. Urban expansion induced loss of natural vegetation cover and ecosystem service values: A scenario-based study in the siliguri municipal corporation (Gateway of North-East India). *Land Use Policy* **2023**, *132*, 106838. [[CrossRef](#)]
31. Chen, J.; Wei, H.; Li, N.; Chen, S.; Qu, W.; Zhang, Y. Exploring the spatial-temporal dynamics of the Yangtze River Delta urban agglomeration based on night-time light remote sensing technology. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 5369–5383. [[CrossRef](#)]
32. Jian, P.; Yanxu, L.; Hong, S.; Pan, X.; Xiaoxu, H.U.; Yanglin, W. Using impervious surfaces to detect urban expansion in Beijing of China in 2000s. *Chin. Geogr. Sci.* **2016**, *26*, 229243. [[CrossRef](#)]
33. Wang, S.; Tian, M.; Ding, Q.; Shao, H.; Xia, S. Study on coupling coordination degree of urbanization and ecological environment in Chengdu-Chongqing economic circle from 2002 to 2018. *Environ. Sci. Pollut. Res.* **2024**, *31*, 3134–3151. [[CrossRef](#)]
34. Zhu, H.; Ou, X.; Yang, Z.; Yang, Y.; Ren, H.; Tang, L. Spatiotemporal dynamics and driving forces of land urbanization in the Yangtze River Delta urban agglomeration. *Land* **2022**, *11*, 1365. [[CrossRef](#)]
35. Guan, C.; Wang, Y.; Keith, M.; Li, Y.; Cao, G. Does local planning of fast-growing medium-sized towns lead to higher urban intensity or to sprawl? Cases from Zhejiang Province. *Cities* **2022**, *130*, 103869. [[CrossRef](#)]
36. Qin, X.; Wei, Y.D.; Wu, Y.; Huang, X. Regional development and inequality within city regions: A study of the Yangtze River delta, China. *Geogr. Rev.* **2023**, *113*, 359–385. [[CrossRef](#)]
37. Fang, C.; Zhao, S. A comparative study of spatiotemporal patterns of urban expansion in six major cities of the Yangtze River Delta from 1980 to 2015. *Ecosyst. Health Sustain.* **2018**, *4*, 95114. [[CrossRef](#)]
38. Solecki, W.D.; Leichenko, R.M. Urbanization and the metropolitan environment: Lessons from New York and Shanghai. *Environ. Sci. Policy Sustain. Dev.* **2006**, *48*, 8–23. [[CrossRef](#)]
39. Yu, Z.; Chen, L.; Li, L.; Zhang, T.; Yuan, L.; Liu, R.; Wang, Z.; Zang, J.; Shi, S. Spatiotemporal characterization of the urban expansion patterns in the Yangtze River Delta region. *Remote Sens.* **2022**, *13*, 4484. [[CrossRef](#)]
40. Liu, Y.; Zhu, J. Spatial restructuring and population shrinkage along with industrialization in counties: The case of Jiangsu Province. *Habitat Int.* **2023**, *135*, 102785. [[CrossRef](#)]
41. Chen, Y.; Deng, A. Evolution and influencing factors of urban built-up areas in the Yangtze River Delta urban agglomeration. *IEEE Access* **2023**, *11*, 133413–133427. [[CrossRef](#)]
42. Yang, Q.; Ding, L.; Wang, L.; Liu, C.; Fan, Y.; Li, Y.; Wang, Y. Influence mechanism of new-type urbanization on urban land use efficiency in the Yangtze River Delta, China. *Chin. Geogr. Sci.* **2023**, *33*, 474–488. [[CrossRef](#)]
43. Guo, J.; Yu, Z.; Ma, Z.; Xu, D.; Cao, S. What factors have driven urbanization in China? *Environ. Dev. Sustain.* **2022**, *24*, 6508–6526. [[CrossRef](#)]
44. Yin, S.G.; Yang, S.; Li, Z.J. Spatial pattern and influencing factors of ecological urbanization in the Yangtze River Delta. *J. Nat. Resour.* **2022**, *37*, 1494–1506. [[CrossRef](#)]

45. Sun, M.; Wang, J.; He, K. Analysis on the urban land resources carrying capacity during urbanization—A case study of Chinese Yangtze River Delta. *Appl. Geogr.* **2020**, *116*, 102170. [[CrossRef](#)]
46. Kang, J.; Xu, W.; Yu, L.; Ning, Y. Localization, urbanization and globalization: Dynamic manufacturing specialization in the Yangtze River Delta mega-city conglomeration. *Cities* **2020**, *99*, 102641. [[CrossRef](#)]
47. Xi, Q.; Ji, P. Does the development zone promote population urbanization? Evidence from China. *Land Use Policy* **2023**, *131*, 106736. [[CrossRef](#)]
48. Storper, M.; Scott, A.J. Rethinking human capital, creativity and urban growth. *J. Econ. Geogr.* **2009**, *9*, 147–167. [[CrossRef](#)]
49. Mendez, P.; Atienza, M.; Modrego, F. Urbanization and productivity at a global level: New empirical evidence for the services sector. *Reg. Sci. Policy Pract.* **2023**, *15*, 1981–1998. [[CrossRef](#)]
50. Yang, Y.; Bao, W.; Wang, Y.; Liu, Y. Measurement of urban-rural integration level and its spatial differentiation in China in the new century. *Habitat Int.* **2021**, *117*, 102420. [[CrossRef](#)]
51. Liu, M.; Li, Q.; Bai, Y.; Fang, C. A novel framework to evaluate urban-rural coordinated development: A case study in Shanxi Province, China. *Habitat Int.* **2024**, *144*, 103013. [[CrossRef](#)]
52. Lin, X.; Wang, Y.; Wang, S.; Wang, D. Spatial differences and driving forces of land urbanization in China. *J. Geogr. Sci.* **2015**, *25*, 545–558. [[CrossRef](#)]
53. Zhou, X.; Wang, Y.C. Spatial-temporal dynamics of urban green space in response to rapid urbanization and greening policies. *Landsc. Urban Plan.* **2011**, *100*, 268–277. [[CrossRef](#)]
54. Ma, R.; Wang, T.; Zhang, W.; Yu, J.; Wang, D.; Chen, L.; Jiang, Y.; Feng, G. Overview and progress of Chinese geographical human settlement research. *J. Geogr. Sci.* **2016**, *26*, 1159–1175. [[CrossRef](#)]
55. Ye, Y.; Liu, Y.; Liu, J.; Weng, Z.; Chang, Z.; Li, X. Driving Factors of Urban–Rural Human Settlement Changes and a Multi-Scenario Simulation Based on SDGs. *Ecosyst. Health Sustain.* **2025**, *11*, 0300. [[CrossRef](#)]
56. Vlahov, D.; Galea, S. Urbanization, urbanicity, and health. *J. Urban Health* **2002**, *79*, S1–S12. [[CrossRef](#)] [[PubMed](#)]
57. Fan, P.; Ouyang, Z.; Nguyen, D.D.; Nguyen, T.T.H.; Park, H.; Chen, J. Urbanization, economic development, environmental and social Changes in transitional economies: Vietnam after Doimoi. *Landsc. Urban Plan.* **2019**, *187*, 145–155. [[CrossRef](#)]
58. Mckenzie, K. Urbanization, social capital and mental health. *Glob. Soc. Policy* **2008**, *8*, 359–377. [[CrossRef](#)]
59. Hussain, M.; Imitiyaz, I. Urbanization concepts, dimensions and factors. *Int. J. Recent Sci. Res.* **2018**, *9*, 23513–23523.
60. Li, B.; Chen, C.; Hu, B. Governing urbanization and the new urbanization plan in China. *Environ. Urban.* **2016**, *28*, 515–534. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.