Integrating Multisource Geographic Big Data to Delineate Urban Growth Boundary: A Case Study of Changsha

Xing Gao, Nan Xia, Sudan Zhuang, Xin Zhao, Jiale Liang, Ziyu Wang, and Manchun Li

***Abstract*—An urban growth boundary (UGB) is an important policy tool used to control urban sprawl, which can effectively balance the urban construction needs, residents' quality of life, and urban ecological protection. Current studies of UGB delineation and its indicators have paid little attention to human factors, such as human activities and economic vitality, and weights for evaluation indicators have been determined highly subjectively. In response to these problems, this article integrated multisource geographic Big Data to construct a total of 30 natural, human, and ecological evaluation indicators. The GIS technology and machine-learning (ML) approach were combined to determine indicator weights with an officially manually drawn 2035 UGB as the reference, aiming to reduce the subjectivity. The suitability score was then calculated from indicator grading 4–1 and related weights, and high suitability (>2.15) regions were eventually delineated as the UGB. Results showed a delineated UGB of 1528.06 km^2 with an overall accuracy of over 93% and high consistence with reference data for 2035 in the Chinese city of Changsha. The geographic Big Data totally contributed more than 33.72%, which mainly characterized role of human elements, and a 5–6 percentage point reduction in accuracy was found without these data. Compared with existing articles, our delineated UGB had higher accuracy and closer spatial pattern to the reference data, verifying the effectiveness and reasonability of ML-based weight setting approach and index system with geographic Big Data. The proposed method can provide scientific and accurate framework for UGB delineation, which can promote the territorial spatial planning and sustainable urban development.**

***Index Terms*—Geographic Big Data, GIS spatial analysis, random forest (RF), territorial spatial planning, urban growth boundary (UGB).**

I. Introduction

IN RECENT years, China has been undergoing rapid urbanization, which has given rise to a series of problems, including urban sprawl, environmental degradation, and severely imbalanced land resources [1]. These problems threaten ecological security and the human living environment. In response, in 2006, China proposed developing the urban growth boundary (UGB) as an important policy tool to prevent unlimited urban sprawl, protect open spaces on the edges of cities, and encourage a compact urban layout [2], [3]. In 2019, during a new round of territorial spatial planning, the definition of UGB was clarified as a spatial control boundary of an area with primarily urban functions, such as a city, town, or development zone, where urban development and construction can be concentrated for a certain period of time [4]. The delineation of UGB is the basis of territorial spatial planning in China, and setting this kind of maximum boundary for urban expansion should consider multiple factors, such as natural conditions, agricultural production, environmental capacity, and economic development [4], [5]. By delineating UGB, the unlimited urban expansion and the disorderly development can be avoided, and the spatial patterns and spatial resources of cities can be optimized and rationally allocated [6], [7]. Meanwhile, it can also help to guide local governments to revitalize the urban land stock, promote the intensive urban construction, improve the urban living environment, and strengthen ecological conservation and farmland protection [8], [9]. Overall, scientific delineation of UGB is an important means to promote new urbanization, ecological civilization construction, and modernization of the national governance capacity [10], [11].

UGB was first proposed in the "smart growth" movement in the USA, where UGB delineation in the western city of Portland has become a classic case [12]. More recently, India and European countries, such as Switzerland, have recognized the effectiveness of UGB in urban planning [13]. Research on UGB delineation in China began relatively late, with two main categories of methods: spatial simulation and comprehensive evaluation. Spatial simulation of urban expansion relies on cities' historical development dynamics to construct models, such as cellular automata (CA) and artificial neural networks, to simulate the future land use layout and predict the boundaries of construction land [14], [15]. This approach focuses on internal drivers of urban expansion and economic development, such as transportation, location conditions, and population, lacking sufficient attentions to the demand for and carrying capacity of urban resources and the environment [16], [17]. As another type of method, comprehensive evaluation is to identify areas

Manuscript received 6 September 2023; revised 28 November 2023; accepted 11 April 2024. Date of publication 16 April 2024; date of current version 1 May 2024. This work was supported by the National Natural Science Foundation of China under Grant 42101415 and Grant 42230113; in part by the Research and Development Program of China under Grant 2022YFC3800804-01; and in part by the Ministry of Education of Humanities and Social Science Project under Grant 21YJCZH181. *(Corresponding author: Nan Xia.)*

The authors are with the Jiangsu Provincial Key Laboratory of Geographic Information Technology, Nanjing University, Nanjing 210023, China, and also with the School of Geography and Ocean Science, Nanjing University, Nanjing 210023, China (e-mail: xinggao@smail.nju.edu.cn; xianan@nju.edu.cn; zhuangsudan@smail.nju.edu.cn; zhaoxin@smail.nju.edu.cn; ljl0715@smail.nju.edu.cn; 502022270086@smail.nju.edu.cn; limanchun@nju.edu.cn).

This article has supplementary downloadable material available at https://doi.org/10.1109/JSTARS.2024.3389503, provided by the authors.

Digital Object Identifier 10.1109/JSTARS.2024.3389503

with high development potential and delineate UGB by approaches, such as spatial overlay analysis, fuzzy mathematics, and global sensitivity analysis [18], [19]. This method focuses on the carrying capacities of resource and environmental, natural conditions, spatial suitability of urban construction, protection of ecological land, soil properties, etc. [20]. Meanwhile, it can also account for specific characteristics of the spatial structure and internal drivers of urban development, including urban resource endowments, infrastructure, public services, and population distribution [21]. Compared to the spatial simulation method, the comprehensive evaluation method has the advantages of intuitiveness, clearness, and good operability. It provides multidimensional, multilevel, and multiperspective evaluation tools for effectively studying urban development patterns under tight spatial and resource constraints [22].

Most previous articles have focused on natural and ecological factors when selecting influence indicators of UGB delineation [23], [24]. For human factors, only population and transportation indicators were frequently considered, failing to fully consider the socioeconomic factors related to the urban development potential, such as human activities and economic vitality [25], [26]. Residents' daily activities, such as consumption and movement, contribute to the demand for regional construction land, which can significantly shape the spatial structure of cities [27]. Economic vitality is the primary indicator of urban dynamism, reflecting the present and future economic development, and is an important driver of urban development and expansion [28], [29]. These two factors are often characterized with dynamic features, and most traditional static data with outdated information and poor spatial resolution cannot represent, such as statistical and survey data [30]. The emergence of geographic Big Data covering such topics as human behavior and earth observations provides high-resolution, wide-coverage, and real-time dynamic data for urban dynamic feature sensing [31]. Human behavior Big Data includes the tremendous volume of spatiotemporal behavioral information generated in people's daily lives including cell phone signals, Weibo check-ins, housing rents, traffic operations, and social media data [32], [33]. These data have the advantages of large sample size, high spatial and temporal resolution, expressive spatial differences, and descriptiveness of interactions, which can be used in research, such as urban spatial expansion, resident behavioral characteristics, and separation of employment and housing [34], [35]. Furthermore, a type of earth observation Big Data known as nighttime light (NTL) remote sensing can show the intensity of lights at night on Earth's surface, and have strong correlations with human activities and economic development, which have been successfully applied to studies on urban sprawl and urban area identification [36], [37]. Combinations of geographic Big Data and deeper analysis of human factors have important significance for evaluations of urban areas and the accurate expression of city boundaries, which can improve the accuracy of UGB delineation.

The comprehensive-evaluation-based UGB delineation method requires consideration of many influencing factors, with the assignment of indicator weights directly related to the accuracy of the delineation results. The current methods for assigning weights primarily use the Delphi analytical hierarchy method of expert consultation or the entropy weights method. However, these methods are highly subjective when faced with multidimensional and multisource indicators and their complex nonlinear relationships with UGB [38]. With the rapid development of AI technology, machine learning (ML) has attracted widespread attention from scholars for its powerful learning and computational capabilities that can enable intelligent analysis of data trends and patterns [39]. By selecting relevant samples and training the model to explore nonlinear relationships between influencing factors and geographical phenomena, ML can efficiently clarify the importance of each indicator to the sample data while minimizing errors caused by subjective experience and obtaining more objective and scientific indicator weights [40]. At present, ML-based weight determination methods have been applied in the indicator construction and composite assessment for purposes, such as evaluating performance, detecting spatial patterns, and selecting locations [41], [42], [43]. Moreover, classification, evaluation, and simulation accuracy can be improved by support vector machine, random forest (RF), extreme gradient boosting, and other related ML algorithms used in studies on land use/cover classification and urban sprawl simulation [44], [45]. Thus, by combining the comprehensive evaluation method with ML, the influence of multidimensional impact factors on UGB sample data can be explored to determine the weights of each factor, which can achieve objective, reasonable, and scientifically accurate UGB delineation. Nevertheless, due to a lack of UGB sample data and of clarity in the overall framework, few existing articles on UGB delineation based on the comprehensive evaluation method have incorporated ML algorithms, and there is an urgent need to seek its potentials and broaden the approach to intelligent UGB delineation.

Overall, for the comprehensive evaluation method of UGB delineation, human factors, such as human activities and economic vitality, and the potentials of ML algorithms for determining factor weights have not been fully explored in existing articles. Thus, the Chinese city of Changsha, with a rapidly developing economy, was selected as the study area to carry out objective and empirical UGB delineation by the following means: 1) integrating multisource geographic Big Data to characterize human activities and economic vitality and to construct a comprehensive evaluation index system of natural, human, and ecological factors; and 2) combining ML and multifactor spatial overlay methods to improve the determination of indicator weights to comprehensively evaluate the regional development potential for UGB delineation. The article aims to provide new framework for UGB delineation that can balance the needs of the natural environment, ecological security, and economic development, and to provide theoretical support for high-quality and sustainable urbanization.

The rest of this article is organized as follows. Section II introduces the study area and materials. Section III thoroughly describes the methods of combining multifactor overlay analysis and ML algorithms to delineate the UGB. Section IV analyzes the results in aspects of indicator weight and the UGB delineation. Following this, Section V discusses the results. Finally, Section VI concludes this article.

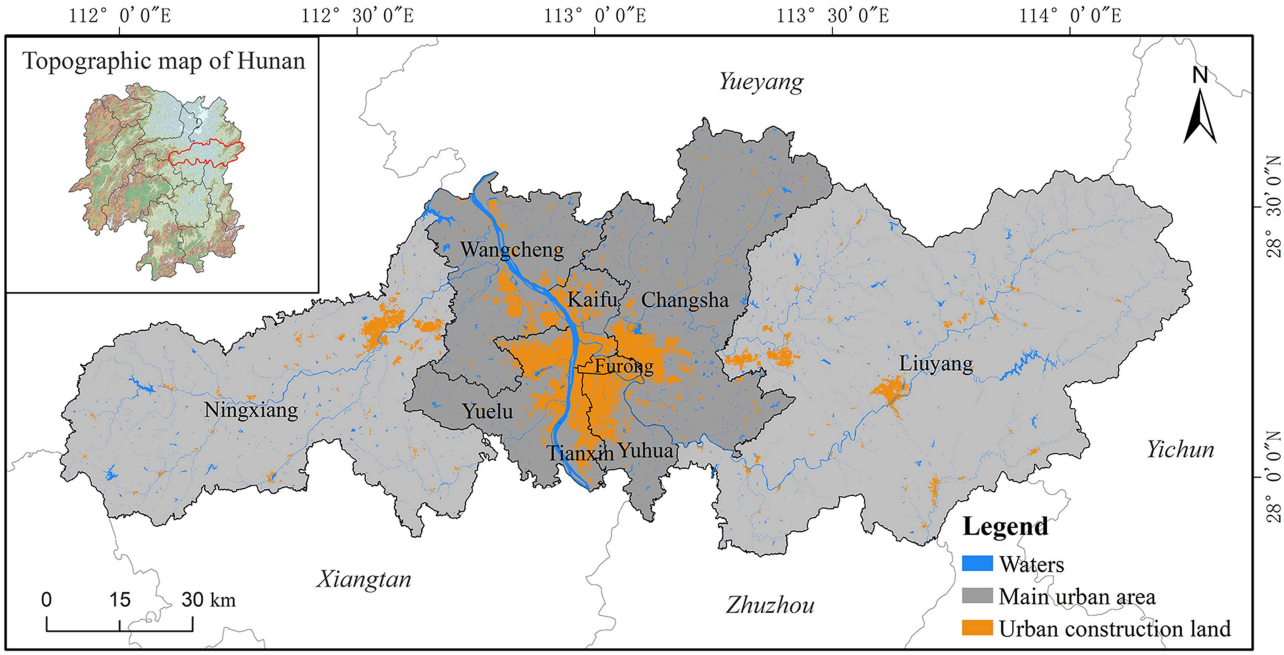


Fig. 1. Location of study area and spatial distribution of current urban construction land in 2019.

II. Study Area and Materials

A. Study Area

The city of Changsha is located in the northeastern part of Hunan Province, in the transition zone between the hills of central Hunan and the coastal plain. The city has nine administrative districts (or counties and county-level cities) (Fig. 1): two county-level cities (Liuyang City and Ningxiang City) and main urban area including Furong, Tianxin, Yuelu, Kaifu, Yuhua, Wangcheng Districts, and Changsha County. The total area of Changsha is 11816.14 km^2. It had a resident population of 10 239 300 at the end of 2021 and an urbanization level of 83.16%. It has favorable natural conditions, with a monsoon-influenced humid subtropical climate consisting of four distinct seasons, rain and heat in the same season, and annual precipitation of 1300–1600 mm. Most of the city's river system belongs to the Xiang River Basin, which provides abundant water resources and has a total storage capacity of 3 billion m^3 [46]. Changsha itself is in the Xiang–Liuyang River Basin. It has undulating terrain, consisting of 30.7% mountains, 19.3% hills, 28.6% sloped terrain, and 21.4% plains. Mount Yuelu is in the west of the city, the Liuyang River passes through the east and the Xiang River through the middle, with the island Juzizhou in the center of the river. As the capital of Hunan Province, Changsha had an estimated GDP growth of 4.8% in 2022, ranking first in central China. It is an important node city in the middle reaches of Yangtze River Urban Agglomeration and the Yangtze River Economic Belt. In recent years, Changsha is undergoing a period of rapid urbanization, with the urban built-up area growing from 272.39 km^2 in 2010 to 560.8 km^2 in 2020 [47]. To achieve high-quality sustainable development, the city is struggling to achieve a balance between natural ecological protection and economic and social development.

B. Materials

The data involved in the article primarily include UGB reference data, spatial survey, natural, and environmental data, and Earth observation and human behavior Big Data. A 300 × 300 m grid was used as the basic unit for UGB delineation.

1) UGB Reference Data (Hereinafter Referred to as UGB$_r$): The vector-format UGB by Changsha Natural Resources and Planning Bureau in 2021 was used as reference data, which was manually mapped based on land demand, policy requirements, planning objectives, current status of ecological agriculture, and natural features. It is planned for 2035, and covers an area of approximately 1651.27 km^2. The Eliminate tool in ArcGIS mapping generalization was used to remove fine fragmented surfaces of area less than 0.09 km^2 from the UGB$_r$ and convert it into raster data to facilitate subsequent model training and weight determination.

2) Spatial Survey Data: The whole of Changsha was classified into six land use types (grassland, arable land, forest land, water, construction land, and unused land) according to 12 primary categories based on data from China's third national land survey in 2019 (referred to as third Survey). The declassified vector-format data were provided by Changsha Natural Resources and Planning Bureau.

Population statistics for each district (county and city) in Changsha were obtained from the seventh national census (referred to as census).

Data on value-added secondary industry of each district, county, and county-level city were obtained from the statistical yearbook for 2020 (referred to as industry survey).

3) Natural and Environmental Data: ASTER global digital elevation model (GDEM) V2 data with 30-m resolution were

obtained from the Geospatial Data Cloud (http://www.gscloud.cn/).

Changsha's geological disaster point data during 2018–2022, including landslides, avalanches, debris flows, and ground collapse hazards, were obtained from the geological cloud platform (https://geocloud.cgs.gov.cn/) (referred to as geological disaster).

Meteorological data including annual average $PM_{2.5}$ and PM_{10} concentration monitoring data at 10 stations of Changsha in 2020 were obtained from China Meteorological Data Service Center (http://data.cma.cn/) to reflect air quality and atmospheric environment (referred to as atmosphere).

Soil data were obtained from the Chinese soil dataset (v1.1) of the Harmonized World Soil Database (http://www.ncdc.ac.cn/portal/), from which attributes of soil erodibility and organic carbon content were extracted (referred to as soil).

4) Earth Observation Big Data: NPP–VIIRS-derived NTLs data for 2020 were obtained with 500-m resolution (https://eogdata.mines.edu/products/vnl/) (referred to as night light).

To characterize vegetation cover, Landsat-8 remote sensing data with a resolution of 30 m from 2020 were acquired from the DATABANK data platform, from which the normalized difference vegetation index (NDVI) data were obtained.

5) Human Behavior Big Data: POI/AOI data were obtained from Baidu Maps and Dianping. These data included highway intersections, railway stations, passenger stations, subways, and bus stop POI data; development zone and industrial park POI data to characterize industrial prospects; AOI data of forest and geological parks, scenic areas, and nature reserves; and commercial services POI data, including gourmet food, life services, leisure and entertainment, sports and fitness, shopping, and beauticians.

Weibo check-in data contained the locations from Weibo platform in Changsha in July 2020 (referred to as Weibo).

Property rent data were obtained from residential rent information on Anjuke (https://changsha.anjuke.com/) in September 2021 and from store rent information on 58.com (https://nj.58.com/), where coordinates information was determined by listing names (referred to as Property rent).

All the human behavior Big Data were obtained by web crawler technology, and data cleaning, coordinates correction, and manual sampling checks were performed to ensure the data reliability and the accuracy.

III. Methods

We constructed a multifactor index system consisting of natural, ecological, and human indicators based on the spatial survey, nature, and environment data, and Earth observation and human behavior Big Data. A variety of methods, including GIS spatial analysis and landscape pattern analysis, were used to spatially quantify each indicator. We then constructed a RF model based on the UGB_r to determine the indicator weight. Finally, we conducted a multifactor overlay analysis to evaluate UGB suitability, thereby delineating the UGB. The specific technical process is shown in Fig. 2.

A. Indicator Construction and Quantification

An index system with natural, ecological, and human dimensions was constructed based on existing articles and literature, combined with the characteristics of Changsha and consultation with relevant experts (Table I). The study area was divided into 133 533 300 × 300 m grids, and UGB suitability was analyzed from a single-factor perspective. The scoring criterion for each of 30 evaluation indicators was obtained according to the natural breakpoint method and the actual circumstances, with low-to-high suitability values of 1, 2, 3, and 4 assigned (see Table SI). The grading results of some indicators are shown in Fig. 3.

1) Natural Factors: Terrain: Two indicators *Slope-A1* and *Elevation-A2* were obtained from the GDEM data. Considering that plains occupy a large proportion of the study area and mountains and hills have low elevations, the scores of UGB suitability could be distinguished as 4–1 for the elevation indicator assigned at about 30-m intervals. Urban land classification and planning construction land standards stipulate that the maximum slope of urban construction land should not exceed 25°, and the slope index were assigned 4–1 at intervals of 5° or 10° to reflect the undulating terrain in the study area.

Waters: Two indicators were *distance to water body-A3* and *water density-A4*. Since urban construction is prohibited in important waters and adjacent areas, the least restrictive factor rule was applied. Grids with a distance score of 1 (<100 m), too close to large rivers, lakes, and reservoirs, were removed from subsequent weighted overlay analysis and their comprehensive evaluation scores were set as 1. The intersection analysis tool was used to calculate the density of the water and high water-density areas were assigned low scores.

Landscape pattern: Using the construction land patches from 2019 third Survey, 11 landscape indices, such as radius of gyration (GYRATE), shape index (SHAPE), and contiguity index (CONTIG) were calculated in the FRAGSTATS platform. The relationship between the 11 normalized landscape indices and UGB_r was obtained through binary logistic regression, and the *comprehensive landscape index (CLI)-A5* [48] characterizing the landscape pattern of UGB was obtained. Higher CLI scores indicated more suitable urban form and spatial pattern for development.

Geological disaster: Geological disaster-prone areas have a harsh natural environment, so urban construction needs to be avoided. The *distance to geological hazard points-A6* indicator was obtained through buffer analysis.

2) Human Factors: Land use situation: We calculated the *proportion of urban land-B1* indicator based on vector data from third survey and the stipulation that if more than 30% of the grid is urban land, it belongs to the urban land grid. We used ArcGIS focal statistics to analyze the proportion of urban land and show the current situation of urban construction.

Location: The *distance to regional centers-B2* indicator was constructed based on the location of district and county government offices as central markers, with urban construction tending to gather toward urban centers.

Industry prospects: *Industrial park density-B3* index was constructed based on the point data of high-tech development zones,

TABLE I
INDEX SYSTEM FOR URBAN GROWTH BOUNDARY DELINEATION

First -level indicator	Second -level indicator	Third -level indicator	Indicator description and calculation	Data source
Nature (A)	Terrain	Slope-A1	Average slope of each grid ($\sum_{i=1}^{n} Slope_i \times S_i / S_{grid}$)	I-DEM
		Elevation-A2	Average elevation of each grid ($\sum_{i=1}^{n} Elevation_i \times S_i / S_{grid}$)	I-DEM
	Waters	Distance to water body-A3	Distance to the nearest important water bodies	II-Third survey
		Water density-A4	Proportion of the water bodies of rivers, lakes, reservoirs, etc. in each grid to the grid area ($S_{water\ body}/S_{grid}$)	II-Third survey
	Landscape pattern	Comprehensive landscape index (CLI)-A5	Comprehensive values of landscape indices, such as GYRATE, SHAPE, and CONTIG ($\mathrm{CLI} = 0.08\times\mathrm{ZGYRATE} + 0.326\times\mathrm{ZPARA} + 0.059\times\mathrm{ZSHAPE} - 0.173\times\mathrm{ZFRAC} - 0.037\times\mathrm{ZCIRCLE} + 0.599\times\mathrm{ZCONTIG}$ [50] Note: Z represents the normalized value)	II-Third survey
	Geological disaster	Distance to geological hazard points-A6	Distance to the nearest geological disaster points (landslides, avalanches, debris flows, and ground collapses)	I-Geological disaster
Humanity (B)	Land use situation	Proportion of urban land-B1	Proportion of urban land within each grid ($S_{urban\ land}/S_{grid}$)	II-Third survey
	Location	Distance to regional centers-B2	Distance to the district-level government offices	II-Third survey
	Industry prospects	Industrial park density-B3	Ratio of the number of high-tech development zones, science parks, and industrial parks in each grid to the grid area ($N_{industrial\ park}/S_{grid}$)	III-POI
	Transportation conditions	Distance to traffic nodes-B4	Distance to the nearest traffic nodes, such as railway and passenger station, and highway intersections	III-POI
		Density of transport lines-B5	Ratio of the length of railway, subway, county, and township roads, town, and village roads in each grid to the grid area ($L_{transport\ line}/S_{grid}$)	II-Third survey
		Density of public transport stations-B6	Ratio of the number of subway and bus stations in each grid to the grid area ($N_{public\ transport\ station}/S_{grid}$)	III-POI
	Public service facilities	Density of SECH facilities-B7	Ratio of the number of scientific, education, cultural, and health (SECH) facilities in each grid to the grid area ($N_{SECH\ facilitie}/S_{grid}$)	II-Third survey
		Density of C&B facilities-B8	Ratio of the number of commercial and business (C&B) facilities in each grid to the grid area, such as food, life services, leisure, fitness, shopping, and beauty ($N_{C\&B\ facilitie}/S_{grid}$)	III-POI
	Human activity	Resident population density-B9	Average density of permanent resident population in the district where the grid is located	II-Census
		Intensity of human activity-B10	Ratio of the number of monthly Weibo check-in records in each grid to the grid area ($N_{Weibo\ check-in\ record}/S_{grid}$)	III-Weibo
	Economic vitality	Value-added of secondary industry-B11	Value-added of the secondary industry in the district where the grid is located	II-Industry survey
		Residential rent-B12	Monthly average residential rent per unit area	III-Property rent
		Store rent-B13	Monthly average store rent per unit area	III-Property rent
		Nighttime economic vitality-B14	Annual average VIIRS night light index	IV-Night light
Ecology (C)	Ecological land situation	Distance to nature reserves-C1	Distance to the nearest nature reserves, such as forest park, geological park, scenic area, etc.	III-POI
		Distance to other ecological lands-C2	Distance to the nearest other ecological lands, such as forest, garden, grassland, wetland, etc.	II-Third survey
		Ecological land coverage rate-C3	Proportion of ecological land area in each grid to the grid area ($S_{ecological\ land}/S_{grid}$)	II-Third survey
	Cultivated land condition	Distance to cultivated land-C4	Distance to the nearest cultivated land	II-Third survey
		Cultivated land coverage rate-C5	Proportion of cultivated land area in each grid to the grid area ($S_{cultivated\ land}/S_{grid}$)	II-Third survey
	Vegetation cover	NDVI-C6	Annual average NDVI value of each grid	IV-NDVI
	Soil properties	Soil organic matter content-C7	Estimation based on soil organic carbon content information in the HWSD soil dataset	I-Soil
		Soil erosion sensitivity-C8	Geometric mean value of sensitivity analysis results R of erodibility, topography, and coverage factors ($\sqrt[3]{R_{\mathrm{erodibility}} \times R_{topography} \times R_{\mathrm{coverage}}}$)	I- Soil
	Atmospheric environment	$PM_{2.5}$ index-C9	Annual average value of $PM_{2.5}$ concentration monitoring by Kriging spatial interpolation	I-Atmosphere
		PM_{10} index-C10	Annual average value of PM_{10} concentration monitoring by Kriging spatial interpolation	I-Atmosphere

Note: S represents square, L represents length, and *N* represents number. I represents natural and environmental data, II represents spatial survey data, III represents human behavior big data, IV represents earth observation big data.

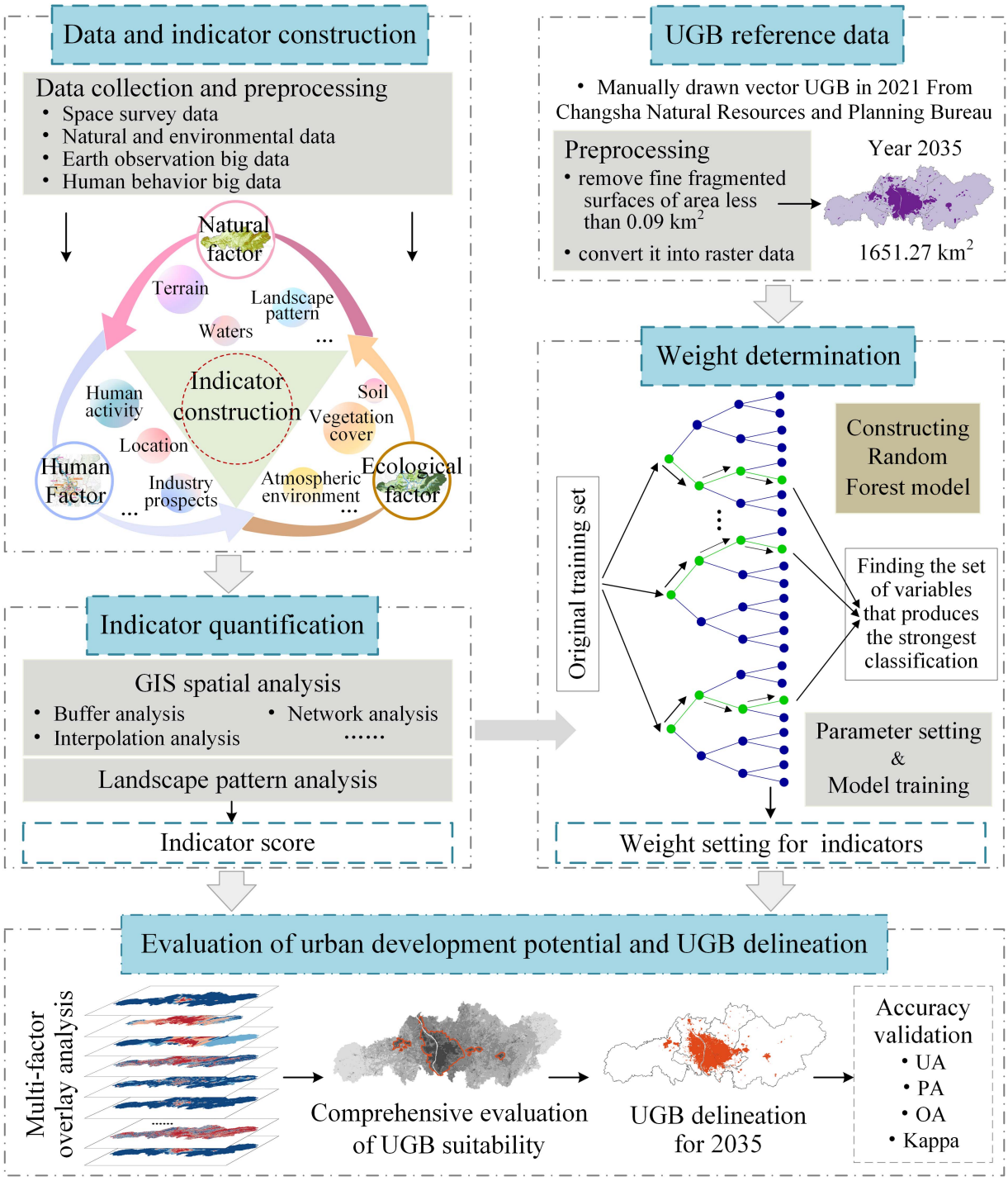


Fig. 2. Technical process of UGB delineation.

science and technology parks, and other industrial parks, as areas with concentrations of technology and industry that can promote the construction of surrounding cities.

Transportation conditions: Three indicators were *distance to traffic nodes-B4*, *density of transport lines-B5*, and *density of public transport stations-B6*. Greater proximity to a transportation corridor facilitates movement and increases the probability for urban sprawl [49].

Public service facilities: Two indicators *density of SECH facilities-B7* and *density of C&B facilities-B8* were constructed using data from the third survey and POI data. A concentration of residential activities, high mobility, and better infrastructure promotes the urban construction.

Human activity: Resident population data from the seventh census were used to construct the *resident population density-B9* indicator. The number of Weibo check-in records in each grid

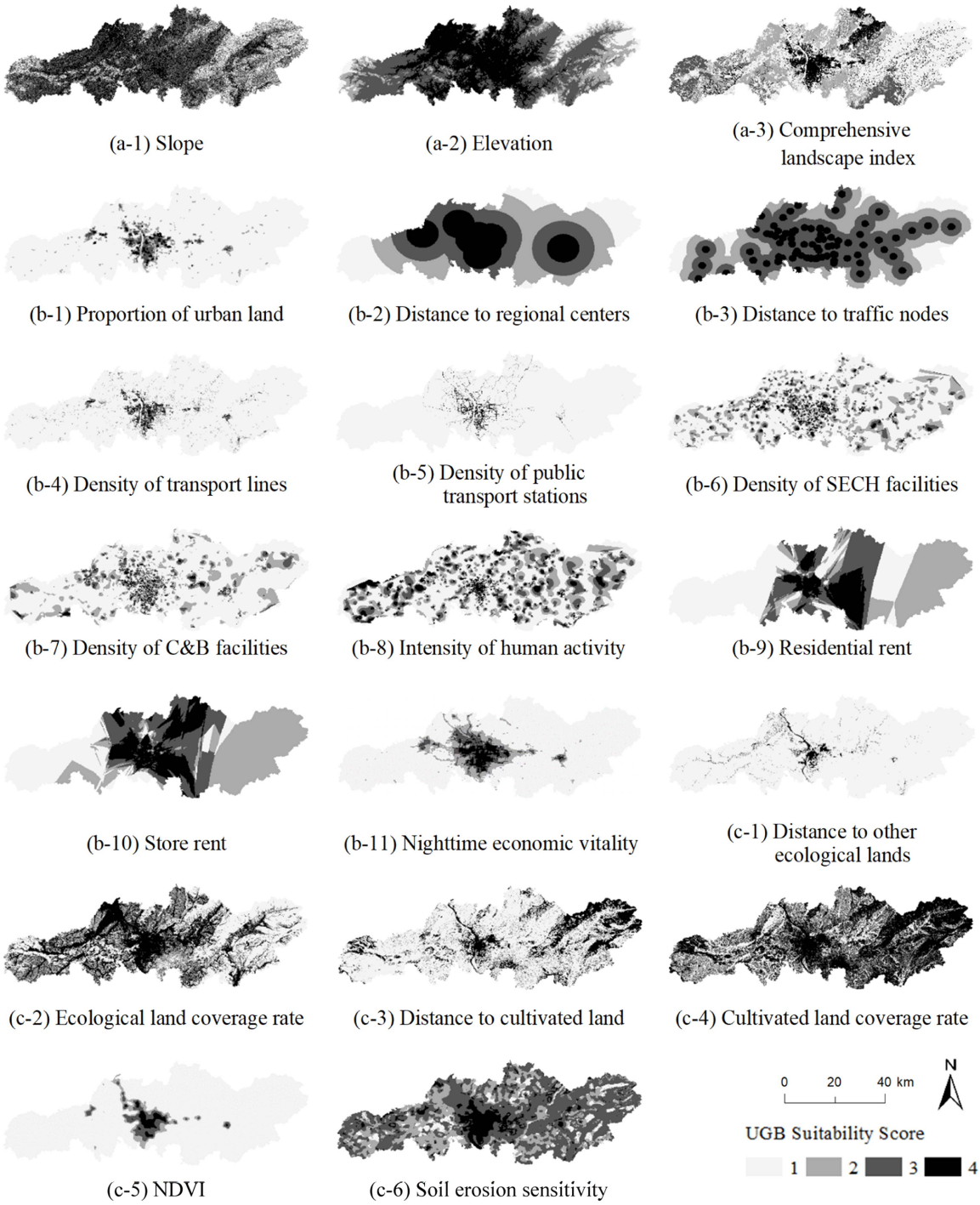


Fig. 3. Evaluation of the urban development boundary suitability score of partial indicators: (a), (b), and (c) represent natural, human, and ecological factors, respectively.

were counted to construct the *intensity of human activity-B10* indicator, with high values representing high intensity of the human activity.

Economic vitality: This was represented by four indicators: *value-added of secondary industry-B11*, *residential rent-B12*, *store rent-B13*, and *nighttime economic vitality-B14*. *B11* can characterize the level of regional industrial development, and is obtained from statistical yearbooks of each district and county. To characterize the land rental level, a score for each grid was obtained by Kriging interpolation using the monthly rental price per m^2 for each residential and commercial location. The NTLs data were used to directly characterize the regional nighttime economic dynamics [51].

3) Ecological Factors: Ecological land situation: The *distance to nature reserves-C1* indicator was the distance to a forest park, geological park, scenic area, nature reserve, or other AOI, and the *distance to other ecological lands-C2* indicator was based on the distance to forest, garden, grassland, wetland,

or similar. The *ecological land coverage rate-C3* was also constructed. Ecological land can provide essential materials for human survival, which needs to be protected and is generally not suitable for urban development.

Cultivated land condition: Changsha has a large proportion of cultivated land, so this is analyzed separately using the indicators *distance to cultivated land-C4* and *cultivated land coverage rate-C5*. Cultivated land is ecologically sensitive and protected by national policies, not suitable urban construction.

Vegetation cover: The *NDVI-C6* index is a quantitative description of vegetation cover and abundance, with a higher value indicating denser vegetation. The cutoff points of 0.4, 0.5, and 0.6 were used to assign a UGB suitability score of 4–1. Grids with a score of 1 were evaluated using the least restrictive factor rule to avoid important ecological areas, such as forests from being classified as UGB.

Soil properties: The *soil organic matter content-C7* indicator was constructed based on organic carbon content data from the soil dataset, which characterizes the degree of fertility of the soil, as organic matter contributes to the development and formation of the soil. The *soil erosion sensitivity-C8* indicator was constructed by calculating the geometric mean of sensitivity analyses of erodibility attributes from the soil dataset as well as topography and coverage factors [52]. The lower the organic matter content and the higher the erosion sensitivity, the more fragile the soil environment in the area, indicating greater sensitivity to human activities, which required strict control of construction activities. Therefore, lower organic matter content and higher erosion sensitivity meant lower UGB suitability.

Atmospheric environment: Based on data from 10 monitoring stations in Changsha, the $PM_{2.5}$ index-C9 and PM_{10} index-C10 indicators at the grid scale in the study area were obtained using Kriging interpolation, where areas with severe air pollution are not suitable as UGB.

B. Weight Determination by Machine-Learning Algorithm

1) Construction of Random Forest Model: RF is a combinatorial algorithm that uses decision trees as base classifiers [53]. A decision tree is a tree-like prediction model consisting of nodes representing features or attributes and directed edges representing the values of the features [54]. The RF is constructed based on the training subset, which is formed by random sampling with put-back from the original dataset. At each internal node, *mtry* predictor variables are randomly selected from original variables as feature variables of the split node, and the best split method is selected to split the node to complete decision training. Decision trees are constructed using a selected subset of features on a training subset, iterating until *ntree* decision trees are generated [55]. After the construction, each decision tree gives the prediction of new samples, and RF determines the final classification result by majority voting [56]. The process of splitting and judging the decision tree is usually based on the purity or impurity of features. First, the best feature is selected as the dividing standard of the current node based on criteria including information gain, Gini index, and mean square error. After determining best features, node splitting is performed with different specific splitting methods until termination conditions are met, such as a maximum depth is reached or the number of samples in a leaf node is less than a predefined threshold [57].

Evaluating the importance of variables in RF to obtain the weights of each variable requires the use of out-of-bag (OOB) data. Since the Bootstrap method was used for random sampling with replacement when constructing decision trees [57], for each tree there were approximately 36.8% samples that were not involved in tree generation, and these are the OOB data [58]. The OOB data are brought into the trained RF model and the calculated percentage of wrong scores is the OOB error, which can be used to measure the importance of the variables [59]. This article used the decrease in average precision, which changes the value of a variable into a random number, to analyze the degree of reduction in the predictive accuracy of the RF with other variables unchanged. A larger value indicates that the variable is more important [53].

To determine the weight of each indicator, the RF model first must be trained on UGB_r. The areas belonging to the UGB in the training data were assigned a value of 1 and other regions were assigned a value of 0 as the UGB status value. Using the random stratified sampling method, about 10% of sample points of UGB and other areas in the training data were extracted, and the sample points' corresponding UGB status values were read using the sample function of ArcGIS. Raster data of the evaluation and grading results of the 30 indicators in the index system were unified into tif format and input to the RF model to obtain the spatial variable dataset, thereby forming the original training set X. Scikit-learn, an open-source toolkit for ML in Python, was used to input the original training set X to train the RF model.

2) Model Training and Parameter Setting: During the training process, the two abovementioned parameters *ntree* (number of decision tree) and *mtry* (number of predictor variables) needed to be set. The RF model combines the predictions of multiple decision trees, and the final prediction results are obtained by voting, so the *mtry* and *ntree* settings have an impact on the prediction accuracy [60]. Since the test sample OOB in the RF can be evaluated internally without the need to cross-validate it or use a separate test set, the OOB misclassification rate is an unbiased estimate of the RF generalization error. Thus, this article used the OOB misclassification rate to measure the model's classification accuracy during parameter setting [60].

Improving prediction accuracy by making adjustments to determine the appropriate parameters is a key step in model training. The randomization of the model is better when the number of decision trees is higher; i.e., larger *ntree* setting within the allowable range means lower error of the model [61]. As shown in Fig. 4, the *ntree* value was set to 100 for testing, and accuracy increased and then decreased as *mtry* increased. When *mtry* was set to 16 for testing, as *ntree* increased, the classification accuracy increased then leveled off. Based on the computational efficiency and computing speed of the model, *ntree* was set to 100 and *mtry* to 16.

After training the RF, variable importance was measured in the generated model using OOB data, and the contribution of

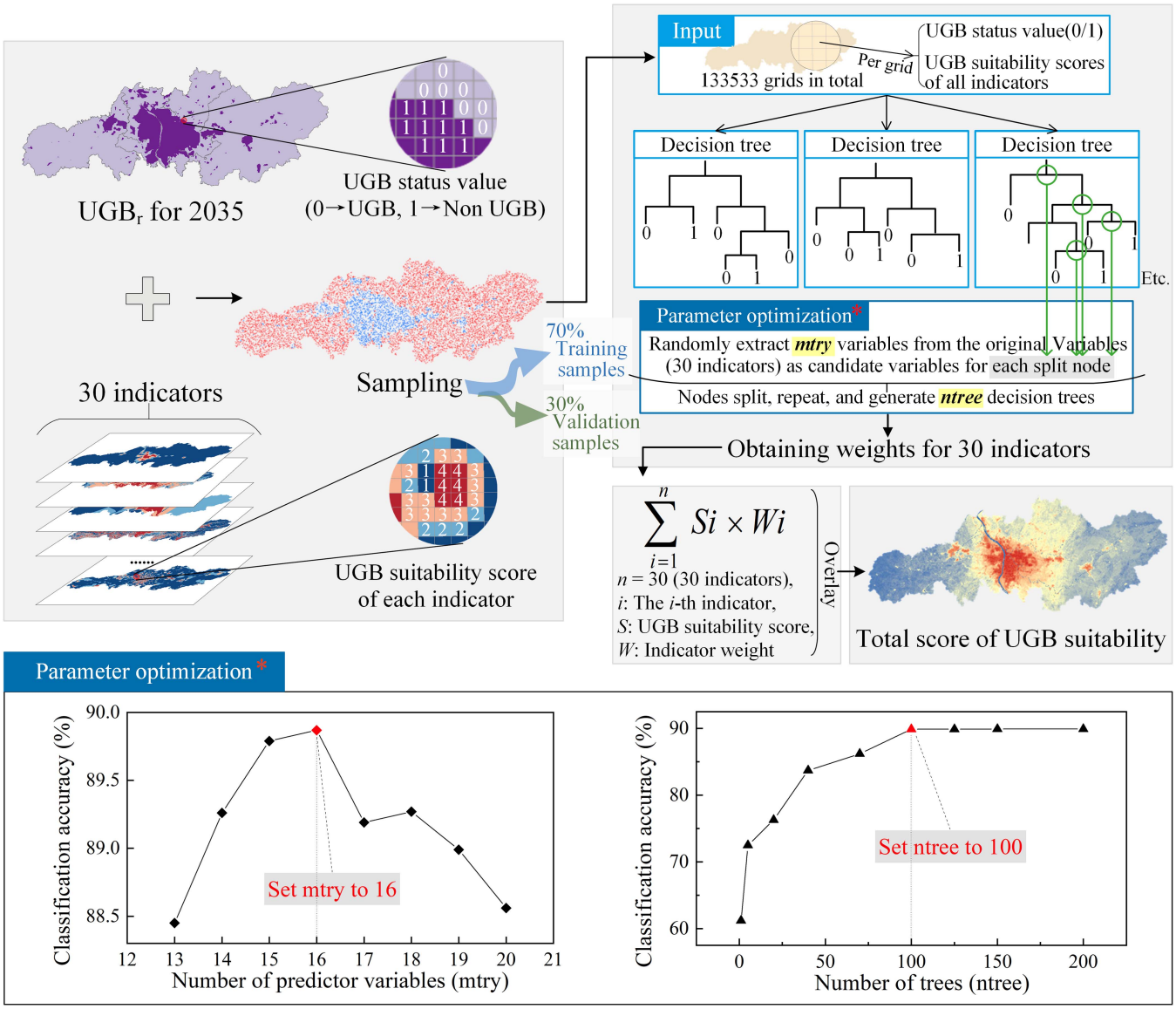


Fig. 4. Schematic diagram of weight determination by random forest model.

each variable to the model could be obtained [41]. Models, such as logistic regression, require linear independence among spatial variables, which is often difficult to satisfy. In comparison, RF, which is a natural nonlinear modeling tool, obtains variable weight values that are more consistent with reality and can evaluate the variables importance more effectively. To improve training and computational speed, multithreaded and parallel independent operations were used to construct the RF model [62].

C. Comprehensive Evaluation of UGB Suitability and Validation

After obtaining the evaluation index system and the results of the 1–4 suitability grading for each indicator, multifactor overlay analysis was conducted according to the index weights determined by the RF model. ArcGIS was used to overlay the suitability scores of each indicator according to the indicator weights, giving the final UGB suitability score for each grid, thus enabling a comprehensive evaluation of the UGB suitability.

The results were evaluated by comparing with the UGB$_r$. Accuracy was validated by user's accuracy (UA), producer's accuracy (PA), and overall accuracy (OA), and the Kappa coefficient (K) to verify the accuracy and reliability of UGB delineation [63]. UA, PA, and OA were calculated based on the statistical concepts of true positive (TP), true negative (TN), false positive (FP), and false negative (FN), using the following equations:

$$\mathrm{UA} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{1}$$

$$\mathrm{PA} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{2}$$

$$\mathrm{OA} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{3}$$

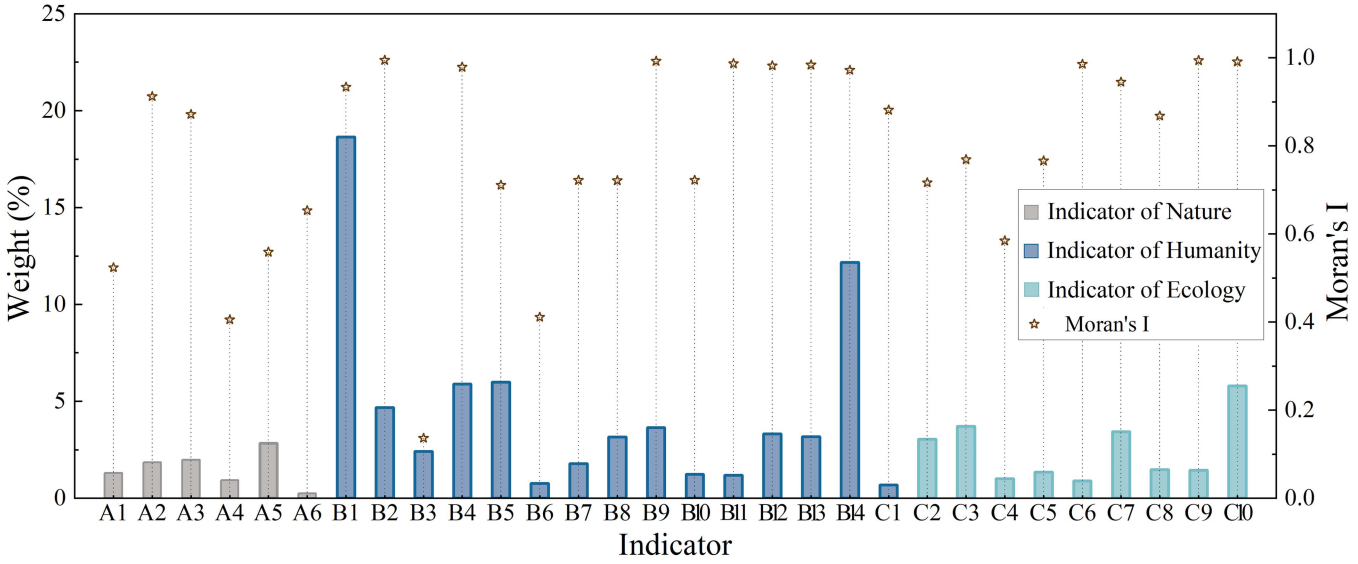


Fig. 5. Indicator contributions by random forest model and Moran's I value of 30 indicators.

where TP is the number of grids attributed to UGB_r that the model classified as UGB, TN is the number of grids not attributed to UGB_r that the model classified as other regions, FP is the number of grids not attributed to UGB_r that the model classified as UGB, and FN is the number of grids attributed to UGB_r that the model classified as other regions.

Different suitability scores were used as thresholds for classifying UGB, and the OA and Kappa coefficient of the UGB delineated under different thresholds are discussed. The highest accuracy score was selected to determine the threshold for classifying UGB in this article. If the grid suitability score was greater than this score, it was designated as UGB, while those with a lower score were not deemed suitable to be UGB. Based on this, the final UGB delineation results of this experiment were obtained, and UA and PA were calculated to assist in the verification of the results.

IV. Results

A. Indicator Weight and Contribution Analysis

In this article, 2000 and 10 000 sample points from 24 597 UGB grids and 108 936 other grids were selected from 2035 UGB_r data, approximately 10%, and then 70% of 12 000 samples formed the training set, 30% for validation set. The spatial variable dataset was obtained based on the data of each indicator after evaluation, and the trained RF model was used to measure the weight of each indicator (Fig. 5). The global Moran's I value of each index was calculated based on the index evaluation grading results (Fig. 3) to reflect the spatial correlation among the evaluation units, to observe the spatial distribution characteristics of high–high and low–low clustering of the development of each factor in the city [64]. Results showed that the Moran's I value of all 30 indicators were greater than 0 and the significance test results were $p < 0.05$, indicating that all the indicators had positive spatial aggregation, which was consistent with the agglomeration principle of urban development. The Moran's I value was greater than 0.9 for several indicators, such as *distance to traffic nodes-B4* and *nighttime economic vitality-B14*, indicating a high level of overall spatial agglomeration.

There were 6 natural, 14 human, and 10 ecological indicators in the constructed index system. Compared with natural and ecological elements, human elements played a larger role in UGB delineation, so they had higher weights. Among the 30 indicators identified in the experiment, *proportion of urban land-B1* and *nighttime economic vitality-B14* had the highest weights—both exceeding 10%, indicating the greatest ability to reflect urban development potential. Meanwhile, nine indicators, including *distance to traffic nodes-B4*, *density of transport lines-B5*, *density of C&B facilities-B8*, *residential rent-B12*, *soil organic matter content-C7*, and *PM_{10} index-C10*, had weights that accounted for more than 3%, indicating a relatively significant impact (Fig. 5). From the weighting ratios of the indicators, the geographic Big Data totally contributed more than 33.72% to UGB delineation. The greater the suitability for urban development, the higher the intensity of human activity and economic vitality. Among these, NTLs data and housing rent data had significant advantages in evaluating urban development potential. Weibo check-in data also played a guiding role in the UGB delineation for its ability to reflect human activities. In addition, better infrastructure and public facilities equated to a higher possibility of urban development, and POI related indicator showed a great influence on the UGB delineation.

B. Urban Growth Boundary Delineation Results and Analysis

1) Urban Growth Boundary Suitability and Threshold Setting: A comprehensive evaluation of the urban development potential of the study area was conducted based on the constructed evaluation index system and the index weights obtained from model training, and the suitability scores of all grids delineated as 2035 UGB were then obtained. Most of the scores were

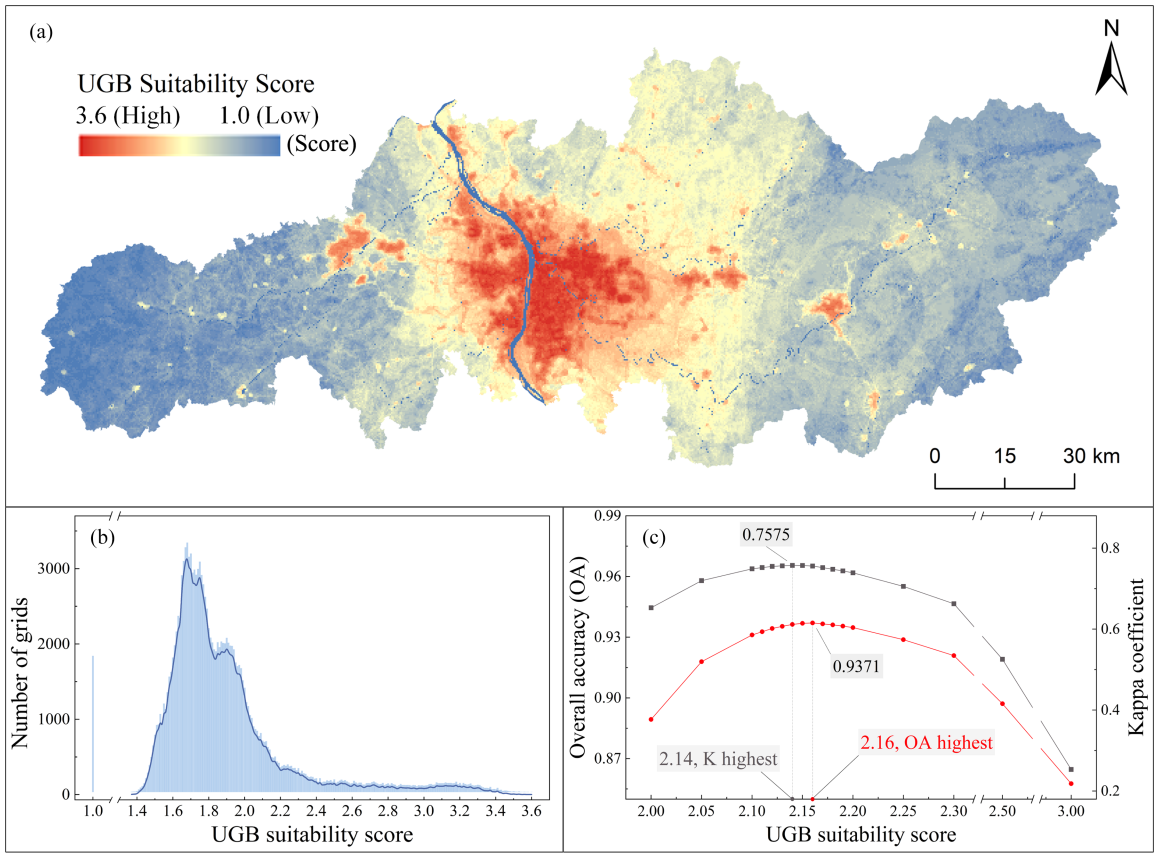


Fig. 6. (a) Spatial distribution of urban development boundary suitability score for 2035 (divided into 30 levels). (b) Frequency histogram of urban development boundary suitability scores. (c) Statistics of delineated accuracy under different suitability scores.

concentrated in the range of 1.6–2.1, with the number of scores between 1 and 2 accounting for approximately 77.09% of the total number of grids, scores between 2 and 3 accounting for about 19.96%, and scores above 3 accounting for about 2.94% [Fig. 6(b)]. The high UGB suitability areas were primarily concentrated in the current built-up areas and their surrounding areas on both sides of the Xiang River and in the central parts of the Liuyang and Ningxiang city, while low suitability scores were primarily concentrated in the mountainous areas in the east and west of the study area [Fig. 6(a)].

Different UGBs were delineated under each threshold values of suitability score with equal interval. The accuracy was analyzed by comparing different delineation results with the UGB_r reference data under threshold value, discussing the corresponding OA and Kappa, and finding the suitability score with the highest delineation accuracy to determine the threshold value for delineating the UGB. Since the suitability scores ranged 1–3.6, the 2–3 point region was selected for accuracy verification, and the range was narrowed down to the 2–2.5 interval by the accuracy results of 2, 2.5, and 3. Finally, statistical analysis was performed at 0.01 intervals in the range of 2.1–2.2, yielding a maximum OA of 93.71% at a threshold of 2.16 and a maximum K of 0.7575 at 2.14 [Fig. 6(c)]. The comprehensive statistical results showed that both OA and K values were closest to the maximum (OA was 93.69% and K was 0.7573) when the suitability score was 2.15, which was determined as the threshold value. The results of UGB delineation based on multisource geographic Big Data (hereinafter referred to as UGB_g) were thus obtained (Fig. 7), and the area of 2035 UGB_g was 1528.06 km^2, about 2.25 times of 2019 current urban construction land.

2) Accuracy of Urban Growth Boundary: UGB_g was primarily concentrated in the main urban area of Changsha along the Xiang River. The eastern part of Liuyang City includes the Dawei and Jiuling mountain ranges, the central part is covered by the Lianyun Mountains, and the western part of Ningxiang City includes the Weishan Mountain area, and these areas are ecological spaces and not suitable for urban construction. Comparison between UGB_g and UGB_r showed that the overall morphology of the two was relatively consistent, with an overlapping area of 1424.60 km^2, indicating that more than 93% of the UGB_r was correctly identified (Fig. 7). The PA of UGB_g delineation was 86.27% and the UA was 93.22% (Table II). The area of misclassification and omission was a total of 330.13 km^2, indicating relatively small errors and that the constructed index system for UGB delineation was reasonable and the model was highly accurate. Compared with the UGB_r area of 1651.27 km^2, the UGB_g area of 1528.06 km^2 delineated in this article was somewhat smaller in scope, particularly in the east-central and south-central parts of Wangcheng District and some areas in the central part of Yuelu District, which were not delineated as development boundaries. The current status of these areas is mostly vegetation, hills, and agricultural land. Some small areas

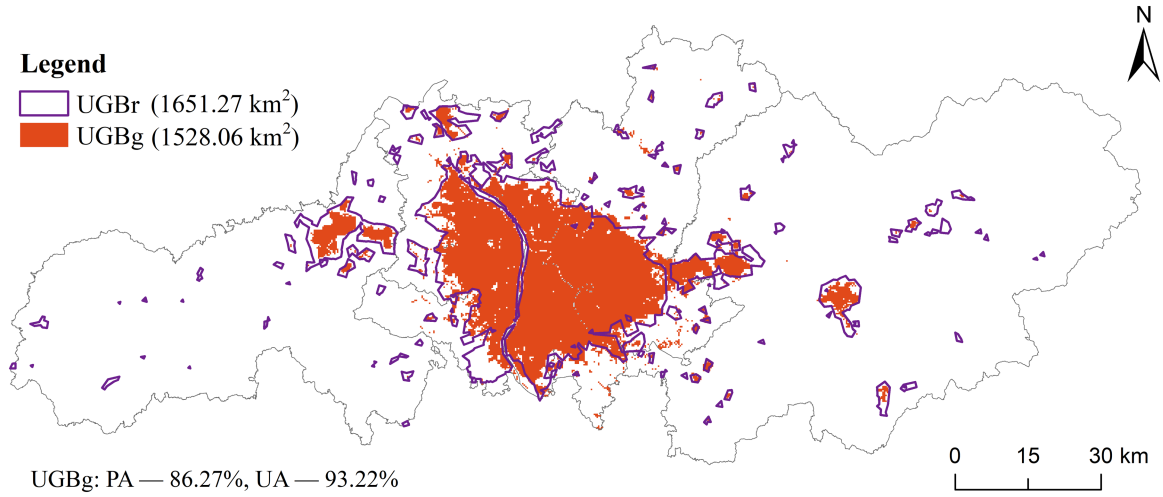


Fig. 7. Comparison between the urban growth boundary reference data (UGB_r) and the urban growth boundary delineated by experiment (UGB_g).

TABLE II
URBAN GROWTH BOUNDARY DELINEATED BY EXPERIMENT (UGB_G), URBAN GROWTH BOUNDARY DELINEATED WITHOUT GEOGRAPHIC BIG DATA (UGB_N), AND URBAN GROWTH BOUNDARY DELINEATED WITHOUT NIGHT LIGHT DATA (UGB_V) DELINEATION ACCURACY AND AREA

	PA	UA	OA	Kappa	Delineated area
UGB_g	86.27%	93.22%	93.69%	0.7573	1528.06 km^2
UGB_n	82.31%	87.75%	88.26%	0.6893	1446.60 km^2
UGB_v	84.98%	90.42%	91.01%	0.7059	1292.17 km^2

scattered in Ningxiang and Liuyang cities were also failed to be included in the UGB delineation.

V. DISCUSSION

A. *Spatial Pattern Analysis of Urban Growth Boundary*

Orientation analysis was used to explore the differences in the spatial distribution of delineated 2035 UGB_g in terms of direction and scale and to identify the urban development direction and growth trend in relation to current urban construction land in 2019. Taking the commercial center of Changsha, Wuyi Square, as the central point, the area of Changsha was divided in 16 directions at an interval of 22.5°, and the ratio of UGB_g to its total area in each direction was calculated to generate the wind-rose diagram, with current construction land and UGB_g as the base map [Fig. 8(a)].

The spatial distribution of UGB_g was consistent with the overall spatial pattern of Changsha: "one axis, one belt, four corridors, as well as one core area, two sub-centers, and ten clusters" as proposed in the territorial spatial overall planning of Changsha city (2021–2035), where Fig. 8(b) and (c), respectively, showed transportation hubs, road networks and clusters, economic development zones, ecological barriers, etc. in the overall planning. The wind-rose diagram indicated that UGB_g was primarily distributed in a W-E direction, with the largest proportion of the UGB_g delineated area in the eastern region, based around development clusters, such as the mature Changsha economic and technological development (ETD) zone and Liuyang ETD zone, with further development at the Xingsha Songya Lake Cluster and the Jinyang New City Cluster. In contrast, the distribution of the UGB_g in the N-S direction accounted for a relatively small percentage, with 28.78% of UGB laid out as the core function area of urban construction under the guidance of the Xiang River Comprehensive Service Axis. This is related to Changsha's long east-west distribution and short vertical urban form. Moreover, the urban development framework emphasizes guidance by key transportation hubs and transportation road networks when building the industrial area of Jinyang New City and Jinzhou New City to the east and west of the center, so that urban development extends along the east-west axis. However, the ecological barrier formed by the mountainous areas to the east and west of Changsha also limits the uncontrolled expansion of the city in the east-west direction.

Compared with current urban construction land [Fig. 8(a)], the UGB_g expanded outward in all directions and extended the most in the E–SEE–SE directions, which accorded with the Changsha masterplan in terms of developing a "Changsha–Liuyang–Ningxiang urban development belt that traverses Ningxiang City, the core area of the main city, and Liuyang City." Under the guidance of the Air–Rail Cluster and the Jinyang New City Cluster east of the Xiang River, urban development will gradually penetrate the southeast of Changsha County. As urban

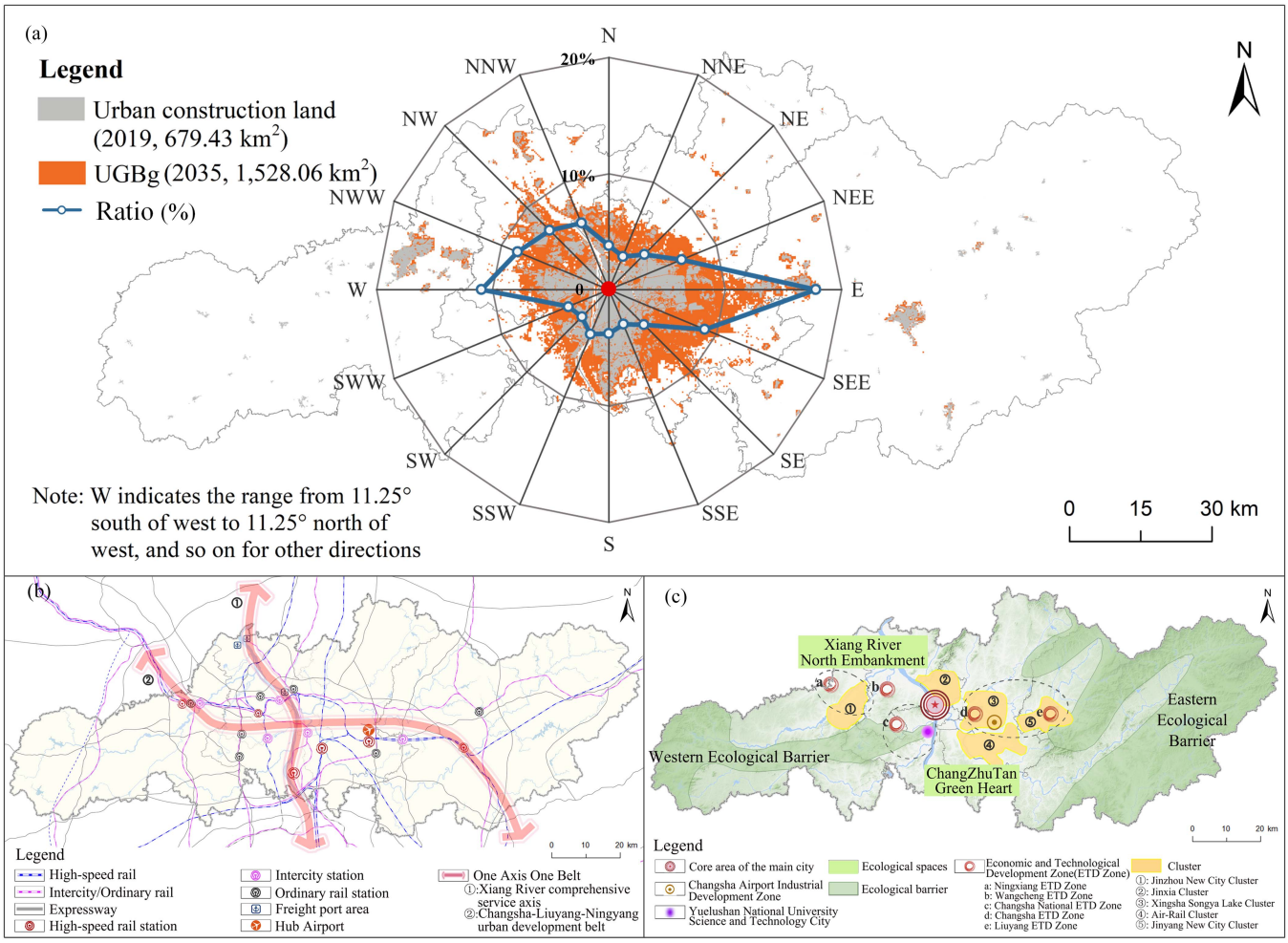


Fig. 8. (a) Wind-rose map of urban growth boundary delineated by experiment (UGB_g) area ratios by orientation and the distribution of 2019 current urban construction land in Changsha. (b) Comprehensive transportation hub, road network, and urban development axis map of Changsha. (c) Territorial spatial pattern, industrial park layout, and ecosystem protection map of Changsha [data from overall land and space planning of Changsha City (2021–2035)].

development needs to avoid the ecological spaces of Xiang River North Embankment and ChangZhuTan Green Heart, expansion in the north–south direction has slowed, and urban construction has focused on upgrading. In the SW and E directions, however, a hub of science and technology innovation has formed around Yuelushan National University Science and Technology City and a hub of aerospace manufacturing has formed around the Changsha and Liuyang ETD Zone. Thus, expansion and multipolar extension of the development boundary have occurred, as is important for healthy urban development.

B. Contributions of Geographic Big Data

Geographic Big Data has provided high spatial—temporal resolution and comprehensive information for the studies of urban dynamics and human behavior, and the laws and trends hidden in the data can be revealed by data mining method, such as ML [31], [33], [35]. This article integrated multiple sources of geographic Big Data, such as Weibo check-ins and property rents to carry out UGB delineation and explored their contribution in UGB delineation. We broadly classified geographic Big Data into two categories: Earth observation and human behavior Big Data [65]. A control experiment was constructed to delineate "UGB without geographic Big Data" (UGB_n), which did not use two types of geographic Big Data in UGB delineation but used only spatial survey and natural and environmental data. Using the same RF model and parameter optimization method, UGB_n delineated an area of 1446.60 km^2 (Fig. 9), with a PA of 82.31%, UA of 87.75%, OA of 88.26%, and K of 0.6893 (Table II). Compared with the delineation of UGB_g, the delineation of UGB_n was less effective, with a 5–6 percentage point reduction in accuracy, indicating the effectiveness of multisource Big Data.

With reference to Google Earth images, UGB_n, UGB_g, and UGB_r, were compared (Fig. 9), and the results of the suitability scores of delineated UGB_n and UGB_g were calculated. The scores were normally distributed but with evident differences. The peak of grid scores of UGB_g was around 1.7, while the peak of UGB_n was around 2.1. The UGB_n was primarily distributed around the old city center in Furong and Tianxin Districts, spreading significantly southward of Yuhua District, while the delineated areas on both the east (eastern Changsha County) and

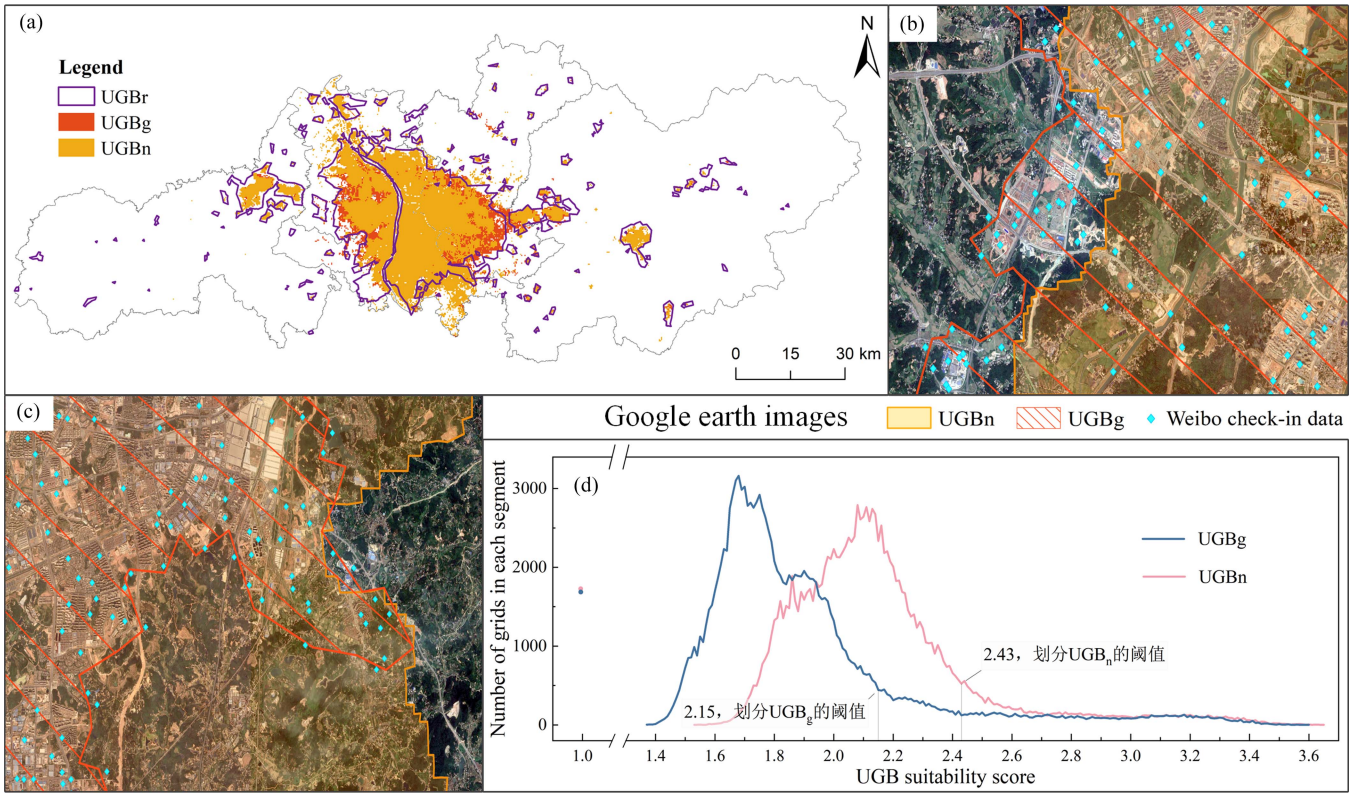


Fig. 9 (a) Urban growth boundary reference data (UGB_r) versus urban growth boundary delimited with or without geographic Big Data (UGB_g and UGB_n, respectively). Comparison of UGB_n and UGB_g supplemented by remote sensing data and Weibo check-in data for example regions in (b) western part and (c) southern part. (D) Histogram comparison chart of UGB_n and UGB_g suitability score.

west (western Wangcheng and Yuelu Districts) were smaller, excluding some suitable. The explanation was that the UGB_n delineation relied only on predominantly natural and land use factors, appropriately as they were the basis for urban development, but these physical characteristics could not represent the socioeconomic attributes of human activities. Geographic Big Data contain rich information on human activities, which can explore the "human–land" relationship from a more comprehensive perspective. The Weibo check-in data indicated that residents' activities in the eastern area of Changsha County and Furong District were widespread [Fig. 3(b-8)], and POI data also reflected the well-developed transport infrastructure and commercial service points in these regions [Fig. 3(b-3), (b-5), and (b-7)]. Furthermore, the property rent and NTLs data could show higher economic vitality in Wangcheng and Yuelu Districts in the western part of the central urban area [Fig. 3(b-9), (b-10), and (b-11)], which had a greater concentration of traffic nodes and industrial parks with more developable land.

For large suitable areas on the east and west sides, UGB_g notably extended to the west of the central city and with an evident small, noncontiguous cluster of development to the east, which accorded with large-scale development of the Air–Rail cluster. And because geographic Big Data, such as Weibo check-in data, were not integrated, part of the densely populated area on the west side of UGB_n was not included within the development boundary [Fig. 9(b)]. In addition, UGB_n extended to the south of Yuhua District, where a large area of ecological land was unsuitable for urban construction [Fig. 9(c)], only based on the spatial survey and natural and environmental data. The human behavior Big Data with a finer scale and wider spatial and temporal scope show low levels of human activity, poor infrastructure, and low economic vitality in this region [such as Weibo check-in data in Fig. 3(b-8)]. NDVI data also revealed that the area had an abundance of vegetation cover where urban construction needs to be avoided; therefore, the southern part of Yuhua District is not suitable for designation as UGB. Generally speaking, geographic Big Data had significant advantages in portraying the "human–land" relationship, and could play an important supportive role in UGB delineation.

Fig. 5 showed that the *nighttime economic vitality-B14* indicator characterized by NTL data had the largest weight share among the geographic Big Data. NTL data with 500-m resolution could characterize subtle economic differences at the grid scale better than economic statistics [36]. NTL data contained rich information on socioeconomic activities which was significantly positively correlated with urbanization, thus played a prominent role in the evaluation of UGB suitability [66]. To further quantify and analyze the contribution of NTL data, "UGB without NTLs" (UGB_v) was defined as removing NTL data from the evaluation model to observe the change in the accuracy of UGB delineation. We focused on the main urban area of Changsha because there was little difference in the delineation results for Liuyang and Ningxiang (Fig. 10). The delineated area of UGB_v was 1292.17 km^2, which was much

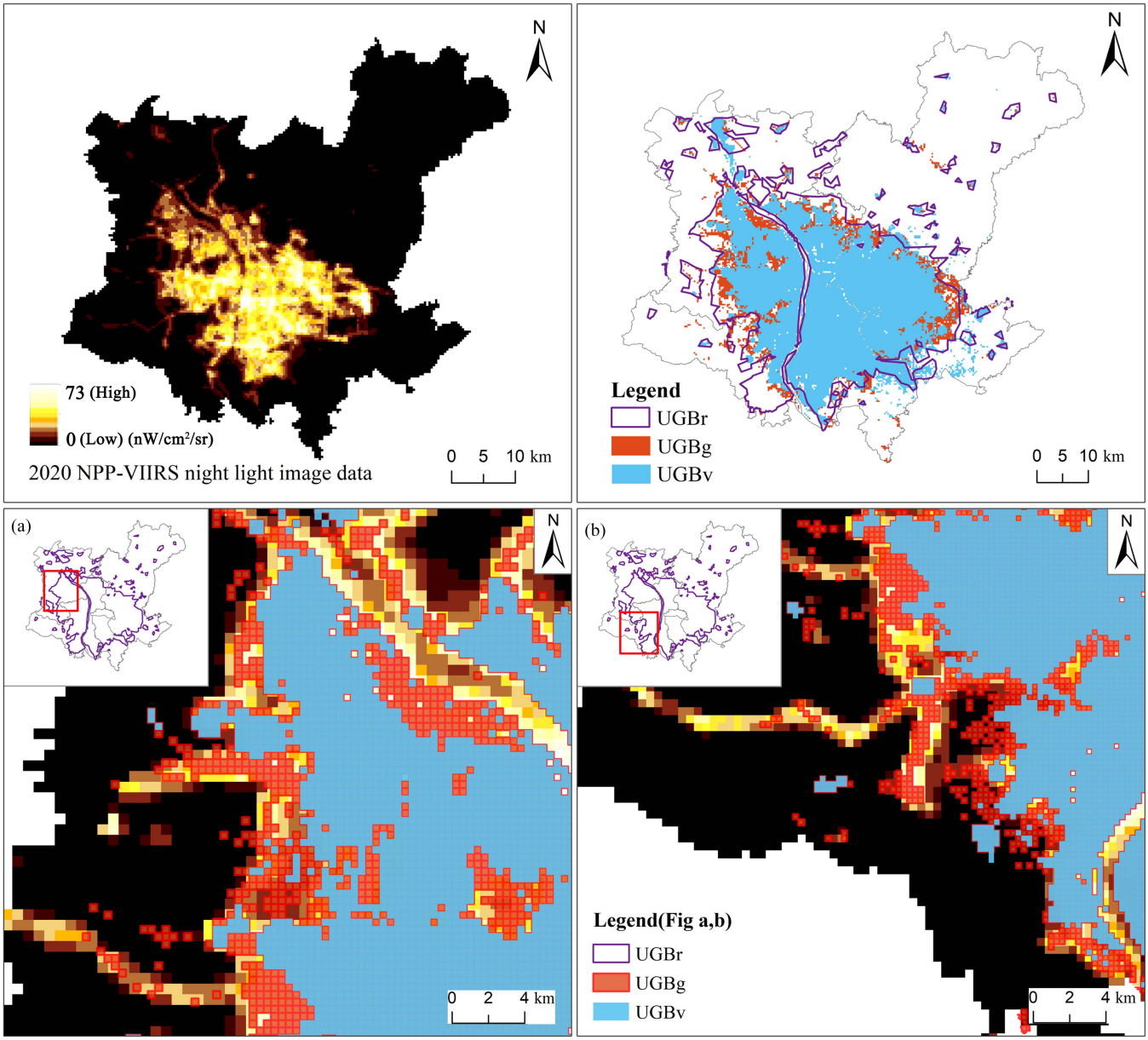


Fig. 10. Urban growth boundary reference data (UGB_r) versus urban growth boundary delimited with or without nightlight data (UGB_g and UGB_v, respectively) for example regions in (a) northwestern part and (b) southwestern part.

smaller than UGB_r and UGB_g. The results of UGB_v had a PA of 84.98%, UA of 90.42%, OA of 91.01%, and K of 0.7059, so accuracy decreased by 2–3 percentage points (Table II). The local comparisons showed that the unidentified areas of UGB_v were mostly areas with high NTL intensity, and the lack of NTL data in the comprehensive evaluation could lead to the failure to identify areas with high urban development potential. Notably, areas such as waters with saturated problem and isolated suburban roads outside urban areas were not designated as UGB_g despite with high NTL value. Therefore, integrating geographic Big Data into the UGB delineation could improve the identification of urban features and human activities, allowing scientific and objective evaluation of urban development potential.

C. Comparative Analysis With Existing Methods

Comparison with existing methods could prove the effectiveness of our proposed method, and different weight determination method was first compared. The analytical hierarchy process method of expert consultation or the entropy method is usually used to determine index weights for the land use suitability evaluation related to UGB delineation. Faced with 30 indicators in the domains of nature, humanity, and ecology, the expert scoring method can encounter difficulties of accurately estimating the weight distributions of many indicators for one expert. Moreover, there may be large differences in the scores of different experts. It is difficult to meet the inconsistency ratio requirement of the analytical hierarchy process

method, and the determination of weights is highly subjective [67]. For the entropy method, it relies on objective data and is only affected by sample data, lacking subjective guidance. The construction of indicators, such as humanities and ecology, are deeply influenced by human factors, so the entropy method cannot comprehensively and correctly represent the real situation of data, and cannot reflect the relationship between various indicators [68]. Comparative experiments were conducted using the entropy method to delineate UGB. According to the index system constructed in the previous section, the source data were normalized and non-negative processed, and the entropy value were calculated according to the formula of the information entropy [69]. Then, the difference coefficient and weight of the indicators were calculated, and the entropy method was used to analyze the processed data in SPSS to obtain the final weight value of indicators. Accordingly, evaluation index system and related indicator weight were combined to delineate UGB in Changsha. Comparing the delineation result with UGB_r, the overlap area was 1341.60 km^2 and the OA was about 87.93%, which was about 5 percentage points lower than our proposed method.

For the other category of UGB delineation method, spatial simulation methods, such as CA models, suffer from the difficulty of acquiring data. Model construction for land use simulation in the near, medium, and long term and related model validation require a long time-series of historical land use data, and insufficient data affect the accuracy of the simulation. Moreover, such models struggle to capture the macro socioeconomic drivers of urban growth, especially the human decision-making process, which means they fail to reflect the land demand of urban construction [70]. Comparative experiments were also conducted based on the CLUMondo model to delineate UGB, which was widely applied in the urban studies [71]. First, according to the spatial policy and restriction document, we added the restricted area layer to CLUMondo model to extract the areas where land use changes were allowed. Then, the model's simulation scenario parameters were set based on the land use data and land use change driving factor files. In this step, Landsat images of 2000, 2015, and 2020 were adopted to classify land use types into agricultural land, ecological land, water area, and construction land. The quantitative change, change rate, and transfer matrix of land use were analyzed accordingly. Based on the indicators in Section III-A, the driving factor files with parameters of the simulation scenario were further generated. Finally, based on land use demand, the difference with simulated land use data was compared in a constant iterative calculation until land use demand was satisfied for CLUMondo model. In this experiment, according to the average growth rate of Changsha's economic and social development in past 10 years, the demand for three parameters of food production, built-up area, and forest area in 2035 was calculated, and the 2010 land use data were used as a benchmark to simulate the scale of construction land in 2035. The simulated construction land was directly deemed as UGB, and correctly identified grids was calculated for consistency check compared with UGB_r. Simulated results showed an overlapping area of 1390.67 km^2 with UGB_r, with the OA of 90.72%, slightly lower than our proposed method.

Overall, in contrast to these two types of methods commonly used in UGB delineation, the present article used ML model to obtain weights, which could reduce the subjectivity of weight determination and guaranteed the objectivity and scientificity of the results. The use of multisource geographic Big Data allowed this article to consider the influence of human behavior and economic vitality on urban development while also accounting for natural factors and ecological protection. This comprehensive analysis provided a more scientific and practical reference for decision-making about the scale and spatial pattern of urban development. However, indicators of human factors in terms of development potential, strategy, and forecasting were lacking in this article. Consequently, the fact that peripheral areas tend to develop and expand outward was not considered during the delineation process. Some geographic Big Data were not included in our index system due to data acquisition difficulties, such as mobile phone signaling data, taxi trajectories data, and land surface temperature data [72]. In addition, relatively low accuracy of the data used for some evaluation indicators resulted in an inability to accurately evaluate some fragmented and small areas with high development potentials, especially geographic Big Data. All the abovementioned issues require more attentions in future articles to further improve the performance of UGB delineation.

D. Suggestions on Urban Growth Boundary Delineation

Urban development is closely related to natural, human, and ecological elements. Integration of the full range of elements can provide multidimensional, multilevel, and multiperspective comprehensive features for the accurate delineation of UGB. Attentions should be paid to adhering to the combination of reverse and positive constraints, fully respecting the natural geographical pattern, and avoiding limiting factors, such as the bottom line of environment and disaster risk in the process of delineation. Meanwhile, the UGB delineation should make full use of natural geography and feature boundaries, such as rivers, hills, and transportation infrastructure, with the form as complete as possible for easy identification and management. Moreover, when extending the proposed method to other cities or regions, differences in the size, spatial structure, development status, and natural resources should be fully considered for the UGB delineation.

The scientific delineation of UGB also requires a combination of multisource geographic Big Data, which can provide high-resolution, wide-coverage, real-time dynamic data to improve the accuracy of UGB delineation. Geographic Big Data, such as human behavior Big Data and NTL remote sensing data, can represent the difference in human activities and economic vitality, which are essential to the UGB delineation. Integration and deeper analysis of geographic Big Data can reveal spatial pattern and trends for urban development, and then allow scientific and objective evaluation of its potentials, which has significance for the accurate expression of urban areas.

The precise implementation of UGB delineation results should be promoted continually by local governments and the Ministry of Natural Resources and Planning. There is a need to strictly regulate the fact that no centralized construction of cities, no planning and construction of development zones, and industrial parks shall be carried out outside UGB. Meanwhile, terrestrial spatial planning within UGB needs to be well arranged in a way that ensures both constraint and flexibility.

VI. Conclusion

This article combined spatial survey data, natural and environmental data, and multisource geographic Big Data, such as NTLs, Weibo check-in, and property rent data, to delineate the UGB. A total of 6 natural, 14 human, and 10 ecological indicators formed the UGB suitability index system, with score of 1–4. Using the manually drawn boundary UGB_r as reference data, the RF model was constructed by setting parameters *ntree* as 100 and *mtry* as 16, and the indicator weight was obtained more objectively and scientifically. Based on these, the multifactor overlay analysis was applied to delineate an area of 1528.06 km^2 UGB_g (with geographic Big Data) in 2035, with suitability score of 2.15 as the threshold value. The overall morphology of the delineation results of was consistent with UGB_r, and the overlapping area reached 1404.60 km^2, with OA of 93.69, PA of 86.27%, and UA of 93.22%. The spatial distribution of UGB_g was consistent with the territorial spatial overall pattern, which primarily distributed in a W-E direction. And it extended the most in the E–SEE–SE directions compared with current urban construction land, which accorded with the Changsha master-plan. The geographic Big Data totally contributed more than 33.72% to UGB delineation, indicating its advantages, where NTLs data and housing rent data had the largest contributions. Comparison analysis showed that the accuracy of UGB_n (without geographic Big Data) and UGB_v (UGB without NTLs data) delineation results decreased by about 5.43 and 2.68 percentage, respectively. By comparison with the entropy method and CLUMondo model urban simulation method, the proposed method was about 5 and 3 percentage points higher in the OA, respectively, which showed a better performance and superiority of the method.

Our proposed UGB delineation model with multisource geographic Big Data provides a scientific, accurate, and novel framework for predicting the potential region of urban development, which can contribute to territorial spatial planning and realization of sustainable and high-quality development. The optimization of the index system, the combination with more types of geographic Big Data, and improvements of the ML model in UGB delineation can be the further articles in the future.

Acknowledgment

The authors would like to thank Prof. Penghui Jiang of the Nanjing Agricultural University. Sincere thanks are given for the comments and contributions of anonymous reviewers and members of the editorial team.

References

[1] L. Sun, J. Chen, Q. Li, and D. Huang, "Dramatic uneven urbanization of large cities throughout the world in recent decades," *Nature Commun.*, vol. 11, Oct. 2020, Art. no. 5366, doi: 10.1038/s41467-020-19158-1.

[2] Ministry of Housing and Urban Rural Development of the People's Republic of China (MOHURDPRC), Urban Planning-Making Method. 2006. [Online]. Available: https://www.gov.cn/gongbao/content/2006/content_454832.htm

[3] H. Wang, X. Ning, H. Zhang, and Y. Liu, "Urban expansion analysis of China's prefecture level city from 2000 to 2016 using high-precision urban boundary," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 7514–7517, doi: 10.1109/IGARSS.2019.8898249.

[4] "General office of the CPC central committee and the state council issued the guidelines on overall delineating and implementing the three lines of control in territorial space planning," General Office of the CPC Central Committee and General Office of State Council, 2019. [Online]. Available: https://www.gov.cn/xinwen/2019-11/01/content_5447654.htm

[5] R. Tan, Y. Liu, Y. Liu, and Q. He, "A literature review of urban growth boundary: Theory, modeling, and effectiveness evaluation," *Prog. Geogr.*, vol. 39, no. 2, pp. 327–338, Feb. 2020, doi: 10.18306/dlkxjz.2020.02.013.

[6] K. Sinclair-Smith, "Methods and considerations for determining urban growth boundaries—An evaluation of the cape town experience," *Urban Forum*, vol. 25, pp. 313–333, Oct. 2014, doi: 10.1007/s12132-013-9207-z.

[7] R. Tan, P. Liu, K. Zhou, and Q. He, "Evaluating the effectiveness of development-limiting boundary control policy: Spatial difference-in-difference analysis," *Land Use Policy*, vol. 120, Sep. 2022, Art. no. 106229, doi: 10.1016/j.landusepol.2022.106229.

[8] J. Olbrich, G. Vich, C. Miralles-Guasch, and L. Fuentes, "Urban sprawl containment by the urban growth boundary: The case of the regulatory plan of the metropolitan region of Santiago of Chile," *J. Land Use Sci.*, vol. 17, pp. 324–338, Jan. 2022, doi: 10.1080/1747423X.2022.2086312.

[9] "Notice of the ministry of natural resources on the management of urban development boundary (Trial)," Ministry of Natural Resources, People's Republic of China, 2023. [Online]. Available: https://www.gov.cn/zhengce/zhengceku/202310/content_6908043.htm

[10] L. Loures, "Post-industrial landscapes as drivers for urban redevelopment. Public versus expert perspectives towards the benefits and barriers of the reuse of post-industrial sites in urban areas," *Habitat Int.*, vol. 45, pp. 72–81, Apr. 2014, doi: 10.1016/j.habitatint.2014.06.028.

[11] M. G. Mortoja and T. Yigitcanlar, "Why is determining peri-urban area boundaries critical for sustainable urban development?," *J. Environ. Plan. Manage.*, vol. 66, no. 1, pp. 67–96, Sep. 2021, doi: 10.1080/09640568.2021.1978405.

[12] M. Stuart, "Regional Planning," in *Growing Smart Legislative Guidebook*, 6th ed. New York, NY, USA: American Planning Association, 2002, pp. 156–180.

[13] B. Weilenmann, I. Seidl, and T. Schulz, "The socio-economic determinants of urban sprawl between 1980 and 2010 in Switzerland," *Landscape Urban Plan.*, vol. 157, pp. 468–482, Aug. 2017, doi: 10.1016/j.landurbplan.2016.08.002.

[14] D. Mitsova, W. Shuster, and X. Wang, "A cellular automata model of land cover change to integrate urban growth with open space conservation," *Landscape Urban Plan.*, vol. 99, no. 2, pp. 141–153, Oct. 2011, doi: 10.1016/j.landurbplan.2010.10.001.

[15] M. Cao, L. Chang, S. Ma, Z. Zhao, K. Wu, and X. Hu, "Multi-scenario simulation of land use for sustainable development goals," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 2119–2127, Feb. 2022, doi: 10.1109/JSTARS.2022.3152904.

[16] Y. Long, H. Han, and Q. Mao, "Establishing urban growth boundaries using constrained CA," *Acta Geographica Sinica*, vol. 64, no. 8, pp. 999–1008, Jan. 2009, doi: 10.11821/xb200908011.

[17] J. J. Arsanjani, M. Helbich, and E. Vaz, "Spatiotemporal simulation of urban growth patterns using agent-based modeling: The case of Tehran," *Cities*, vol. 32, pp. 33–42, Jan. 2013, doi: 10.1016/j.cities.2013.01.005.

[18] H. Luo, B. Ye, and Y. Zhang, "Study of urban land use evaluation of the comprehensive planning—Nanjing city as a case," in *Proc. IEEE Int. Conf. Remote Sens., Environ. Transp. Eng.*, 2011, pp. 4456–4459, doi: 10.1109/RSETE.2011.5965320.

[19] Z. Liu, P. Zhang, G. Li, D. Yang, and M. Qin, "The response of composite ecosystem services to urbanization: From the perspective of spatial relevance and spatial spillover," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 8204–8214, Sep. 2023, doi: 10.1109/JSTARS.2023.3311107.

[20] X. Luo, J. Yang, W. Sun, B. He, and J. Xia, "Suitability of human settlements in mountainous areas from the perspective of ventilation: A case study of the main urban area of Chongqing," *J. Cleaner Prod.*, vol. 310, Aug. 2021, Art. no. 127467, doi: 10.1016/j.jclepro.2021.127467.

[21] Y. E. Fouda and D. M. Elkhazendar, "A criterion for modelling the 'live-and-work' city index using sustainable development indicators," *Int. J. Urban Sustain. Develop.*, vol. 11, no. 1, pp. 24–47, Jun. 2018, doi: 10.1080/19463138.2018.1556161.

[22] D. Dizdaroglu and T. Yigitcanlar, "A parcel-scale assessment tool to measure sustainability through urban ecosystem components: The MUSIX model," *Ecol. Indicators*, vol. 41, pp. 115–130, Jan. 2014, doi: 10.1016/j.ecolind.2014.01.037.

[23] Q. Yang, L. Wang, Y. Li, and X. Qin, "Urban spatial expansion mode based on the construction of landscape ecological security pattern: A case of the coastal area of Jiangsu," *Scientia Geographica Sinica*, vol. 41, no. 5, pp. 737–746, May 2021, doi: 10.13249/j.cnki.sgs.2021.05.001.

[24] S. Wang, Y. Huang, X. Jiang, T. Wang, and Y. Jin, "Identification and optimization of ecological security patterns in the Xiangyang metropolitan area," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 8671–8679, Aug. 2023, doi: 10.1109/JSTARS.2023.3310154.

[25] P. Lin, Y. He, and M. Pei, "Data-driven analysis of traffic volume and hub city evolution of cities in the Guangdong–Hong Kong–Macao greater bay area," *IEEE Access*, vol. 8, pp. 12043–12056, Jan. 2020, doi: 10.1109/ACCESS.2020.2963852.

[26] C. Lu, W. Hong, Y. Wang, and D. Zhao, "Study on the coupling coordination of urban infrastructure and population in the perspective of urban integration," *IEEE Access*, vol. 9, pp. 124070–124086, Sep. 2021, doi: 10.1109/ACCESS.2021.3110368.

[27] G. Pullano, E. Valdano, N. Scarpa, S. Rubrichi, and V. Colizza, "Evaluating the effect of demographic factors, socioeconomic factors, and risk aversion on mobility during the COVID-19 epidemic in France under lockdown: A population-based study," *Lancet Digit. Health*, vol. 2, no. 12, pp. 638–649, Dec. 2020, doi: 10.1016/S2589-7500(20)30243-0.

[28] D. Gilles, "The death and life of great American cities/The economy of cities," *Regional Stud.*, vol. 51, no. 12, pp. 1871–1873, 2017, doi: 10.1080/00343404.2017.1364041.

[29] X. Wang, Y. Li, H. Gao, and Y. Mu, "Research on economic vitality based on combination weighting method and grey relational analysis," in *Proc. IEEE 35th Youth Acad. Annu. Conf. Chin. Assoc. Automat.*, 2020, pp. 584–589, doi: 10.1109/YAC51587.2020.9337688.

[30] J. Yang, H. You, Y. Zhang, and C. Jin, "Research progress on human settlements: From traditional data to Big Data+," *Prog. Geogr.*, vol. 39, no. 1, pp. 166–176, Jan. 2020, doi: 10.18306/dlkxjz.2020.01.016.

[31] J. Ren, J. Yang, F. Wu, W. Sun, X. Xiao, and J. Xia, "Regional thermal environment changes: Integration of satellite data and land use/land cover," *iScience*, vol. 26, no. 2, Feb. 2023, Art. no. 105820, doi: 10.1016/j.isci.2022.105820.

[32] J. Kandt and M. Batty, "Smart cities, Big Data and urban policy: Towards urban analytics for the long run," *Cities*, vol. 109, Feb. 2020, Art. no. 102992, doi: 10.1016/j.cities.2020.102992.

[33] K. V. Katsikopoulos and M. C. Canellas, "Decoding human behavior with Big Data? Critical, constructive input from the decision sciences," *AI Mag.*, vol. 43, no. 1, pp. 126–138, Mar. 2022, doi: 10.1002/aaai.12034.

[34] F. D. Salim, B. Dong, M. Ouf, Q. Wang, I. Pigliautile, and X. Y. Kang, "Modelling urban-scale occupant behaviour, mobility, and energy in buildings: A survey," *Building Environ.*, vol. 183, Oct. 2020, Art. no. 106964, doi: 10.1016/j.buildenv.2020.106964.

[35] N. E. Phillips, B. L. Levy, R. J. Sampson, M. L. Small, and R. Wang, "The social integration of American cities: Network measures of connectedness based on everyday mobility across neighborhoods," *Sociol. Methods Res.*, vol. 50, no. 3, pp. 1110–1149, Aug. 2021, doi: 10.1177/0049124119852386.

[36] Y. Yang, Y. Wang, C. Xiu, X. Xiao, J. Xia, and C. Jin, "Optimizing local climate zones to mitigate urban heat island effect in human settlements," *J. Cleaner Prod.*, vol. 275, Dec. 2020, Art. no. 123767, doi: 10.1016/j.jclepro.2021.127467.

[37] T. H. Kabanda, "Using land cover, population, and night light data to assess urban expansion in Kimberley, South Africa," *South Afr. Geogr. J.*, vol. 104, no. 4, pp. 539–552, Jan. 2022, doi: 10.1080/03736245.2022.2028667.

[38] J. Pena, G. Napoles, and Y. Salgueiro, "Implicit and hybrid methods for attribute weighting in multi-attribute decision making: A review study," *Artif. Intell. Rev.*, vol. 54, no. 5, pp. 3817–3847, Jan. 2021, doi: 10.1007/s10462-020-09941-3.

[39] C. Janiesch, P. Zschech, and K. Heinrich, "Machine learning and deep learning," *Electron. Markets*, vol. 31, pp. 685–695, Apr. 2021, doi: 10.1007/s12525-021-00475-2.

[40] S. D. Nagappan and S. M. Daud, "Machine learning predictors for sustainable urban planning," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 7, pp. 772–780, Jul. 2021, doi: 10.14569/IJACSA.2021.0120787.

[41] H. Zhang et al., "Combing remote sensing information entropy and machine learning for ecological environment assessment of Hefei–Nanjing–Hangzhou region, China," *J. Environ. Manage.*, vol. 325, Jan. 2023, Art. no. 2012658, doi: 10.1016/j.jenvman.2022.116533.

[42] M. Adamiak, I. Jazdzewska, and M. Nalej, "Analysis of built-up areas of small Polish cities with the use of deep learning and geographically weighted regression," *Geosciences*, vol. 11, no. 5, May. 2021, Art. no. 223, doi: 10.3390/geosciences11050223.

[43] L. Zhang, Z. Xue, H. Liu, and H. Li, "Enhanced generalized regression neural network with backward sequential feature selection for machine-learning-driven soil moisture estimation: A case study over the Qinghai–Tibet plateau," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 7173–7185, Jul. 2023, doi: 10.1109/JSTARS.2023.3298946.

[44] S. Talukdar, P. Singha, S. Mahato, S. Pal, Y. A. Liou, and A. Rahman, "Land-use land-cover classification by machine learning classifiers for satellite observations—A review," *Remote Sens.*, vol. 12, no. 7, Apr. 2020, Art. no. 1135, doi: 10.3390/rs12071135.

[45] S. Singh, M. J. Rao, N. K. Baranval, K. V. Kumar, and Y. V. Kumar, "Geoenvironment factors guided coastal urban growth prospect (UGP) delineation using heuristic and machine learning models," *Ocean Coastal Manage.*, vol. 236, Apr. 2023, Art. no. 106496, doi: 10.1016/j.ocecoaman.2023.106496.

[46] M. Mei, Z. Chen, Y. Zhang, and Y. Zhang, "Study on the delineation method of urban growth boundary under the coordination of residential activity space and ecological constraint: A case study of Changsha, China," *Resour. Environ. Yangtze Basin*, vol. 27, no. 11, pp. 2472–2480, Nov. 2018, doi: 10.11870/cjlyzyyhj201811009.

[47] *Changsha Statistical Yearbook 2021*. Beijing, China: China Statistics Press, 2021. [Online]. Available: http://tjj.changsha.gov.cn/tjnj/2022/mulu.html

[48] J. Nowosad and T. F. Stepinski, "Global inventory of landscape patterns and latent variables of landscape spatial configuration," *Ecol. Indicators*, vol. 89, no. 6, pp. 159–167, Feb. 2018, doi: 10.1016/j.ecolind.2018.02.007.

[49] K. C. Clarke, S. Hoppen, and L. Gaydos, "A self-modifying cellular automaton model of historical urbanization in the San Francisco bay area," *Environ. Plan. B-Plan. Des.*, vol. 24, no. 2, pp. 247–261, Apr. 1997, doi: 10.1068/b240247.

[50] X. Li, *Research on Delineation of Urban Growth Boundary Based on Landscape Index*. Beijing, China: China Univ. Geosci., 2019.

[51] B. Zeng, "Spatial pattern evolution and influencing factors of county-level economy of border regions in Hunan-Hubei-Jiangxi based on nighttime light data," *Scientia Geographica Sinica*, vol. 40, no. 6, pp. 900–907, Jun. 2020, doi: 10.13249/j.cnki.sgs.2020.06.005.

[52] Y. C. Li et al., "Assessment and spatial differentiation of sensitivity of soil erosion in three Gorges Reservoir area of Chongqing," *Acta Ecologica Sinica*, vol. 29, no. 2, pp. 788–796, Feb. 2009, doi: 10.3321/j.issn:1000-0933.2009.02.029.

[53] L. Breiman, "Random forest," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.

[54] B. F. Frimpong and F. Molkenthin, "Tracking urban expansion using random forests for the classification of Landsat imagery (1986–2015) and predicting urban/built-up areas for 2025: A study of the Kumasi metropolis, Ghana," *Land*, vol. 10, no. 1, pp. 1–21, Jan. 2021, doi: 10.3390/land10010044.

[55] J. Peters, B. D. Baets Baets, N. E. C. Verhoest, R. Samson, and S. Degroeve, "Random forests as a tool for ecohydrological distribution modelling," *Ecol. Model.*, vol. 207, no. 2–4, pp. 304–318, May 2007, doi: 10.1016/j.ecolmodel.2007.05.011.

[56] L. Dong, H. Du, F. Mao, N. Han, X. Li, and G. Zhou, "Very high resolution remote sensing imagery classification using a fusion of random forest and deep learning technique—Subtropical," *Observ. Remote Sens.*, vol. 13, pp. 113–128, Dec. 2020, doi: 10.1109/JSTARS.2019.2953234.

[57] M. Schonlau and R. Y. Zou, "The random forest algorithm for statistical learning," *Stata J.*, vol. 20, no. 1, pp. 3–29, Mar. 2020, doi: 10.1177/1536867X20909688.

[58] P. Probst, M. N. Wright, and A. L. Boulesteix, "Hyperparameters and tuning strategies for random forest," *Wiley Interdiscipl. Rev.-Data Mining Knowl. Discov.*, vol. 9, no. 3, Jan. 2019, Art. no. e1301, doi: 10.1002/widm.1301.

[59] R. David, "The elements of statistical learning: Data mining, inference, and prediction," *J. Amer. Statist. Assoc.*, vol. 99, no. 466, 2004, Art. no. 567, doi: 10.1198/jasa.2004.s339.

[60] A. Aulia, D. Jeong, I. M. Saaid, D. Kania, M. T. Shuker, and N. A. El-Khatib, "A random forests-based sensitivity analysis framework for assisted history matching," *J. Petroleum Sci. Eng.*, vol. 181, Oct. 2019, Art. no. 106237, doi: 10.1016/j.petrol.2019.106237.
[61] S. V. Stehman, "Selecting and interpreting measures of thematic classification accuracy," *Remote Sens. Environ.*, vol. 62, no. 1, pp. 77–89, Oct. 1997, doi: 10.1016/S0034-4257(97)00083-7.
[62] A. Verikas, A. Gelzinis, and M. Bacauskiene, "Mining data with random forests: A survey and results of new tests," *Pattern Recognit.*, vol. 44, no. 2, pp. 330–349, Aug. 2011, doi: 10.1016/j.patcog.2010.08.011.
[63] V. F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sanchez, "An assessment of the effectiveness of a random forest classifier for land-cover classification," *ISPRS J. Photogrammetry Remote Sens.*, vol. 67, pp. 93–104, Nov. 2011, doi: 10.1016/j.isprsjprs.2011.11.002.
[64] F. Xu, "Improving spatial autocorrelation statistics based on Moran's Index and spectral graph theory," *Urban Develop. Stud.*, vol. 28, no. 12, pp. 92–101, Dec. 2021, doi: 10.3969/j.issn.1006-3862.2021.12.020.
[65] Y. Liu, Q. Liu, M. Deng, and Y. Shi, "Recent advance and challenge in geospatial Big Data mining," *Acta Geodaetica et Cartographica Sinica*, vol. 51, no. 7, pp. 1544–1560, Jul. 2022, doi: 10.11947/j.issn.1001-1595.2022.7. chxb202207034.
[66] H. M. Nguyen and L. D. Nguyen, "The relationship between urbanization and economic growth an empirical study on ASEAN countries," *Int. J. Social Econ.*, vol. 45, no. 2, pp. 316–339, Dec. 2016, doi: 10.1108/IJSE-12-2016-0358.
[67] W. Ho and X. Ma, "The state-of-the-art integrations and applications of the analytic hierarchy process," *Eur. J. Oper. Res.*, vol. 267, no. 2, pp. 399–414, Sep. 2017, doi: 10.1016/j.ejor.2017.09.007.
[68] W. Yuan and Z. Liu, "Study on evaluation method of energy-saving potential of green buildings based on entropy weight method," *Int. J. Glob. Energy Issues*, vol. 45, pp. 448–460, Aug. 2023, doi: 10.1504/IJGEI.2023.132014.
[69] M. Zhang, J. Zhou, and R. Zhou, "Evaluating sustainability of regional water resources based on improved generalized entropy method," *Entropy*, vol. 20, Nov. 2018, Art. no. 715, doi: 10.3390/e20090715.
[70] L. Zhao, J. Yang, C. Li, Y. Ge, and Z. Han, "Progress on geographic cellular automata model," *Scientia Geographica Sinica*, vol. 36, no. 8, pp. 1190–1196, Aug. 2016, doi: 10.13249/j.cnki.sgs.2016.08.009.
[71] Y. Zhao, D. Su, Y. Bao, W. Yang, and Y. Sun, "A CLUMondo model-based multi-scenario land-use change simulation in the Yangtze river delta urban agglomeration, China," *Sustainability*, vol. 14, Nov. 2022, Art. no. 15336, doi: 10.3390/su142215336.
[72] J. Yang et al., "Understanding land surface temperature impact factors based on local climate zones," *Sustain. Cities Soc.*, vol. 69, Jun. 2021, Art. no. 102818, doi: 10.1016/j.scs.2021.102818.

Xing Gao received the B.S. degree in geographic information science from the School of Resource and Environmental Sciences, Wuhan University, Wuhan, China, in 2021. She is currently working toward the M.S. degree in cartography and geographic information system with the School of Geography and Ocean Science, Nanjing University, Nanjing, China.

Her research interests include digital land spatial planning and urban–rural development.

Nan Xia received the B.S. degree in geographic information system in 2014, and the Ph.D. degree in cartography and geographic information system in 2020, both from the School of Geography and Ocean Science, Nanjing University, Nanjing, China.

He is currently an Associate Researcher with the School of Geography and Ocean Science, Nanjing University, Nanjing, China. His research interests include Big Data analysis and mining, land use simulation, and territorial spatial planning.

Sudan Zhuang received the B.S. degree in tourism management from the School of Business Administration, Southwest University of Finance and Economics, Chengdu, China, in 2021. She is currently working toward the M.S. degree in human geography with the School of Geography and Ocean Science, Nanjing University, Nanjing, China.

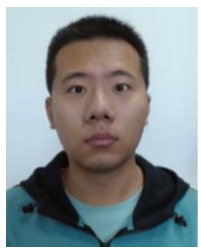
Her research interests include transport and economic development and their correlation.

Xin Zhao is currently working toward the Ph.D. degree in geography with the School of Geography and Ocean Science, Nanjing University, Nanjing, China.

Her research interests include geographic spatiotemporal data mining and the application of geographic information systems.

Jiale Liang received the B.S. degree in city management from the School of Public Administration, Shandong University of Finance and Economics, Jinan, China, in 2019, and the M.S. degree in land resources management from the China University of Geosciences, Wuhan, China, in 2022. He is currently working toward the Ph.D. degree in geography with the School of Geography and Ocean Science, Nanjing University, Nanjing, China.

His research interests include land use change and territorial spatial planning.



Ziyu Wang received the B.S. degree in geographic information science in 2022 from the School of Geography and Ocean Science, Nanjing University, Nanjing, China, where he is currently working toward the M.S. degree in cartography and geographic information system.

His research interests include urban geography and data integration.

Manchun Li received the B.S. degree in cartography in 1987, and the Ph.D. degree in human geography in 1997, both from the School of Geography and Ocean Science, Nanjing University, Nanjing, China.

He is currently a Professor with the School of Geography and Ocean Science, Nanjing University, Nanjing, China. He has authored more than 200 papers in journals and refereed conference. His research interests include digital land spatial planning, urban–rural development, and geographic information technology.