Refined Urban Functional Zones Identification via Empirical Bayesian Kriging: A POI-Weighted Scoring Innovation

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal
IEEE GEOSCIENCE AND REMOTE SENSING SOCIETY SECTION

Received 7 December 2024, accepted 21 January 2025, date of publication 29 January 2025, date of current version 5 February 2025.
Digital Object Identifier 10.1109/ACCESS.2025.3536000

RESEARCH ARTICLE

Refined Urban Functional Zones Identification via Empirical Bayesian Kriging: A POI-Weighted Scoring Innovation

XUAN DU[1], YISHA PAN[1], XIAOYAN YANG[2], LONGGAO CHEN[1], LIANGCHEN LIU[1], AND YING LIN[1]
[1]College of Public Policy and Management (College of Emergency Management), China University of Mining and Technology, Xuzhou 221116, China
[2]College of Geography, Geomatics and Planning, Jiangsu Normal University, Xuzhou 221116, China

Corresponding author: Longgao Chen (chenlonggao@163.com)

This work was supported in part by Jiangsu Office of Philosophy and Social Science under Grant 22GLB028, and in part by the Fundamental Research Funds for the Central Universities under Grant 2022ZDPYSK08.

ABSTRACT The refined identification of Urban Functional Zones (UFZs) is crucial for effective urban planning, resource allocation, and environmental monitoring. However, achieving high precision and comprehensive identification of UFZs has been challenging. The rapid proliferation and open accessibility of multi-source data, such as remote sensing imagery and socio-economic datasets, have introduced new opportunities for the dynamic identification of UFZs. This paper proposes a Point of Interest (POI)-weighted scoring method based on Empirical Bayesian Kriging (EBK), which is empirically demonstrated in Xuzhou, China. We integrate multi-source data, including remote sensing-based land use/cover information, road networks, POIs, and building geometry data, to classify UFZs based on their dependence on buildings. Results indicate that: 1) The EBK interpolation method not only accounts for the spatial and directional distance relationships between known and unknown sample points, but also produces more accurate raster images and provides more precise estimates in unmeasured ranges. 2) The method successfully identifies UFZs for 99.5% of the buildings in the central urban area of Xuzhou. Additionally, the incorporation of multi-source data enables the extraction of traffic and ecological zones, facilitating the identification of non-building-reliant zones within the city, thereby enhancing the completeness and comprehensiveness of urban spatial recognition. 3) The overall accuracy achieved is 84.4%, with a Kappa coefficient of 0.804 for the classification results, which represents a significant improvement in UFZs identification accuracy compared to traditional methods and offers a robust scientific basis for urban planning and resource optimization.

INDEX TERMS Urban functional zones, refined identification, multi-source data, empirical Bayesian kriging, confusion matrix, urban planning.

NOMENCLATURE

ABBREVIATIONS

UFZ	Urban Functional Zone.
POI	Point of Interest.
EBK	Empirical Bayesian Kriging.
RS	Remote Sensing.
KDA	Kernel Density Analyze.
IDW	Inverse Distance Weighting.
NDVI	Normalized Difference Vegetation Index.
OSM	Open Street Map.
Com.	Commercial Zone.
Res.	Residential Zone.
Ind.	Industrial Zone.
Pub.	Public Service Zone.
Tra.	Traffic Zone.
Eco.	Ecological Zone.
Mix.	Mixed Zone.
FD	Frequency Density.

The associate editor coordinating the review of this manuscript and approving it for publication was Gerardo Di Martino.

CR Category Ratio.
OA Overall Accuracy.

I. INTRODUCTION

The acceleration of urbanization has not only led to the diversification of urban spatial structures and functional layouts, but also brought about challenges such as traffic congestion, environmental pollution, and inefficient resource allocation [1], [2]. As the basic unit of urban spatial organization, the precise identification of UFZs not only helps to improve the rationality of urban planning but also has great significance in achieving scientific urban management [3], thus promoting the sustainable development of urban areas [4].

Current research on Urban Functional zones (UFZs) identification primarily encompasses different scale identification units, classification systems for UFZs, and methods for multi-source data fusion. In terms of identification units at various scales, early studies divided functional units at a macro scale using administrative divisions such as counties and districts [5]. Subsequently, most studies have utilized road divisions to segment urban blocks [6], [7], [8] or created grids [9], [10], [11]as recognition units, with accuracy depending on the road hierarchy or the grid cell size. In recent years, some scholars have adopted building geometry as the unit of study [12], [13], enabling more precise identification of UFZs at a finer scale and better reflecting internal variations within blocks. Regarding the classification system of UFZs, [14] divided urban spatial structure into three major categories: employment, recreation, and residence. Based on the "Urban Land Classification and Planning Construction Land Standards", many researchers, have further categorized them into six major types: residential, industrial, commercial and service facilities, public management and public service facilities, green space and squares, roads and transportation facilities. Then, additional categories such as public center [15], leisure and tourism, science and education, healthcare and mixed-use zones have been incorporated. In terms of research methods, traditional identification often relies on Remote Sensing (RS) images, field surveys or questionnaires [16], [17]. However, relying solely on RS images is difficult to fully reflect human activities and socio-economic information comprehensively [18], [19]. Field surveys are labor-intensive and time-consuming, while questionnaire samples may not fully capture the complete information about UFZs [20]. Points of Interest (POIs) can reflect information about the distribution and location of various urban functional entities [21]. Many scholars have allocated scores on the characteristics of building area and public influence or recognition bases on Level-1 of POIs [22], [23], and then calculated Frequency Density to determine functional areas. However, the weights of higher-level classifications can lead to increased valuation errors, and the choice of estimation methods for unknown values can also significantly affect the accuracy of identification. Existing studies have applied known sample points to estimate the values of unknown sample points through methods such as Kernel Density Analyze (KDA) [24], [25], Contour Trees [26], Inverse Distance Weighting Interpolation (IDW) [12] or Kriging Interpolation, but these methods are often limited by bandwidth and minimum raster cell size, making it challenging to reveal the micro-scale differentiation characteristics of various UFZs. As the study area expands, the errors in identification tend to increase. Empirical Bayesian Kriging (EBK) is an interpolation analysis method that combines the semi-variance function of Kriging with the Empirical Bayesian approach. Compared to the traditional methods mentioned earlier, EBK offers the advantages of relatively lower computational complexity and model uncertainty, making it more accurate in spatial data modeling and prediction than other methods.

Recently, socio-economic data that encapsulates diverse human-land interactions, such as taxi trajectory [5], [10], [27], social media check-ins [11], [28], online location [29], mobile phone signals [7], [30] and nighttime light data [31] have provided new perspectives for identifying UFZs. These data sources are characterized by high granularity, broad coverage, and ease of acquisition. Since a single data source avoids the problem of data inconsistency, the information covered is limited and the accuracy is unstable, it is difficult to fully reflect the complexity of UFZs. Therefore, considering the objective characteristics of the information reflected by RS images, the integration of RS images with other multi-source socio-economic data, to fully leverage the advantages of different data, is becoming a hot topic in current research for efficient and accurate identification of UFZs.

Building upon the research of predecessors, this paper proposes a POI-weighted scoring method based on EBK and demonstrates empirically in Xuzhou, China. There are three main parts of contribution: First, the POI-weighted score is based on secondary classification criteria, which enhances the reliability of experimental estimation. Second, the EBK interpolation method has the advantage of considering direction and distance, and provides a better algorithmic mechanism, which increases the precision of estimation. Third, multi-source data including RS-based land use/cover information, road networks, POIs, and building geometry data are used to identify UFZs, which are classified by their dependence on buildings, enhancing the comprehensiveness of urban spatial identification. The aim is to achieve fine-grained identification of UFZs and provide a robust scientific basis for urban planning and resource optimization.

II. STUDY AREA AND DATASETS

A. STUDY AREA

Xuzhou, located in the northern part of Jiangsu Province, is a key central city in Huaihai region. The central urban area stands as the nucleus of the city's economic, cultural, technological, and educational life, with a well-developed commercial, service, and industrial sector that drives the city's economic growth. It boasts a thriving commercial, service, and industrial sector, which propels the city's economic prosperity. Moreover, the district places a strong emphasis on ecological and environmental sustainability, development, adorned with an array of parks or verdant spaces that offer residents ample opportunities for relaxation and recreation. With comprehensive and accessible data available for the central urban area, we select the main central area of Xuzhou as the study area (Fig. 1). According to the "Xuzhou Territorial Spatial Plan (2021-2035)", the central urban area of Xuzhou includes five major districts: Gulou, Yunlong, Jiawang, Quanshan and Tongshan.

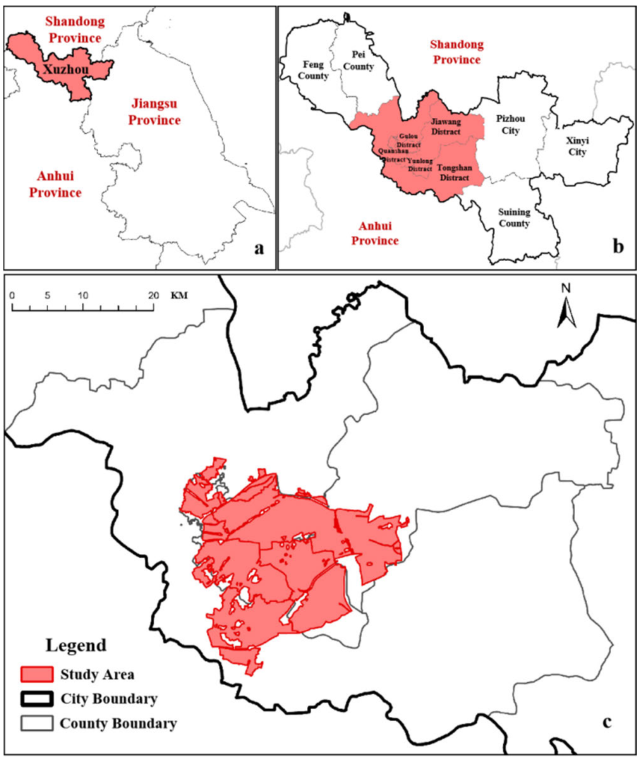

FIGURE 1. Location of the Study Area (a. Xuzhou b. the central urban area of Xuzhou c. the Main Central Area of Xuzhou).

B. DATESETS

Various data for the study area are collected, including basic geographic information, Normalized Difference Vegetation Index (NDVI), 1-meter Resolution Landcover Data (SinoLC-1), Open Street Map (OSM), POI, and Building Geometry data, with data sources as shown in TABLE 1.

C. DATA PREPROCESSING

Data preprocessing primarily involves the preprocessing of OSM, POI, and Building Geometry data (Fig. 2).

For OSM, due to the uneven spatial accuracy and discontinuities in lower-grade roads in certain areas, lower-level roads such as alleys, sidewalks, and footpaths are

TABLE 1. Data sources.

Date Type	Date Sources	Year	Notes
Basic Geographic Information	National Basic Geographic Information Database with a resolution of 1 million https://www.webmap.cn	2021	Include water, settlements and facilities, transportation, pipelines, administrative divisions, landform and soil, vegetation
NDVI	NASA EOSDIS Land Processes Distributed Active Archive Center [32]	2023	/
SinoLC-1	State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing at Wuhan University [33]	2023	/
OSM	Open Street Map Database https://www.openstreetmap.org	2023	Include 158 rail properties and 7,054 road properties
POI	AMap web API	2023	Include 76,074 properties
Building Geometry	https://doi.org/10.5281/zenodo.11397014 [34]	2024	Include 104,029 Polygonal Planes

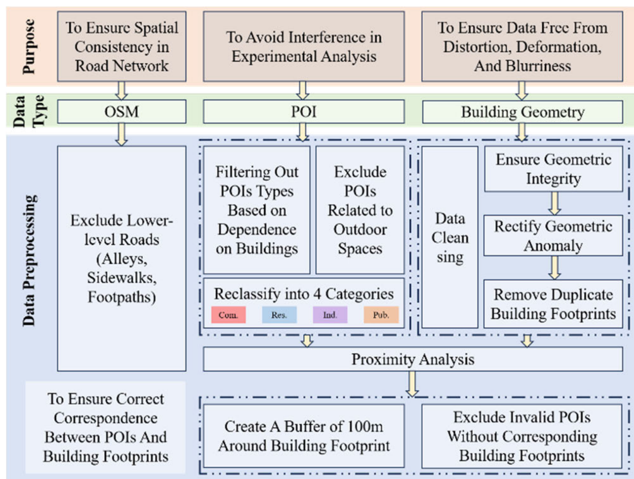

FIGURE 2. Data preprocessing flowchart.

excluded from the raw data to ensure spatial consistency in road network.

For POIs, the initial step involves filtering out POIs based on their dependence on buildings, and excluding points related to outdoor spaces such as golf course, equestrian facility, port and park to avoid interference in the experimental analysis. Subsequently, based on the "Classification of Land Use (GB/T21010-2017) ", POIs are reclassified into 4 major types according to their functions: Commercial Zone (Com.), Residential Zone (Res.), Industrial Zone (Ind.), and Public Service Zone (Pub.). (TABLE 2)

TABLE 2. Reclassification results Of POI data.

UFZs	Level 1 of POIs	Level 2 of POIs
Com.	Hotel accommodation	Economy chain hotel, hotel, hostel, three-star hotel, four-star hotel, five-star hotel
	Food &Beverage	Teahouse, cake & dessert, coffee, foreign food, snack food, Chinese food
	Shopping consumption	Department store, convenience store, supermarket, shopping center, fish-bird-flower market, home appliance, building material, duty-free shop, commercial street, market, sporting good
	Financial institution	ATM, insurance, investment, bank
	Automotive correlation	Charging station, used car, gas station, petrol station, other energy station, auto part, auto repair, auto sales, auto maintenance, car rental, car wash
	Health care	Animal medical, pharmaceutical sale
	Life service	Lottery, telecommunication, barber, photography printing, logistics, laundry, bath, information center, intermediary
	Leisure &Entertainment	KTV, cinema, holiday pension, bar, theater, chess & card, Internet cafe, ~~playground, agritainment~~
	Sports	Bowling, ice sport, fitness center, basketball court, table tennis, billiard room, Taekwondo, tennis court, swimming pool, badminton court, ~~golf course, equestrian facility, campground~~
	~~Tourist attraction~~	~~Red Tour, attraction, aquarium, botanical garden, zoo~~
Res.	Commercial residence	Villa area, commercial & residential building, dormitory, residential area
Ind.	Company &Enterprise	Factory, company
	Commercial residence	Office building, industrial park
Pub.	Science, Education & Culture	Museum, memorial hall, science & technology museum, library, exhibition, adult education, higher education, cultural palace, primary school, art group, kindergarten, vocational & technical education, secondary school, religion, driving school
	Sports &Leisure	Sport complex
	Medical & health care	Emergency center, disease prevention, clinic, specialized hospital, general hospital
	Public service	Radio & television, scientific research institution, public utility, press & publication, archive
	Traffic service	Train station, coach station, ~~subway, port pier, bus station, toll station, parking lot~~
	~~Park Square~~	~~Park, square~~

Note: Underscores represent the types of POI that are removed during data preprocessing

For building geometry data, the first step involves Data Cleansing to ensure the geometric integrity of each building outline, with no missing vertices or disconnected segments. Subsequently, corrections are made for self-intersecting polygons or erroneous vertex coordinates to rectify geometric anomalies. Finally, duplicate building geometry are removed to ensure the data is free from distortion, deformation, and blurriness, which are considered quality issues.

Regarding the spatial deviation issue between POIs and building geometry data, Proximity Analysis are employed in ArcGIS Pro. Based on existing research, POIs within a 100-meter radius of building geometry perimeter are included in the calculations for that building [35]. Additionally, invalid POIs without corresponding building geometry are excluded to ensure the correct correspondence between POIs and building geometry.

III. METHODOLOGY

As shown in Fig.3, the basic framework includes 3 main parts: data collection, refined identification of UFZs, and accuracy assessment. To elaborate, the data collection are as demonstrated in TABLE 1; the specific content of refined identification process is detailed in Section A, which outlines the classification system of UFZs. Sections B and C sequentially present the two identification methods, one relying on buildings and the other not relying on buildings. The methodology for accuracy assessment is presented in Section D.

A. CLASSIFICATION SYSTEM OF URBAN FUNCTIONAL ZONES

We categorize city based on whether UFZs rely on buildings. UFZs based on their dependence on buildings can be divided into 5 major categories: Com., Res., Ind., Pub. and Mixed Zone (Mix.). Non-building-reliant zones can be divided into 2 parts: Traffic Zone (Tra.) and Ecological Zone (Eco.) Ultimately, UFZs are categorized into 7 major classes as outlined above. Specifically, Com. includes spaces comprised of shopping centers, restaurants, banks and similar establishments; Res. encompasses residential communities, dormitories and apartments; Ind. consists of company offices, factory workshops and industrial parks; Pub. involves public facilities, research and educational spaces; Tra. includes railways passing through the city and roads of various classifications; Eco. comprises natural features such as mountains and water as well as various types of green infrastructure; Mix. is defined based on degree of functional blending of different zones [23]. For spaces with ambiguous functional attributes, it can be considered based on the volume of the structure. For instance, the train station has a large volume and serves multiple functions such as commerce, transportation, and public services, so it is more reasonable to define the building volume as Pub. and the railway tracks as Tra.

B. IDENTIFICATION METHOD BASED ON ZONES THAT RELY ON BUILDINGS

1) POI WEIGHT ASSIGNMENT

In general, urban functions with significant influence tend to encompass a larger service area [36]. The relationship between the area of buildings and their functional influence is closely intertwined. Due to the vector nature of POIs, the disparities in the functional area size and radiating influence

TABLE 3. The building area and scoring criteria of various types of POIs.

Building Area Weight / m²	W_1	Level 2 of POIs
0-100	1	Convenience store, fitness, tea, bar, coffee, chess & card, pharmaceutical sale, clinic, community center, barber, logistics, laundry
100-1000	10	Market, gas station, KTV, table tennis, bank, archive, information center, post office, kindergarten, automobile maintenance, company, public utility
1000-3000	30	Department store, basketball court, museum, library, government, press & publication, cinema, tennis, emergency center
3000-5000	50	Shopping mall, dormitory, star hotel, factory, adult education, scientific research
5000-10000	80	Residential area, coach station, holiday pension, specialized hospital, secondary school
>10000	100	Villa, sport complex, hospital, train station, higher education, industrial park, five-star hotel

corresponding to each point cannot be adequately represented. Previous research has established that building area weight and functional influence weight exert a significant impact on POI scoring [22], [23], [37]. We assign scores to these two indicators separately and employ Min-Max Normalization to handle outliers and enhance the stability of the numerical values. Subsequently, a Binary Weight Scoring method is utilized with a 5:5 weight harmonization approach to determine the final weights (W_f) corresponding to various POI. (APPENDIX TABLE).

$$W' = \frac{W - W_{min}}{W_{max} - W_{min}} \tag{1}$$

where W represents the original data, W' denotes the normalized data, W_{max} and W_{min} are the maximum and minimum values in the dataset, respectively.

$$W_f = \frac{W_1{}' + W_2{}'}{2} \tag{2}$$

where W_f represents the final weight for each type of POI, $W_1{}'$ and $W_2{}'$ denotes the normalized value of the building area weight and functional influence weight after the Min-Max Scaling.

a: BUILDING AREA WEIGHT

There is a significant variation in area of different entity objects corresponding to each POI in the city and the variance in scale of POIs leads to differing impacts on building geometry. Due to the limited precision of scoring based on the high-level classification of POIs [22], we determine to start from more granular classifications. First, we calculate the average building area of Level-2 of POIs based on field research results, and then use it as $\boldsymbol{W_1}$ for scoring. The building area of POIs and the six-category scoring criteria are shown in TABLE 3.

b: FUNCTIONAL INFLUENCE WEIGHT

Urban functional influence refers to the area served by the urban functions provided by a particular region or facility [36]. Public service amenities like train stations exert a substantial influence on the entire urban populace, while the impact of standalone dormitories is confined to their immediate vicinity. This discrepancy results in a marked divergence in public awareness between these two types of urban entities. Scholars have previously utilized the "Likert 5-point scale method" for questionnaire surveys [38], but the quantified data from this method often overlooks the uncertainty of influencing factors, and the assessment results may be affected by subjects' expectations of societal norms. Subsequent studies have employed city landmark influence values for weight adjustment [39], overcoming the limitations of subjective cognition towards specific buildings. Additionally, some researchers have utilized topological and statistical indicators to evaluate the service areas on a planar level, objectively measuring the influence of various types of buildings [36]. This paper draws on the methods and data from existing research and further applies the Delphi Method for expert scoring [40]. It integrates the scores from 20 authoritative and representative experts familiar with the field, calculates the average score, and conducts multiple rounds of opinion solicitation, feedback, and revision to ultimately determine the functional influence weight values $\boldsymbol{W_2}$ for Level 2 of POIs. This approach reduces the impact of initial subjective judgments and iteratively converges towards objectivity and accuracy.

2) EMPIRICAL BAYESIAN KRIGING INTERPOLATION

Residential areas, dormitories, and villa districts typically encompass multiple residential buildings, yet POIs on maps are represented by a single vector point. The discrepancy between the coordinates of POIs and the actual location, coupled with the lack of area characteristics inherent to point vector features, results in many building geometries not containing any POIs. In the case of building geometries without POI distribution, previous studies have often employed methods such as KDA or IDW to extrapolate the values of various POIs, but these methods have certain potential limitations. Specifically, KDA, as a non-parametric method, estimates the probability density function without relying on a specific distribution form, making it more suitable for smoothing continuous data. However, its accuracy is easily constrained by processing parameters such as bandwidth or threshold values [24], [25]. IDW is an intuitive and easily understood interpolation method that is sensitive to outliers but fails to accurately capture complex spatial patterns. It is also limited by the minimum grid cell size and is more appropriate for processing medium to small datasets [12], [41].

Research on Kriging dates back to the 1960s, with the algorithm prototype known as Ordinary Kriging. This regression algorithm performs spatial modeling and prediction

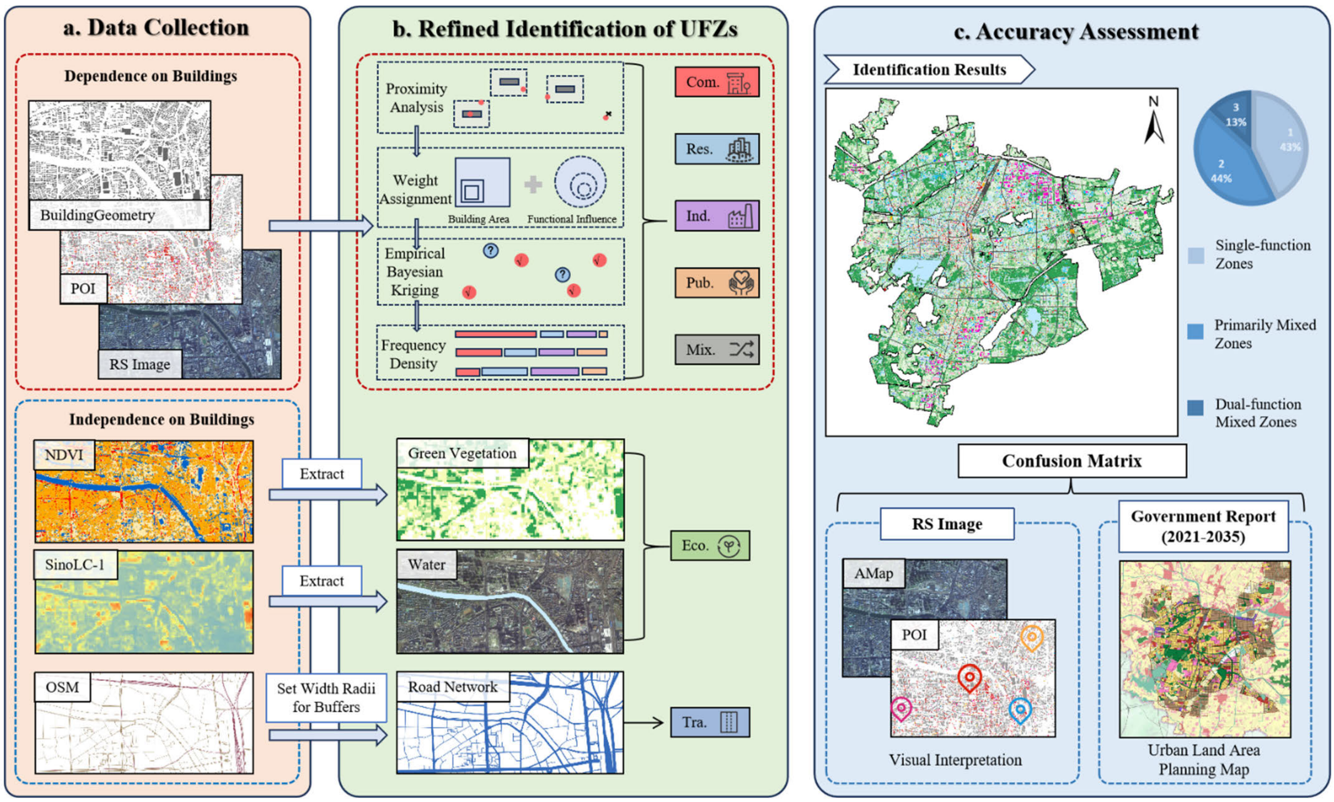

FIGURE 3. Framework for Refined Identification of UFZs (a. Data Collection b. Refined Identification of UFZs c. Accuracy assessment).

of random fields based on covariance functions [42], and is widely applied in geography, environment [43], agriculture [44], mining and meteorological protection [45]. Ordinary Kriging provides optimal linear unbiased estimates but it is characterized by model complexity and requires manual parameter adjustment [46], [47]. Common improvements include Universal Kriging, Co-Kriging, Disjunctive Kriging, etc. [48]. The innovative use of EBK used by few scholars is an interpolation method in geostatistics, which combines Empirical Bayesian statistical methods with Kriging [49]. Unlike other Kriging methods that employ Weighted Least Squares, EBK utilizes Restricted Maximum Likelihood Estimation to determine the parameters of the semivariogram. Empirical Bayesian is an inferential approach that integrates statistical elements with empirical data. It utilizes historical samples to determine Bayesian decision function and estimate prior distribution when the prior is unknown or difficult to specify [50].

Below are presented the fundamental formulas for the semivariogram and the Empirical Bayesian:

$$\begin{cases} \widehat{z_0} = \sum_{i=1}^{n} \lambda_i z_i \\ \min_{\widehat{\lambda_i}} \mathrm{Var}\left(\widehat{z_0} - z_0\right) \\ E\left(\widehat{z_0} - z_0\right) = 0 \end{cases} \tag{3}$$

where λ_i denotes the weight coefficient, which is calculated by the weighted sum of data from all known spatial points to estimate the value at an unknown location. Importantly, these weight coefficients are not simply the inverse of the distances but rather an optimal set of coefficients that ensure the minimal discrepancy between the estimated value $\hat{z_0}$ at the point (x_0, y_0) and the actual value z_0. Additionally, this method must adhere to the condition of unbiased estimation. This approach is recognized as the semi-variance function technique in Kriging methodology.

$$m(x) = \int_{\theta} p(x|\theta)\pi(\theta)d\theta \tag{4}$$

Let $p(x|\theta)$ denote the conditional density of the sample, where the prior distribution $\pi(\theta)$ of the parameter θ is unknown. The marginal density of the sample is denoted by Equation (5), which $m(x)$is also unknown. If we assume that the prior distribution of θ belongs to a known parametric family, denoted as $\pi(x, \lambda)$, then the marginal density of the sample can be expressed as $m(x, \lambda)$. Based on independently and identically distributed historical samples $X_1, \ldots, X_n$ from $m(x, \lambda)$, classical statistical methods can provide an estimate $\hat{\lambda}$ for λ. This estimate allows us to obtain an estimate of the prior distribution $\pi(\theta, \hat{\lambda})$, which can be used as the Bayesian solution in an Empirical Bayesian decision function. This method is known as the parametric Empirical Bayesian method. If the Bayesian solution to the decision problem can be expressed as $\psi(\mathrm{x}, \varphi(\mathrm{m}))$, where ψ and φ are known functions, then $\varphi(\mathrm{m})$ can be estimated using independent samples from $\mathrm{m}(\mathrm{x})$. This allows us to estimate the Bayesian solution $\psi(\mathrm{x}, \varphi(\mathrm{m}))$ and use it as the empirical Bayesian decision function for the decision problem. This approach is known as the non-parametric Empirical Bayesian method.

Therefore, compared to the methods used by existing scholars, EBK demonstrates the following advantages: First, the construction of subsets and simulation processes can automatically compute parameters, reducing the complexity of model building and proving more effective for processing small datasets [51]. Second, the capability to account for the spatial autocorrelation among existing data by integrating local changes and overall trends provides a basis for uncertainty. Third, simulating multiple semivariograms allows for a more flexible adaptation to the spatial variability of data, leading to a more accurate prediction of standard errors [52]. Fourth, with Empirical Bayesian as its theoretical foundation, the key advantage lies in providing estimates of posterior means by aggregating historical data, thus enabling more precise statistical inference within Bayesian framework, and demonstrating robust performance across special types of data and applications [50].

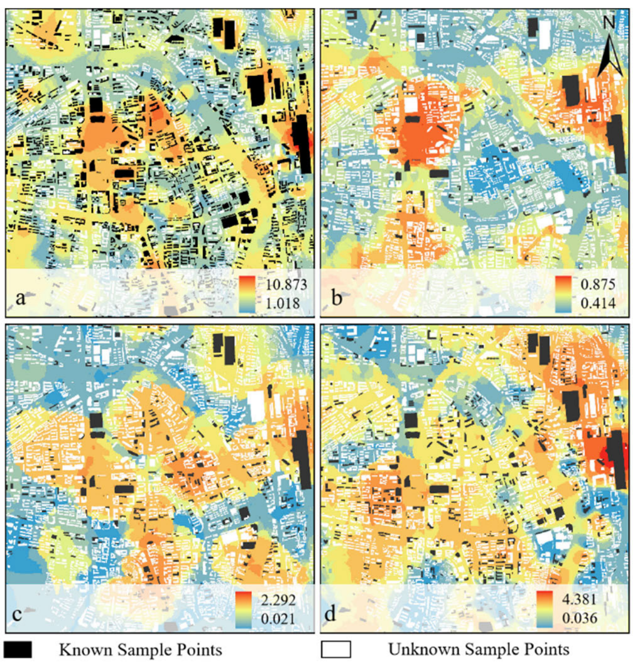

FIGURE 4. EBK interpolation results based on known sample points (a. Com. b. Res. c. Ind. d. Pub).

In ArcGIS Pro, EBK can be executed using Geostatistical Wizard and geoprocessing tools. To elaborate, the input data must initially include x, y, and z coordinates, which are essential for three-dimensional interpolation. Subsequently, subsets are created, with the specification of the number of points per model, with a default of 100 points, to which the software automatically assigns the points into subsets of the designated size. Setting the overlap coefficient between subsets, the larger the value, the smoother the prediction surface, and the longer the processing time. The number of semivariogram simulations for each subset can be set; for input data with larger measurement errors, the number of simulations should be increased. The type of output surface can be chosen from the predicted value surface, the predicted value standard error surface, the probability surface exceeding the critical value, and the quantile surface. For data that is not normally distributed, transformation methods can be selected using the empirical method and the log empirical method. The type of semivariogram can be chosen from linear, power, or thin plate spline functions. Select the appropriate maximum/minimum adjacent element and search radius length based on the input data. Fig. 4 represents the EBK distribution map of various functional values based on known building contours, and then the "Extract Multi Values to Points" spatial analysis tool is used to extract the estimated values to unknown sample points. This map can clearly reflect the changes in the strength of the spatial distribution relationships of these four types of functional values, with higher accuracy.

3) FREQUENCY DENSITY AND CATEGORY RATIO

Given that a single building geometry may encompass multiple types of POIs, it implies that each building has the potential to contain various functional spaces. We utilize two indicators, Frequency Density (FD) and Category Ratio (CR), to identify and assess UFZs that rely on buildings. FD measures the proportion of occurrences of a specific value or within a particular range relative to the total dataset. CR quantifies the relative proportions of different categories within dataset. Therefore, for each building geometry, the types of POIs contained within it can be used to construct FD and CR metrics, which in turn help identify and assess the functional attributes associated with building geometry.

$$\begin{cases} \mathrm{F_i} = \dfrac{\mathrm{n_i}}{\mathrm{N_i}} (\mathrm{i} = 1, 2, 3, \ldots, \mathrm{n}) \\ \mathrm{C_i} = \dfrac{\mathrm{F_i}}{\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{F_i}} (\mathrm{i} = 1, 2, 3, \ldots, \mathrm{n}) \end{cases} \tag{5}$$

where i denotes the type of POIs; n represents the total number of types of POIs; $\mathrm{n_i}$ indicates the number of POIs of type i within a building geometry; $\mathrm{N_i}$ is the total number of POIs of type i in the dataset; $\mathrm{F_i}$ denotes the FD of POIs of type i; and $\mathrm{C_i}$ represents the proportion of FD of POIs of type i relative to the total FD of all POIs types in the building geometry.

Based on the calculated FD and CR, and following the threshold of 0.50 as referenced from previous studies [12], [37], the determination process for specific functional zones is outlined in Fig. 5. The Specific process is described as follows: If any FD of functional value in the building geometry exceeds 0.50, the building is classified as the zone with the specific function; otherwise, it is deemed Mix. When the building geometry possesses one FD of functional value that falls between 0.30 and 0.50, it is identified as Mix. that dominated by the specific function; when one building geometry has two FD of functional value that all range from 0.30 to 0.50, it is determined to be Mix. that dominated by these two specific functions; when there are two more FD of functional value between 0.30 and 0.50, it is classified as a multi-functional Mix.

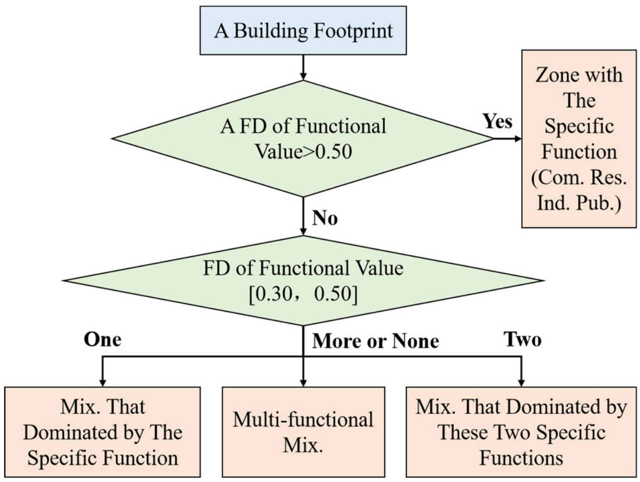

FIGURE 5. The determination process for zones that rely on buildings using FD and CR.

C. IDENTIFICATION METHOD BASED ON ZONES THAT NOT RELY ON BUILDINGS

1) FOR TRAFFIC ZONE

Urban roads in cities carry the main transportation function. OSM, being vector line data, does not possess width attributes. Referred to "Urban Comprehensive Transportation System Planning Standards (GB/T51328—2018) ", different radii for buffers are be set based on the road classification. The relationship between OSM road hierarchy and the buffer zone radius is outlined in TABLE 4. The integrated polygon data formed can be utilized to represent Tra. of city.

2) FOR ECOLOGICAL ZONE

Urban Ecological Zone (Eco.) typically encompasses two major components: water and green vegetation. SinoLC-1 characterize landcover features at a 1m resolution with high accuracy [33], depicting land use types that include forests, grasslands, arable land, wetlands, water, and 11 other categories. Among these, areas with an attribute value of "9" represent regions that are permanently covered by water, including naturally occurring lakes, rivers, and streams, as well as artificially created reservoirs, ditches, water management facilities, ponds, and aquaculture farms. Therefore, it is reasonable to extract these areas from SinoLC-1 as the water in Eco of city.

NDVI is a metric used to assess vegetation coverage and health. It is calculated using the difference or ratio of reflectance values between the near-infrared and red-light bands, with values typically ranging from -1 to 1 [32].

TABLE 4. The relationship between OSM road hierarchy and the radius of buffer zones.

Road Hierarchy	Road Width/m	Buffer Radius/m
Railway	20	10
Motorway	30	15
Primary Road	40	20
Secondary Road	20	10
Tertiary Road	14	7
Cycleway /Footway /Pedestrian	8	4

TABLE 5. NDVI Index classification by natural breaks method.

NDVI Range	Type
[-0.28，0.015]	Water
[0.015，0.14]	Built-up Area
[0.14，0.18]	Barren Land
[0.18，0.27]	Shrub and Grassland
[0.27，0.36]	Sparse Vegetation
[0.36，0.74]	Dense Vegetation

$$\mathrm{NDVI} = \frac{\mathrm{NIR} - \mathrm{Red}}{\mathrm{NIR} + \mathrm{Red}} \tag{6}$$

where NIR represents the reflectance in the near-infrared band, and Red represents the reflectance in the red-light band. Positive NDVI values indicate the presence of vegetation, with the value increasing with greater vegetation coverage; a value of 0 suggests the presence of rock or bare soil, where NIR and Red reflectance are nearly equal; negative values indicate surface cover such as water or snow, which reflects visible light strongly. Referring to the studies of existing scholars and based on Natural Breaks Method for detailed classification as shown in TABLE 5 [53], [54], we can identify and demarcate the urban green vegetation of Eco. by filtering data with NDVI values ranging between 0.18 and 0.74.

D. CONFUSION MATRIX

Confusion Matrix is a tool used to evaluate the accuracy of a classification model by depicting the relationship between the true attributes of sample data and the predicted results in matrix form. It can be represented as an N × N dimensional CM (C, M) [16].

$$\mathrm{CM}\left(\mathrm{C}, \mathrm{M}\right) = \begin{bmatrix} \mathrm{cm}_{11} & \mathrm{cm}_{12} & \ldots & \mathrm{cm}_{1i} & \ldots & \mathrm{cm}_{1N} \\ \mathrm{cm}_{21} & \mathrm{cm}_{22} & \ldots & \mathrm{cm}_{2i} & \ldots & \mathrm{cm}_{2N} \\ \vdots & \vdots & & \vdots & & \vdots \\ \mathrm{cm}_{i1} & \mathrm{cm}_{i2} & \ldots & \mathrm{cm}_{ii} & \ldots & \mathrm{cm}_{iN} \\ \vdots & \vdots & & \vdots & & \vdots \\ \mathrm{cm}_{N1} & \mathrm{cm}_{N2} & \ldots & \mathrm{cm}_{Ni} & \ldots & \mathrm{cm}_{NN} \end{bmatrix} \tag{7}$$

where each column of CM represents the predicted categories, with the sum of each column indicating the number of instances predicted as that category; each row represents the true categories of the data, with the total number in each row indicating the count of instances that are truly of that category. For example, cm_{11}represents instances where the true class is Type 1 and the predicted class is also Type 1, while cm_{12}represents instances where the true class is Type 1 but the predicted class is Type 2. Confusion Matrix is commonly used to evaluate the accuracy of classification results using Overall Accuracy (OA) and the Kappa coefficient. The formulas are as follows:

$$\begin{cases} OA = \sum_{i=1}^{n} \frac{X_{ii}}{N} (i = 1, 2, 3, \cdots\cdots, n) \\ Kappa = \frac{\left(N \sum_{i=1}^{n} X_{ii} - \sum_{i=1}^{n} X_{i+} X_{+i}\right)}{N^2 - \sum_{i=1}^{n} X_{i+} X_{+i}} \\ \qquad (i = 1, 2, 3, \cdots\cdots, n) \end{cases} \tag{8}$$

where N represents the total number of samples, n denotes the number of rows and columns in the confusion matrix, X_{ii} is the number of samples in the i-th row and i-th column of CM, and X_{i+} and X_{+i} represent the total number of samples in the i-th row and i-th column, respectively.

IV. RESULTS AND CONCLUSION

A. ESTIMATION RESULTS OF EBK INTERPOLATION

The EBK method is employed to calculate the unknown spatial values for four types of functional areas and Fig. 6 illustrates the distribution characteristics of these functional spaces. Specifically, driven by economic agglomeration effect, Com. typically exhibits a pronounced clustering tendency, forming dense, small-scale agglomeration points in central urban areas. Res. displays a clustered distribution pattern in the central region, while areas farther from the city center, with comparatively lower land costs, tend to develop into larger residential clusters. Small-scale Ind. are inclined to cluster in the central regions, taking advantage of the economic vibrancy and accessibility. In contrast, Ind. that may generates environmental pollution or noise are strategically distributed on a larger scale around the periphery of the central urban area to mitigate any adverse effects on residential areas. Pub. which provides essential services to the community are predominantly situated in proximity to Res.

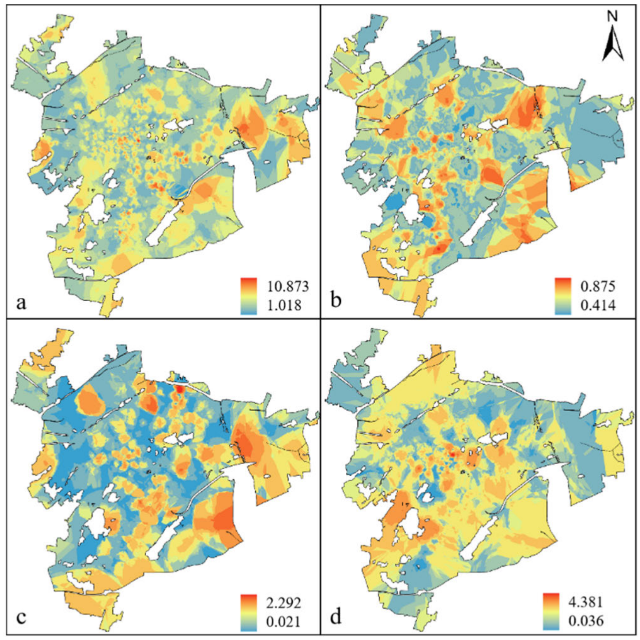

FIGURE 6. EBK interpolation results based on known POI (a. Com. b. Res. c. Ind. d. Pub).

B. IDENTIFICATION RESULTS OF URBAN FUNCTIONAL AREAS

1) RESULTS OF OVERALL CLASSIFICATION

Based on refined identification method for UFZs we proposed, both Proximity Analysis and EBK interpolation method account for the spatial correspondence between POIs and building geometries, addressing the issue of POIs not being entirely distributed in building geometries. Within the study area, 103,615 out of 104,092 building geometries were identified with UFZs, achieving a recognition rate of 99.5%, which is a significant improvement over traditional methods of building functional zones identification. By using OSM to set up Tra. with different levels and NDVI with the Natural Breaks Method to effectively extract Eco. in the city that do not rely on buildings have also been efficiently identified, enhancing the comprehensiveness of urban space recognition.

As shown in Fig. 7, the refined classification results of UFZs are as follows:

(1) Tra., which plays a crucial role in residents' travel, accounts for 9.98% of the total area of main central area of Xuzhou, with accurate position and width that generally match; Eco., including parks, green spaces, nature reserves, etc., accounts for 79.76% of the total area, with the remaining 10.26% of the area being the 5 types of UFZs that rely on buildings.

(2) Among the 5 types of UFZs that rely on buildings, Com., Res., Ind., Pub. and Mix. in main central area of Xuzhou account for 14.67%, 47.79%, 17.95%, 5.92%, and 13.26% of total area, respectively. This means that nearly half of the city's buildings is primarily for residential use, and nearly a fifth is for industrial purposes, which is consistent with Xuzhou's past as a well-known resource-based city and its current gradual transformation. The significant presence of Com. and Ind., along with the rational distribution of Pub. and Mix., together form the city's diversified spatial structure.

(3) According to the specific distribution of UFZs that rely on buildings, Com. (Fig. 7a) are mostly distributed in a point or strip pattern, typically situated at the periphery of urban block and forming commercial podiums that enclose Res. They are strategically positioned near traffic arteries or street corners, aligning with the general prerequisites for commercial development, including convenient transportation, high population density, strong accessibility, and extensive service reach. The epicenter of such commercial activity is often concentrated in the vibrant heart of the city, specifically in the central business district. Res. and

predominantly residential Mix. (Fig. 7b) are the most widely distributed UFZs in the city, accounting for 40-60% of urban area, generally aggregated within the blocks enclosed by roads with strong privacy, well-developed infrastructure and good ecology. Both residential communities and urban villages conform to the characteristics of residential space cluster distribution in most cities. Ind. (Fig. 7c) include companies, factories and industrial parks. Due to the large area required by industrial factories, which may cause pollution to surrounding environment, most enterprises set them up on the outskirts of the city to reduce costs, forming clustered layouts, which is in line with Weber's theory of industrial location. Xuzhou is dominated by east winds, and the city's industrial space is noticeably concentrated in the northeastern and southern parts of the city, located in the downwind direction, which does not cause air pollution. Pub. (Fig. 7d) include stations, hospitals, schools, etc., which need to consider the convenience of coverage, and are mostly scattered and evenly distributed to meet basic service radius requirements of population. However, Pub. such as government offices are generally independent buildings and are more concentrated in old city.

(4) Buildings identified as single-functional zones, primarily mixed-use zones and dual-function mixed-use zones account for 42.62%, 43.69% and 13.25% of the total building area, respectively. This means that more than half of urban buildings has the attribute of Mix., indicating that the mixed use of UFZs is a relatively common phenomenon, which is conducive to mitigating urban sprawl, reducing traffic pressure and improving land use efficiency. Etymologically speaking, Mix. refer to a single building bearing multiple types of functions in the vertical direction, achieving the superposition of multiple functions within a limited space, maximizing the use of land resources, and reducing the pressure of urban expansion. In the main central area of Xuzhou, Mix. dominated by residence, commerce-residence, residence-industry and residence-facility account for 42.63%, 1.67%, 1.21%, and 9.90% of the total building area, respectively, with the remaining Mix. accounting for a smaller proportion. Specifically, Mix. dominated by residence and commerce-residence are mostly distributed in the more prosperous areas of the city center, such as catering services at the bottom and residential or office space above, which helps to achieve multifunctionality and flexibility of space with high spatial utilization efficiency; Mix. dominated by residence-industry break the traditional boundaries between Res. and Ind. in urban planning, helping to reduce citizens' commuting time and improve quality of life, mostly planned in the city's high-tech development zones; Mix. dominated by residence-facility help to improve quality of life for residents and promote sustainable community development, with large areas and mostly distributed around residential space clusters, mainly due to the radiation effect of public service functions such as community centers, health stations and kindergartens located around Res. Multifunctional Mix. usually refer to areas that integrate multiple functions mostly located near transportation hubs, urban renewal areas, educational research institutions, and policy-guided areas.

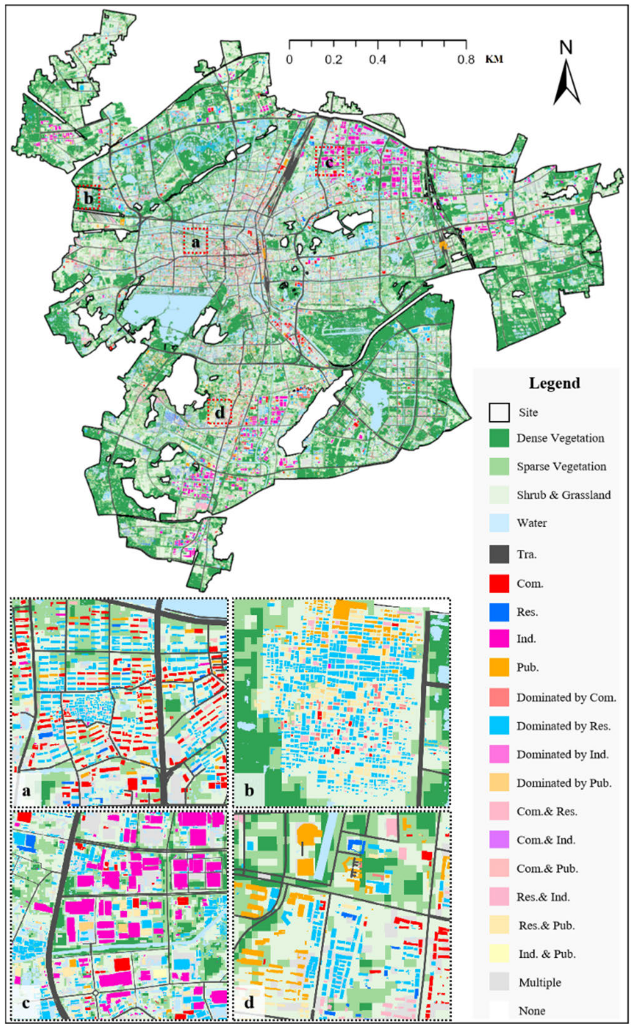

FIGURE 7. Refined Identification Results of UFZs (a.Com. b. Res. c. Ind. d. Pub.)

2) RESULTS OF LOCAL CLASSIFICATION

As illustrated in Fig. 8, we randomly select a block in study area and compare it with the current situation shown on AMap. The map delineates that the block is composed of the Fuli City residential community to the north, the State Grid to the south, and an array of shops to the east. Since the block is predominantly residential, traditional identification methods would simplistically classify the entire area under Res. category, which lacks the granularity. However, by applying the refined identification method outlined earlier, we can assign specific functional attributes to both built and non-built spaces within the block. On the eastern side of the block, buildings that house a variety of eateries and hotels are designated as Com.; the central buildings collectively form a residential complex, thus earning the classification of Res.; the podium of the building in the north of the residential area has a community center, leading to its identification as

Mix. dominated by residential-public service; the State Grid in the south is a row of buildings, mainly identified as Pub., with the remaining ancillary areas classified as Mix. based on EBK interpolation, which may occasionally introduce minor inaccuracies. Overall, the functional spaces within the block exhibit a notable degree of inter-connectivity, meaning that in addition to Res., the block typically encompasses supporting infrastructure such as supermarkets and service centers, which are integral to the community's daily life and overall functionality. Jiefang Road to the east of the block stands as a pivotal urban arterial, spanning 30 meters in width and accommodating bidirectional traffic across three lanes, while Fengming Road to the south functions as a secondary urban thoroughfare, measuring 15 meters wide with two lanes for bidirectional traffic. The remaining roads are narrower, typically under 10 meters wide, and serve as block roads, with Tra. identification being relatively accurate. The entire Res. boasts a high rate of greenery, characterized by a rich diversity of vegetation types, so the remaining spaces are identified as Eco.

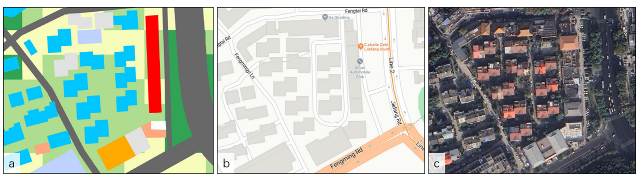

FIGURE 8. Comparison of classification results within a block with the actual situation on AMap (a. Classification results by method in the paper b. Road network map on AMap c. Current situation on satellite AMap).

3) ACCURACY ASSESSMENT

Confusion Matrix is established to quantitatively assess the accuracy of the identification method. A total of 500 points are randomly selected from the study area, and the types of real UFZs are manually determined and the accuracy of the identification is then evaluated using Confusion Matrix. The actual UFZs of the random points are determined based on visual interpretation of high-resolution RS images, inquiries on AMap and combined with on-site surveys. The results indicated that OA of this method reached 84.4%, with a Kappa coefficient of 0.804. To ensure the innovative application of the EBK interpolation in this paper, KDA and IDW interpolation methods are compared in the study area for identification and accuracy assessment. Fig. 9 represents the OA of these two methods, which are 57.5% and 81.2%, respectively. These results suggest that the research method presented in this paper has improved the accuracy of UFZs identification, with a higher degree of conformity to the actual functions of buildings, particularly in the enhancement of Ind. and Mix. Additionally, Kappa value has improved, indicating that the identification function is almost perfect.

As shown in TABLE 6 and Fig. 10, the error matrix of classified UFZs reveals that the sensitivity of experimental method for Com., Res., Ind., Pub., Tra., Eco. and Mix. are 80.6%, 73.9%, 72.7%, 76.1%, 87.0%, 96.4% and 78.3%,

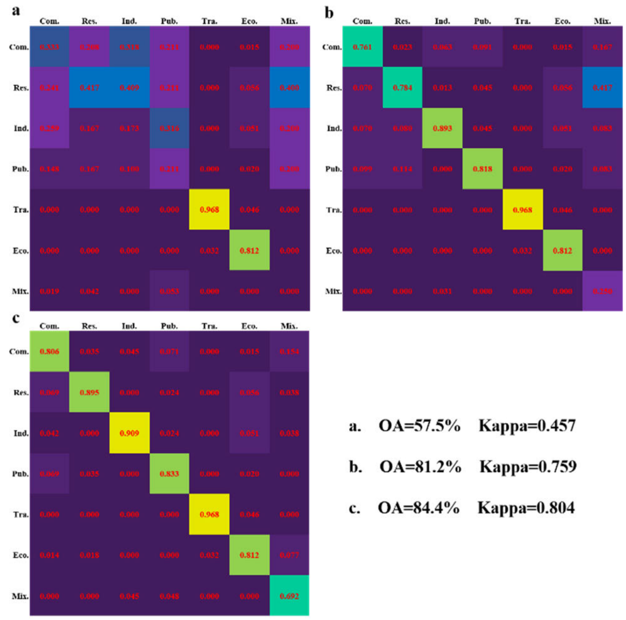

FIGURE 9. Comparation of confusion matrix with different interpolation methods (a. KDA; b. IDW; c. EBK).

TABLE 6. Comparison of automatic recognition and manual interpretation.

Manual Interpretation	Automatic Recognition						
	Com.	Res.	Ind.	Pub.	Tra.	Eco.	Mix.
Com.	58	2	2	3		3	4
Res.	5	51		1		11	1
Ind.	3		40	1		10	1
Pub.	5	2		35		4	
Tra.					60	9	
Eco.	1	1			2	160	2
Mix.		1	2	2			18
Sensitivity (%)	80.6	73.9	72.7	76.1	87.0	96.4	78.3
Accuracy (%)	80.6	89.5	90.9	83.3	96.8	81.2	69.2
OA(%)	84.4						
Kappa	0.804						

respectively; the accuracies are 80.6%, 89.5%, 90.9%, 83.3%, 96.8%, 81.2% and 69.2%, respectively. These results indicate that in the refined identification of UFZs, the method presented in the paper has the strongest capturing ability for Eco., followed by Tra. The identification accuracies for Ind. and Tra. are the highest, both exceeding 90%. The accuracies for Com., Res., Pub. and Eco. are relatively high, while the accuracy for Mix. is comparatively less significant.

Fig. 11 delineates instances of misclassifications stemming from the methodology we proposed. The initial misalignment between Tra. and imagery in Fig. 11a and b can be attributed to the establishment of six-tier buffer zones for categorizing myriad road types from OSM. Subsequently, the omission and incompleteness of buildings in Fig. 11e and h are a direct consequence of interface web API limitations affecting the architectural geometry data. These two factors are primarily

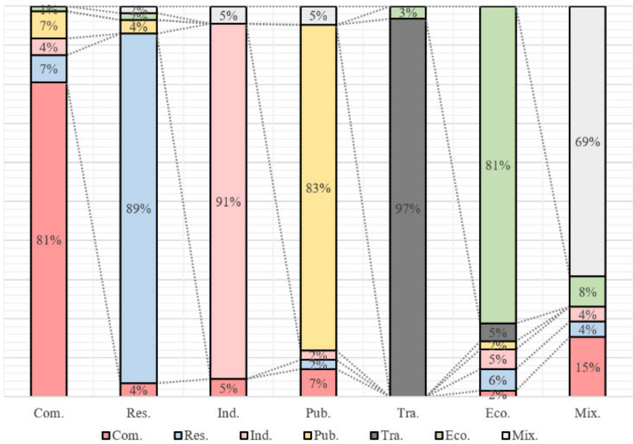

FIGURE 10. Error result of 7 types of UFZs.

responsible for the misclassification of building-dependent UFZs as non-building-dependent UFZs.

Furthermore, an in-depth analysis reveals the causes behind the misclassifications as other building-dependent UFZs. The sporadic identification of small-scale buildings on vacant land in Fig. 11c is a result of the outdated updates to building geometry data. The train station identified as a space encompassing Pub., Res. and Com. in Fig. 11b arises from the extraction process, which fragments large structures into multiple smaller entities. The misidentification of a single instance in an urban village, depicted in Fig. 11d, as a Mix. is due to buffer threshold may lead to erroneous classification or exclusion of POIs in Proximity Analysis. Collectively, these three factors can influence intensity values of specific functions within zoning units, thereby amplifying errors during EBK interpolation process. In addition, the functional zoning discerned through manual interpretation of RS imagery and on-site surveys inherently possesses subjectivity, which is especially likely to affect the classification of high-rise buildings with diverse internal functions.

In detailed analysis of specific UFZs relying on buildings, Res. in Fig. 11d may have been misidentified as Com. and Pub. due to the uniform assignment of building area weight scores to "residential areas" that exhibit varying characteristics. Com. similar to Wanda Plaza, depicted in Fig. 11e, are categorized as Res., Ind. or Mix. owing to the limitations of 2D zoning vertical projections, which fail to fully represent buildings with lower commercial podiums and upper residential or office spaces. In Fig. 11f, some small factories are classified as Res. or Mix. due to the internal diversity of functions in large industrial parks with numerous buildings that are susceptible to interference during interpolation. Pub. such as kindergartens and schools which often irregularly occupy a corner or part of a block in Fig. 11g, have their functional boundaries inaccurately defined by interpolation methods, potentially leading to the misidentification of Pub. as Res.

In UFZs not relying on buildings, the identification accuracy of Tra. is notably high. Misclassifications of Eco. can be attributed to two primary factors: Firstly, NDVI imagery is susceptible to the effects of cloud cover and atmospheric conditions, leading to suboptimal index performance. Secondly, the incompleteness of the established classification system results in the misidentification of outdoor land uses, such as squares and playgrounds in Fig. 11g and h, which are supplemented by multi-source data and erroneously classified as aquatic or vegetative Eco.

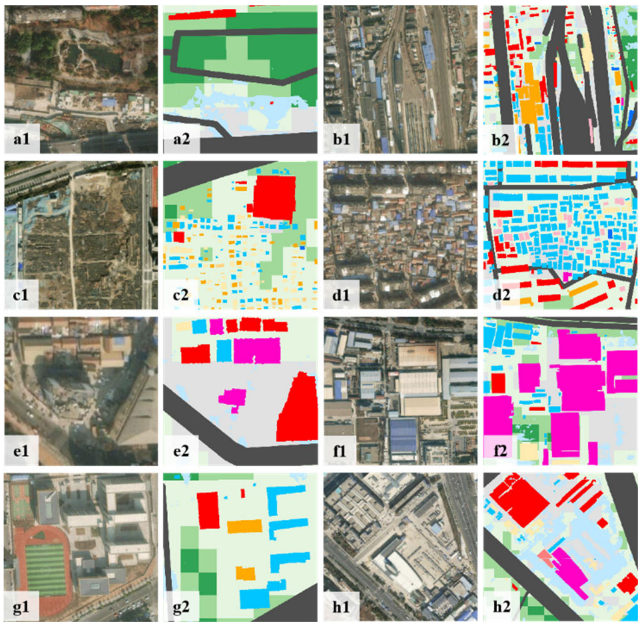

FIGURE 11. Analysis of cases of misjudgment in the process of refine identification (1: Current situation on satellite AMap; 2: Classification results by method in the paper).

V. DISCUSSION AND LIMITATION

The refined identification of UFZs is essential for urban planning, resource allocation, and environmental monitoring, serving as a foundation for promoting sustainable urban development. In the field of urban planning management, identifying UFZs allows for a more precise analysis of public service demands across different areas, thereby optimizing resource allocation. For instance, high-density Res. may require a greater number of educational and medical facilities, while Com. may necessitate more transportation and parking infrastructure. This can assist policymakers in planning public service facilities based on the characteristics of UFZs, enhancing service efficiency and resident satisfaction. Additionally, for urban areas experiencing industrial decline, analyzing their potential for functional transformation can inform urban revitalization strategies, such as converting old industrial areas into Com. or Res. while preserving and reinforcing their historical and cultural values. Furthermore, the identification of UFZs can also aid in environmental monitoring and management. For example, Ind. require stricter pollution control measures, while Res. need to incorporate more green and open concepts in their renewal and design. Moreover, this method is highly adaptable in disaster response planning, traffic network and socio-economic analysis, which helps in the effective allocation of rescue resources, optimization of traffic network design, and provides a rich array of application scenarios and data support for a diversified economic structure.

Therefore, it is crucial to seek efficient methods to replace traditional survey approaches. The effective integration of multi-source data combined with EBK interpolation method offers a novel perspective for addressing this issue. To address challenge such as mismatched research scales, identification accuracy, and incomplete functional zone classification in previous UFZs identification methods, a POI-weighted scoring method is proposed based on EBK, which demonstrates its effectiveness in Xuzhou, China.

The study area, encompassing 40,772 hectares, is spatially categorized into seven major classes: Com., Res., Ind., Pub., Tra., Eco., and Mix. The effective identification rate reached 99.5%, with OA of 84.4% and a Kappa coefficient of 0.804. These figures validate the effectiveness of the innovative identification method proposed in this paper. In addition, this paper also conducts comparative experiments with different interpolation methods, confirming the high efficiency and significant advantages of the EBK interpolation method in the process of estimating unknown values.

While the method presented has great potential in the identification of UFZs, it still has some limitations. For instance, there is a lack of building geometry data, especially in urban fringe areas; the classification system is incomplete, not considering spaces like hard-paved squares; and 3D spatial factors such as building altitude is not account for in POI weight assignment. Therefore, future experimental designs can be optimized in the following aspects: Firstly, the integrity and timeliness of data can be assessed from several aspects such as coordinate location, completeness, and semantic information to improve the quality of data acquisition. Secondly, the diversity of classification system can be expanded in conjunction with urban characteristics. In metropolises with a rich variety of internal functional zones, such as Beijing and Shanghai, decision-makers who want to manage urban more efficiently must further supplement functional classification like cultural education, medical health, public utilities and leisure tourism based on government reports. Additionally, advanced data analysis technologies can be explored in the future, such as deep learning or hybrid machine learning methods [1], [28], [55], [56], integrating real-time data streams [36] like sensor networks or mobile data. By introducing algorithms that incorporate 3D spatial factors [13], [57], the identification method of UFZs can be further automated and optimized. In the coming period, we will conduct identification-related studies on more cities to verify its robustness and adaptability, and explore which relevant factors may influence the distribution of UFZs.

APPENDIX I. APPENDIX TABLE

UFZs	Level 1 of POIs	Level 2 of POIs	W_1'	W_2'	W_f
Com.	Accommodation	Chain hotel	0.2965	0.2965	0.2965
		Hotel & Hostel	0.2965	0.2462	0.2714
		Youth hostel	0.0955	0.1457	0.1206
		Three-star hotel	0.4975	0.2965	0.3970
		Four-star hotel	0.7990	0.3970	0.5980
		Five-star hotel	1.0000	0.4975	0.7487
	Food & Beverage	Teahouse	0.0050	0.0452	0.0251
		Cake & dessert	0.0050	0.0050	0.0050
		Coffee	0.0050	0.0050	0.0050
		Foreign food	0.0050	0.0452	0.0251
		Snack	0.0050	0.0050	0.0050
		Chinese food	0.0050	0.0452	0.0251
	Shopping consumption	Department store	0.2965	0.4975	0.3970
		Convenience store	0.0050	0.0050	0.0050
		Supermarket	0.0955	0.3970	0.2462
		Shopping center	0.4975	0.6985	0.5980
		Fish-bird-flower market	0.0955	0.0452	0.0704
		Appliance	0.0955	0.1457	0.1206
		Building material	0.0955	0.1457	0.1206
		Duty-free shop	0.0050	0.0050	0.0050
		Commercial street	0.4975	0.3970	0.4472
		Market	0.0955	0.3467	0.2211
		Sporting good	0.0050	0.0050	0.0050
		Pharmaceutical sale	0.0050	0.0050	0.0050
	Financial institution	ATM	0.0050	0.0452	0.0251
		Insurance	0.0050	0.0452	0.0251
		Investment	0.0050	0.0452	0.0251
		Bank	0.0955	0.0955	0.0955
	Automotive correlation	Charging station	0.0955	0.1457	0.1206
		Gas station	0.0955	0.1457	0.1206
		Auto part & repair & maintenance	0.0050	0.0452	0.0251
		Car wash	0.0050	0.0452	0.0251
	Life service	Lottery	0.0050	0.0050	0.0050
		Telecommunication	0.0050	0.0452	0.0251
		Barber	0.0050	0.0050	0.0050
		Photography & printing	0.0050	0.0050	0.0050
		Logistics	0.0050	0.0050	0.0050
		Laundry	0.0050	0.0050	0.0050
		Bath	0.0050	0.0452	0.0251
		Intermediary	0.0050	0.0050	0.0050
	Leisure &Entertainment	KTV	0.0955	0.0050	0.0503
		Cinema	0.2965	0.3467	0.3216
		Bar	0.0050	0.0050	0.0050
		Theater	0.2965	0.1457	0.2211
		Chess & card	0.0050	0.0050	0.0050
		Internet cafe	0.0050	0.0452	0.0251
		Fitness	0.0050	0.0050	0.0050
		Basketball court	0.2965	0.0955	0.1960
		Table tennis	0.0955	0.0452	0.0704
		Billiard room	0.0050	0.0452	0.0251
		Taekwondo	0.0050	0.0452	0.0251
		Tennis court	0.2965	0.0955	0.1960
		Swimming pool	0.0955	0.0955	0.0955
		Badminton court	0.0955	0.0955	0.0955
Res.	Commercial residence	Villa	1.0000	0.3970	0.6985
		Dormitory	0.4975	0.1457	0.3216
		Residential area	0.7990	0.1960	0.4975

Ind.	Company & Enterprises	Factory	0.4975	0.0955	0.2965
		Company	0.0955	0.0955	0.0955
		Agriculture & forestry & animal &fishery	0.0955	0.0452	0.0704
		Industrial park	1.0000	0.4472	0.7236
Pub.	Education & Culture	Museum	0.2965	0.7487	0.5226
		Memorial hall	0.2965	0.7487	0.5226
		Science & technology	0.2965	0.5980	0.4472
		Library	0.2965	0.5980	0.4472
		Exhibition	0.0955	0.1457	0.1206
		Adult education	0.4975	0.2965	0.3970
		Higher education	1.0000	0.8995	0.9497
		Cultural palace	0.2965	0.3970	0.3467
		Primary school	0.2965	0.2462	0.2714
		Art group	0.0050	0.1457	0.0754
		Training unit	0.0955	0.2965	0.1960
		Kindergarten	0.0955	0.1960	0.1457
		Vocational & technical education	0.7990	0.3970	0.5980
		Secondary school	0.7990	0.3970	0.5980
		Driving school	0.0955	0.0452	0.0704
		Community Center	0.0050	0.2965	0.1508
		Religion	0.0955	0.2965	0.1960
	Sports	Sport complex	1.0000	0.5980	0.7990
	Medical & health care	Emergency center	0.2965	0.3970	0.3467
		Disease prevention	0.0955	0.2462	0.1709
		Clinic	0.0050	0.0452	0.0251
		Animal medical	0.0050	0.0452	0.0251
		Holiday pension	0.7990	0.2965	0.5477
		Specialized hospital	0.7990	0.4975	0.6482
		General hospital	1.0000	0.9497	0.9749
	Public service	Radio & Television	0.2965	0.3970	0.3467
		Scientific research institution	0.4975	0.2965	0.3970
		Public utility	0.0955	0.0955	0.0955
		Post office	0.0955	0.0955	0.0955
		Press & publication	0.2965	0.2464	0.2714
		Information center	0.0955	0.0955	0.0955
		Archive	0.0955	0.2462	0.1709
	Traffic service	Train station	1.0000	1.0000	1.0000
		Coach station	0.7990	0.9497	0.8744

REFERENCES

[1] W. Xu, J. Wang, and Y. Wu, "Multi-dimension geospatial feature learning for urban region function recognition," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2022, pp. 5832–5835, doi: 10.1109/IGARSS46834.2022.9884450.

[2] M. Zheng, H. Wang, Y. Shang, and X. Zheng, "Identification and prediction of mixed-use functional areas supported by POI data in Jinan City of China," *Sci. Rep.*, vol. 13, no. 1, p. 2913, Feb. 2023, doi: 10.1038/s41598-023-30140-x.

[3] W. Tu, J. Cao, Y. Yue, S.-L. Shaw, M. Zhou, Z. Wang, X. Chang, Y. Xu, and Q. Li, "Coupling mobile phone and social media data: A new approach to understanding urban functions and diurnal patterns," *Int. J. Geographical Inf. Sci.*, vol. 31, no. 12, pp. 2331–2358, Jul. 2017, doi: 10.1080/13658816.2017.1356464.

[4] Y. Zheng, S. Du, X. Zhang, L. Bai, and H. Wang, "Estimating carbon emissions in urban functional zones using multi-source data: A case study in Beijing," *Building Environ.*, vol. 212, Mar. 2022, Art. no. 108804, doi: 10.1016/j.buildenv.2022.108804.

[5] Y. Gu, L. Jiao, T. Dong, Y. Wang, and G. Xu, "Spatial distribution and interaction analysis of urban functional areas based on multi-source data," *Geomatics Inf. Sci. Wuhan Univ.*, vol. 43, no. 7, pp. 1113–1121, Jul. 2018, doi: 10.13203/j.whugis20160192.

[6] Y. Hu and Y. Han, "Identification of urban functional areas based on POI data: A case study of the Guangzhou economic and technological development zone," *Sustainability*, vol. 11, no. 5, p. 1385, Mar. 2019, doi: 10.3390/su11051385.

[7] B. Xue, X. Xiao, J. Li, B. Zhao, and B. Fu, "Multi-source data-driven identification of urban functional areas: A case of Shenyang, China," *Chin. Geographical Sci.*, vol. 33, no. 1, pp. 21–35, Feb. 2023, doi: 10.1007/s11769-022-1320-2.

[8] G. Luo, J. Ye, J. Wang, and Y. Wei, "Urban functional zone classification based on POI data and machine learning," *Sustainability*, vol. 15, no. 5, p. 4631, Mar. 2023, doi: 10.3390/su15054631.

[9] Z. Gao, W. Sun, P. Cheng, G. Yang, and X. Meng, "Identify urban functional zones using multi feature latent semantic fused information of high-spatial resolution remote sensing image and POI data," *Remote Sens. Technol. Appl.*, vol. 36, no. 3, pp. 618–626, Jun. 2021, doi: 10.11873/j.issn.1004-0323.2021.3.0618.

[10] Y. Gao, J. Wang, G. Li, and J. Yan, "Time series dynamic graph embedding: A method for precision identification of urban functions," *J. Geo-inf. Sci.*, vol. 24, no. 10, pp. 1968–1981, Oct. 2022, doi: 10.12082/dqxxkx.2022.220057.

[11] H. Zhou and H. Fan, "Research on the identification of urban function areas and the hot spots based on crowdsourcing geographic data," *Eng. J. Wuhan Univ.*, vol. 55, no. 4, pp. 417–426, Apr. 2022, doi: 10.14188/j.1671-8844.2022-04-013.

[12] Y. Cao, J. Liu, Y. Wang, L. Wang, W. Wu, and F. Su, "A study on the method for functional classification of urban buildings by using POI data," *J. Geo-inf. Sci.*, vol. 22, no. 6, pp. 1339–1348, Jun. 2020, doi: 10.12082/dqxxkx.2020.190608.

[13] Z. Li, W. He, J. Li, and H. Zhang, "Identifying every building's function in large-scale urban areas with multi-modality remote-sensing data," 2024, *arXiv:2405.05133*.

[14] X. Niu, L. Di, X. Song, and Q. Zhang, "Understanding urban spatial structure of Shanghai central city based on mobile phone data," *Urban Planning Forum*, vol. 12, no. 3, pp. 1–13, 2015, doi: 10.3969/j.issn.1000-3363.2014.06.009.

[15] Y. Lin, H. Huang, X. Gu, Y. Xie, J. Li, and C. Wang, "Identification and evaluation of urban public centers from temporal and spatial perspectives: A case study of six districts in Guangzhou," in *Proc. ICCREM*, Dec. 2022, pp. 594–607, doi: 10.1061/9780784484562.065.

[16] J. Li and C. Huang, "Analysis of changes in urban land use in built-up areas of Changsha," *South Archit.*, vol. 12, pp. 19–31, Dec. 2023, doi: 10.3969/j.issn.1000-0232.2023.12.003.

[17] P. Chen, G. Qi, and Y. Zhong, "Progress and perspectives of urban functional region identification," *Bull. Surveying Mapping*, vol. 5, pp. 90–95, May 2024, doi: 10.13474/j.cnki11-2246.2024.0516.

[18] X. Huang, J. Yang, J. Li, and D. Wen, "Urban functional zone mapping by integrating high spatial resolution nighttime light and daytime multi-view imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 175, pp. 403–415, May 2021, doi: 10.1016/j.isprsjprs.2021.03.019.

[19] B. Liu, Y. Deng, M. Li, J. Yang, and T. Liu, "Classification schemes and identification methods for urban functional zone: A review of recent papers," *Appl. Sci.*, vol. 11, no. 21, p. 9968, Oct. 2021, doi: 10.3390/app11219968.

[20] S. Du, S. Du, B. Liu, X. Zhang, and Z. Zheng, "Large-scale urban functional zone mapping by integrating remote sensing images and open social data," *GIScience Remote Sens.*, vol. 57, no. 3, pp. 411–430, Apr. 2020, doi: 10.1080/15481603.2020.1724707.

[21] Q. Qin, S. Xu, M. Du, and S. Li, "Identifying urban functional zones by capturing multi-spatial distribution patterns of points of interest," *Int. J. Digit. Earth*, vol. 15, no. 1, pp. 2468–2494, Dec. 2022, doi: 10.1080/17538947.2022.2160841.

[22] F. Jia, J. Yan, and T. Wang, "Research on scoring evaluation model and functional regions identification constructed by big data," *Sci. Surveying Mapping*, vol. 46, no. 8, pp. 172–178, Aug. 2021, doi: 10.16251/j.cnki.1009-2307.2021.08.024.

[23] X. Hu, X. Li, and F. Wei, "A POI data-based study of the urban functional areas of Chongqing and their mix degree recognition," *J. Southwest Univ., Natural Sci. Ed.*, vol. 43, no. 1, pp. 164–173, Jan. 2021, doi: 10.13718/j.cnki.xdzk.2021.01.019.

[24] R. Xu, Z. Chen, F. Li, and C. Zhou, "Identification of urban functional zones based on POI density and marginalized graph autoencoder," *ISPRS Int. J. Geo-Inf.*, vol. 12, no. 8, p. 343, Aug. 2023, doi: 10.3390/ijgi12080343.

[25] C. Tang, J. Chen, Y. Xie, J. Tang, and C. Zhou, "Integrated identification and spatial differentiation mechanism of urban built-up area in Changsha-Zhuzhou-Xiangtan urban agglomeration based on the multi-source big data," *Econ. Geography*, vol. 44, no. 1, pp. 66–76, Jan. 2024, doi: 10.15957/j.cnki.jjdl.2024.01.007.

[26] S. Wei and J. Pan, "Identification of urban spatial structure in Zhengzhou city based on nighttime light and microblog check-in data," *Remote Sens. Technol. Appl.*, vol. 37, no. 3, pp. 771–780, Jan. 2022, doi: 10.11873/j.issn.1004-0323.2022.3.0771.

[27] S. Ma, J. Zhang, X. Chen, and G. Liao, "Identifying urban functional structures using time-series taxi data," *J. Jilin Univ., Eng. Technol. Ed.*, vol. 22, no. 2, pp. 1–14, Oct. 2023, doi: 10.13229/j.cnki.jdxbgxb.20230447.

[28] Z. Wang, J. Bai, and R. Feng, "A multi-feature fusion method for urban functional regions identification: A case study of Xi'an, China," *ISPRS Int. J. Geo-Inf.*, vol. 13, no. 5, p. 156, May 2024, doi: 10.3390/ijgi13050156.

[29] S. Chen, H. Zhang, and H. Yang, "Urban functional zone recognition integrating multisource geographic data," *Remote Sens.*, vol. 13, no. 23, p. 4732, Nov. 2021, doi: 10.3390/rs13234732.

[30] J. Yang, "Urban Pulse: Research on urban dynamic structure based on multi-source big data," *Planners*, vol. 21, pp. 64–71, Sep. 2020.

[31] C. Fang, L. Zhou, X. Gu, X. Liu, and M. Werner, "Data-driven city: An innovative approach to urban area delineation," Jun. 2024, doi: 10.21203/rs.3.rs-4642145/v1.

[32] K. Didan, "MOD13A3 MODIS/Terra vegetation indices monthly L3 global 1km SIN grid V006," NASA EOSDIS Land Processes Distrib. Act. Arch. Center, Tech. Rep., 2024, doi: 10.5067/MODIS/MOD13A3.006.

[33] Z. Li, W. He, M. Cheng, J. Hu, G. Yang, and H. Zhang, "SinoLC-1: The first 1 m resolution national-scale land-cover map of China created with a deep learning framework and open-access data," *Earth Syst. Sci. Data*, vol. 15, no. 11, pp. 4749–4780, Oct. 2023, doi: 10.5194/essd-15-4749-2023.

[34] Y. Che, X. Li, X. Liu, Y. Wang, W. Liao, X. Zheng, X. Zhang, X. Xu, Q. Shi, J. Zhu, H. Zhang, H. Yuan, and Y. Dai, "3D-GloBFP: The first global three-dimensional building footprint dataset," *Earth Syst. Sci. Data*, vol. 16, no. 11, pp. 5357–5374, Nov. 2024, doi: 10.5194/essd-16-5357-2024.

[35] B. Xue, B. Zhao, and J. Li, "Evaluation and enhancement methods of POI data quality in the context of geographic big data," *Acta Geographica Sinica*, vol. 78, no. 5, pp. 1290–1303, May 2023, doi: 10.11821/dlxb202305014.

[36] T. Pei, X. Yan, H. Shu, X. Wang, X. Chen, S. Guo, J. Jiang, C. Song, and J. Chen, "Geographical flows: A fresh perspective on quantifying urban function," *Urban Informat.*, vol. 3, no. 1, p. 14, May 2024, doi: 10.1007/s44212-024-00044-w.

[37] G. Zheng, Y. Qiao, and S. Sun, "Research on the identification of urban functional areas based on POI data," *Geospatial Inf.*, vol. 21, no. 10, pp. 58–61, Oct. 2023. [Online]. Available: https://kns.cnki.net/kcms2/article/abstract?v=-f9lWFCLl9A-bVGbZRQAxVRYhV5NmlPgvOpOauVBz6ogh9SYugKzLnc8RcCmhNvMyBUxymOOhEQ0tgpLskBR2dhW6F-3LiBgPpoDmqcwRa0MczfZzbP07ur3j5ZwDxAI9A6lw-fqKINPEeAotW46PR9cMyO5lG3v1VR5IsRgxuPKJnUOOP0OWQ==&uniplatform=NZKPT&language=CHS

[38] W. Zhao, Q. Li, and B. Li, "Extracting hierarchical landmarks from urban POI data," *J. Remote Sens.*, vol. 15, no. 5, pp. 973–988, 2011, doi: 1007-4619(2011)05-973-16.

[39] J. Wu and S. Wei, "Research on identification system of urban functional area based on mixed sampling," *Intell. Comput. Appl.*, vol. 14, no. 1, pp. 185–190, Jan. 2024.

[40] G. Ismail and N. Taliep, "The Delphi method," in *Handbook of Social Sciences and Global Public Health*. Cham, Switzerland: Springer, May 2023, pp. 1–19, doi: 10.1007/978-3-030-96778-9_66-2.

[41] R. Zhang, Y. Zhang, J. Guo, Q. Feng, and T. Liang, "Comparison of spatial interpolation methods for precipitation distribution in Xin Jiang region," *Pratacultural Sci.*, vol. 35, no. 3, pp. 521–529, Mar. 2018, doi: 10.11829/j.issn.1001-0629.2016-0608.

[42] J. Chilès and N. Desassis, "Fifty years of Kriging," in *Handbook of Mathematical Geosciences*. Cham, Switzerland: Springer, Jun. 2018, pp. 589–612, doi: 10.1007/978-3-319-78999-6_29.

[43] H. Gao, Z. Yong, J. Liu, X. Cheng, S. Guo, Y. Jiang, and H.-X. Tan, "Spatial and temporal distribution and source variation of heavy metals in cultivated land soil of Xiangzhou district based on EBK interpolation prediction and GDM Model," *Environ. Sci.*, vol. 43, no. 11, pp. 5180–5191, Nov. 2022, doi: 10.13227/j.hjkx.202202068.

[44] V. P. Samsonova, Y. N. Blagoveshchenskii, and Y. L. Meshalkina, "Use of empirical Bayesian Kriging for revealing heterogeneities in the distribution of organic carbon on agricultural lands," *Eurasian Soil Sci.*, vol. 50, no. 3, pp. 305–311, Mar. 2017, doi: 10.1134/s1064229317030103.

[45] J. Zhang, Q. Zhang, D. Xu, and Y. Ding, "EBK and LSTM model in the application of meteorological drought time prediction," *Yellow River*, vol. 42, no. 8, p. 6, 2020, doi: 10.3969/j.issn.1000-1379.2020.08.016.

[46] S. Liu, Y. Hu, Q. Yan, J. Ma, and X. Jin, "Application of expert Kriging algorithm in the spatial interpolation," *Geol. Resour.*, vol. 20, no. 4, pp. 292–294, Aug. 2011, doi: 10.13686/j.cnki.dzyzy.2011.04.018.

[47] Y. Tang, Y. Liang, J. Mai, and J. Lin, "Building function recognition based on random forest and multi-source spatial big data," *Intell. Building City Inf.*, vol. 6, pp. 1–12, Jun. 2022.

[48] J. Kleijnen, "10 Kriging: Methods and applications," in *System- and Data-Driven Methods and Algorithms*, Oct. 2021, Oct. 2021, pp. 355–370, doi: 10.1515/9783110498967-010.

[49] M. Zaresefat, R. Derakhshani, and J. Griffioen, "Empirical Bayesian Kriging, a robust method for spatial data interpolation of a large groundwater quality dataset from the western Netherlands," *Water*, vol. 16, no. 18, p. 2581, Sep. 2024, doi: 10.3390/w16182581.

[50] S. Petrone, S. Rizzelli, J. Rousseau, and C. Scricciolo, "Empirical Bayes methods in classical and Bayesian inference," *METRON*, vol. 72, no. 2, pp. 201–215, Aug. 2014, doi: 10.1007/s40300-014-0044-1.

[51] A. Gribov and K. Krivoruchko, "Empirical Bayesian Kriging implementation and usage," *Sci. Total Environ.*, vol. 722, Jun. 2020, Art. no. 137290, doi: 10.1016/j.scitotenv.2020.137290.

[52] M. Lu and Y. Dong, "Research on the difference of DEM interpolation accuracy by semi-variance function model in EBK algorithm," *Surveying Mapping Technol. Equip.*, vol. 1, no. 3, 2018, doi: 10.3969/j.issn.1674-4950.2018.01.007.

[53] J. Ou, W. Luo, and C. Liu, "Research on land use/cover classification based on GF-1 and multi-source data combination," *J. South China Normal Univ., Natural Sci. Ed.*, vol. 51, no. 5, pp. 92–97, Oct. 2019, doi: 10.6054/j.jscnun.2019089.

[54] C. Kwan, D. Gribben, B. Ayhan, J. Li, S. Bernabe, and A. Plaza, "An accurate vegetation and non-vegetation differentiation approach based on land cover classification," *Remote Sens.*, vol. 12, no. 23, p. 3880, Nov. 2020, doi: 10.3390/rs12233880.

[55] R. Cao, W. Tu, C. Yang, Q. Li, J. Liu, J. Zhu, Q. Zhang, Q. Li, and G. Qiu, "Deep learning-based remote and social sensing data fusion for urban region function recognition," *ISPRS J. Photogramm. Remote Sens.*, vol. 163, pp. 82–97, May 2020, doi: 10.1016/j.isprsjprs.2020.02.014.

[56] N. Xu, J. Luo, T. Wu, W. Dong, W. Liu, and N. Zhou, "Identification and portrait of urban functional zones based on multisource heterogeneous data and ensemble learning," *Remote Sens.*, vol. 13, no. 3, p. 373, Jan. 2021, doi: 10.3390/rs13030373.

[57] S. Du, Y. Zhang, W. Sun, and B. Liu, "Quantifying heterogeneous impacts of 2D/3D built environment on carbon emissions across urban functional zones: A case study in Beijing, China," *Energy Buildings*, vol. 319, Sep. 2024, Art. no. 114513, doi: 10.1016/j.enbuild.2024.114513.

XUAN DU was born in Taiyuan, Shanxi, China, in 2000. She received the Bachelor of Engineering degree in architecture from China University of Mining and Technology, Xuzhou, China, in 2023, where she is currently pursuing the master's degree in land resource management.

She participated in rural planning, cultivated land quality assessment, increase and decrease linkage, green energy-saving building design, and other projects. Her research interests include national spatial planning, urban space carbon emissions, architecture, and geographic big data identification. She was awarded the National Scholarship in 2021, the Pingmei Scholarship, and the First-class Scholarship of China University of Mining and Technology in 2020.

YISHA PAN received the master's degree in land resource management from Henan Polytechnic University, Jiaozuo, China, in 2024. She is currently pursuing the Ph.D. degree with China University of Mining and Technology, Xuzhou, China. Her research interests include land use carbon emissions, focusing on the effects of human and non-human factors on carbon emissions from land use.

XIAOYAN YANG received the master's degree in management and the Ph.D. degree in engineering from China University of Mining and Technology, in 2007 and 2018, respectively. She is currently an Associate Professor with Jiangsu Normal University. Her research interests include land use and land management.

LONGGAO CHEN was born in Suzhou, Anhui, China, in 1976. He received the bachelor's degree in economics from Nanjing Agricultural University, in 1999, and the master's degree in management and the Ph.D. degree in engineering from China University of Mining and Technology, in 2005 and 2009, respectively.

He is currently a Professor and the Doctoral Supervisor. He has published more than 70 academic articles in domestic and international journals, such as *Environmental Impact Assessment Review*, *Habitat International*, and *Journal of Applied Engineering Sciences*. He has undertaken over 30 projects, including the National Natural Science Foundation of China, Jiangsu Province Social Science Fund, Jiangsu Province University Natural Science Fund, and various commissioned projects from enterprises and institutions. His research interests include interdisciplinary fields based on remote sensing and GIS spatial information technology, environmental impact assessment in spatial planning, national spatial planning, carbon effects, and the ecological use of land.

LIANGCHEN LIU was born in Huaibei, Anhui, China, in 2000. He received the bachelor's degree in land resource management from Anhui Normal University, Wuhu, China, in 2022. He is currently pursuing the master's degree with China University of Mining and Technology, Xuzhou, China. His research interests include return farmland to forest, national spatial planning, and land use carbon emissions.

YING LIN was born in Sanming, Fujian, China, in 1999. She received the bachelor's degree in land resource management from China University of Mining and Technology, Xuzhou, China in 2022, where she is currently pursuing the master's degree. Her research interests include governance of rural residential areas, national spatial planning, and land use carbon emissions.