Analyzing Land Use Mixing Degree using a Vector Data Cube with Hierarchical Cell: A Case Study of Seoul, the Republic of Korea

Jiyoung Kim*, Hyojoong Kim*, Jiwon Lee*

* Social Eco Tech Institute, Konkuk university

Abstract. The type of urban land use is a very important factor in urban planning. Therefore, to quantitatively calculate the land use mixing degree, this study applied information entropy to point of interest data to perform the time-series analysis of the land use status. The visual comparison of the cell divided according to the land use mixing degree obtained using the proposed method with the current data provided by the map service portal confirmed the significance of the proposed method.

Keywords. Point of Interest (POI), National Address Information, Information Entropy, Grid Partitioning

1. Introduction

The population decline is emphasizing the need for compact cities as a new strategy in urban planning to ensure the current level of sustainability and efficiency in land use. In large cities with such high density and complex land use, various activities can be concentrated and active interactions can occur. This degree of land use mix could influence transport mode choice and shape commuting behaviors. As a variety of land uses for different functions occur within close proximity, non-auto commuting is more encouraged by the mixed land-use (Christian et al. 2011). The ability to precisely map urban land use types can significantly aid urban planning and urban system understanding. To measure the urban function mixing degree, scholars mainly calculate and analyze it from two aspects. The first is in terms of data, and it involves the use of multiple heterogeneous sources to identify and evaluate the construction of urban functional areas (Long et al. 2013, Wu et al., 2018, Kang et al., 2018, Gao et al. 2019). The second is in terms of the research methods, and it involves the use of different meas-

LBS 2023
18th Conference on Location Based Services
November 20-22, Ghent, Belgium

Published in “Proceedings of the 18th International Conference on Location Based Services (LBS 2023)”, edited by Haosheng Huang, Nico Van de Weghe and Georg Gartner, LBS 2023, 20-22 November 2023 Ghent, Belgium.

This contribution underwent single-blind peer review based on the paper.
https://doi.org/10.34726/5703

urement methods to measure the urban function mixing degree, and these methods are mostly objective weighting methods, such as the entropy method, spatial entropy, mean square deviation, or land use mixing index method. These objective evaluation methods avoid the subjective judgments of researchers (Xia et al. 2020).

Previous studies apply one-size units, which are set for the whole study area, and are often inaccurate owing to spatial heterogeneity (Jing et al. 2022). The rapid evolution of cities has created new challenges in urban planning and management. Accordingly, the accurate evaluation of the mixed use of urban areas is critical, particularly at a fine scale (Cui et al. 2020). To achieve this, among various measurement methods for measuring the land use mixing degree, this study applied the information entropy (IE) method to analyze the mixing degree of each cell, and the cell was hierarchically divided according to the land use mixing degree to enable its analysis on a fine scale.

Next, the functional change of the city was examined by analyzing the land use mixing degree in a fine-scale time series. In this process, it is necessary to structure the time-series land use mixing degree generated in hierarchical cells. Therefore, in this study, a data cube recently used in the Earth observation (EO) domain was applied to refer to the time series multi-dimensional (n-D) array. The EO data, a prime source of big data, comprise large-scale and multi-source geospatial data that are acquired from orbital sensors, in-situ measurements, and simulation models (Baumann et al. 2018, Wagemann et al. 2018). An EO data cube can be defined as a time-series multi-dimensional (space, time, and data type) stack of spatially aligned analysis-ready pixels (Nativi et al. 2017). Most EO data cubes are focused on images, and discussions have emerged recently on vector data cubes or geo data cubes that can accommodate geospatial data (Gao et al. 2022, Open Geospatial Consortium 2022).

Therefore, in this study, we proposed a method for analyzing the change in land use in a city by structuring the hierarchically-analyzed land use mixing degree, that is, data cubes with different cell sizes in a time series (Figure 1).

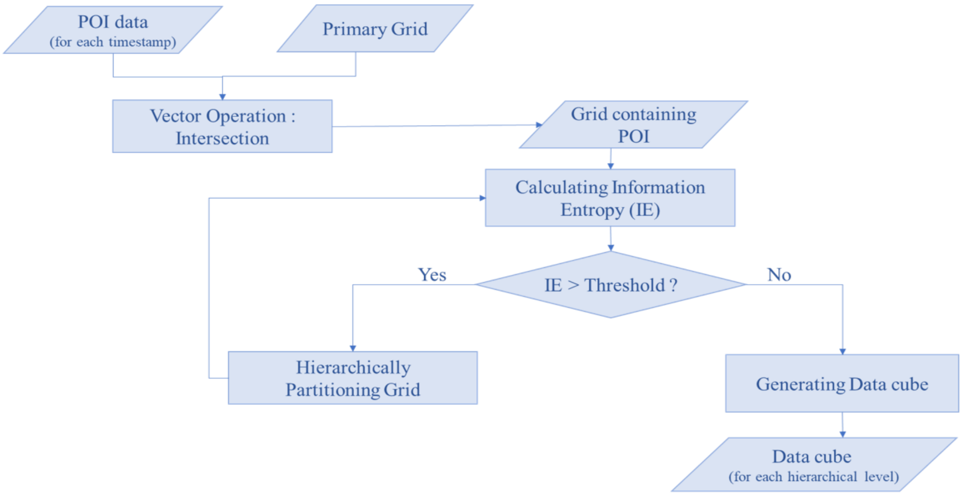


Figure 1. Workflow of this study.

2. Method

2.1. Hierarchical cells partitioning based on IE

2.1.1. IE calculation

IE is a physical concept, introduced by Shannon, to measure the disorder of a system, and to measure the complexity and balance between systems (Guo et al. 2015, Jia et al. 2016, Wei et al. 2018). In the process of analyzing the land use mixing degree, the point of interest (POI) is one of the core elements of the urban structure and form. Therefore, in this study, to calculate the IE, we combined the basic principles of IE with the obtained Seoul POI data. The specific measurement steps are as follows:

[Step 1] Identify the type of POI and count the total number of different POI types. POIs are divided into 10 categories. The total number of POIs of type k in a city is A_K (k = 1, 2, L 10), and the total number of POIs in the city is A ($A = \sum_{k=1}^{10} A_k$).

[Step 2] Calculate the IE in the city. This consists of two parts: measuring the IE of each latitude/longitude cell and measuring the IE of each cell. In this study, the dimension of the latitude/longitude cells was 1000 m × 1000 m, and there was no change in the total number of POI types in each cell. Let be the total number of POIs of type k in cell m, where M is the total number of cells in Seoul, ROK. Accordingly, the proportion of POI type k in the cell can be calculated as expressed in Equation (1):

$$P_k^m = A_k^m / \sum_{k=1}^{10} A_k^m \tag{1}$$

Based on P_k^m, the IE H^m of grid cell m can be calculated as expressed in Equation (2):

$$H^m = -\sum_{k=1}^{10} P_k^m \times \ln(P_k^m) \tag{2}$$

The complexity of the land use of the cell increases with an increase in the IE.

2.1.2. Hierarchical cells partitioning

The design was based on the hypothesis that smaller cells can more accurately represent the land use mixing degree. Therefore, an appropriate cell size should consider the mix of several functional types. The core conception of the partitioning algorithm is iteration partitioning based on the land use mixing degree measurements. According to a previous study on land use and urban planning (Wang et al. 2018), the typical spatial unit that matches most cities is 1000 m. Therefore, 1000 m was set as the initial cell size value. The partition granularity of spatial cells is determined by number of iterations. After the iteration of the model three times, the granularity size was 125 m. As a small size may result in fragments, which lead to lower identification accuracy, the number of iterations was set to 2. Therefore, the size of the smallest unit was 250 m.

The main steps of the partitioning process are as follows:

[Step 1] The first iteration was performed to create primary cells as Level 1. The study area was divided into primary cells with a dimension of W × W m, and was combined with POI data. The function-mixed degree in the primary cells was calculated using the IE. Here, W is the primary granularity.

[Step 2] Another iteration process was performed to sub-divide the upper level. If the mixed degree of the units is greater than the threshold (Th1), the primary units were divided into W/2n × W/2n m, where n is the number of iterations.

[Step 3] The new-level cells are aggregated by aggregating the unpartitioned primary cells in the last level and newly partitioned cells in Step 2. Therefore, they were composed of the new level cells.

[Step 4] Determining the following loop to partition. This validates the iteration criteria by calculating the mixed degree used to determine if the next loop should be performed. Iterating Steps 2 and 3 generates the new level cells.

2.2. Generation of the vector data cube using the hierarchical cells

Here, raster data cubes refer to data cubes with raster (x and y, or longitude and latitude) dimensions, and vector data cubes are n-D arrays that have (at least) a single spatial dimension that can be mapped to a set of vector geometries. To save a hierarchical cell as a data cube, it cannot be made into one data cube because of the difference in the size of the cell. Therefore, in this study, a data cube was created for each cell size (1000, 500, and 250 m), and a feature array was added to classify the hierarchical cells into levels. Therefore, we modeled the vector data cube with 4-D arrays consisting of an X center, Y center of cell, Time, and Feature (Figure 2(a)). Features include cell Id, IE value, level value, and cell size. In addition, to visualize the land use mixing degree analyzed in time series in the hierarchical cells, a vector data cube with cell geometry as an attribute was created (Figure 2(b)).

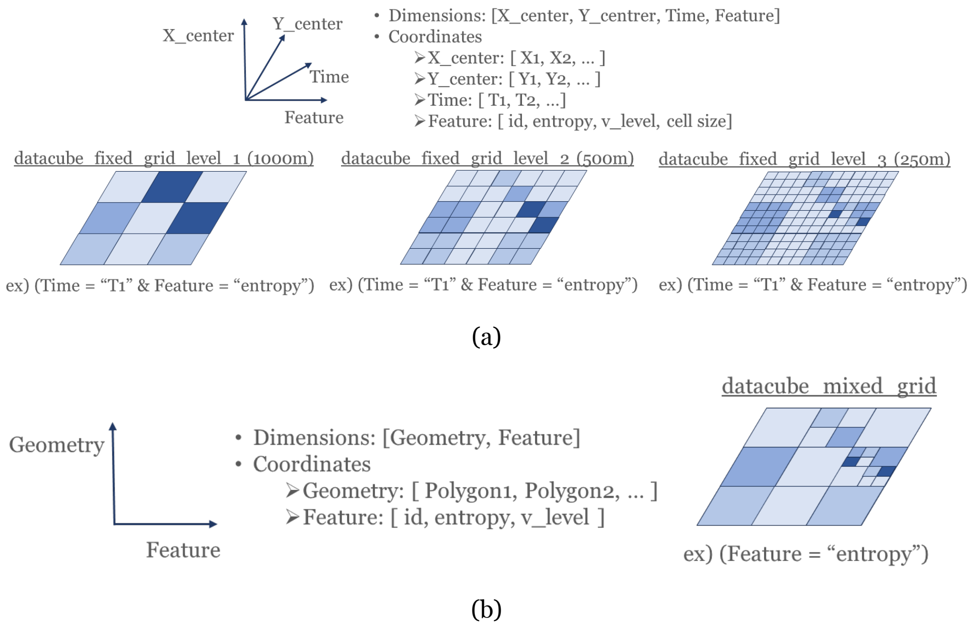


Figure 2. Vector data cube model.

3. Test and Results

3.1. Study area and dataset

The target area of this study was Seoul, the capital of the Republic of Korea (ROK) with an urbanization area ratio of 61.39% (https://data.seoul.go.kr/dataList/569/S/2/datasetView.do) (Figure 3).

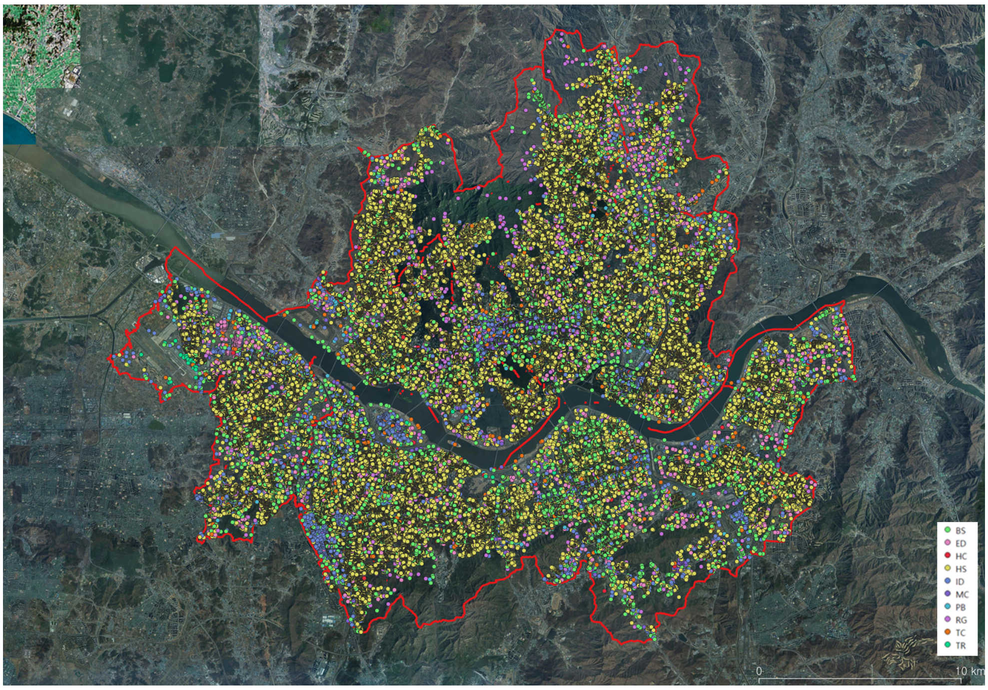


Figure 3. Study area (Seoul, ROK).

The POI data (navigation DB) and the national grid (1000 m) of Seoul provided by the Address information website (https://business.juso.go.kr/addrlink/main.do) were used (Table 1). The navigation DB includes street name address, coordinate values (X, Y) where the address is geocoded to the center of the building, and building (POI) types. Based on this, the POIs to be used for the land use mixing were reclassified into 10 categories (Table 2).

	National grid (1000 m)	POI (Navigation DB)	
Date	2022.05	2021.07	2023.05
Number of features	701	608,325	597,599

Table 1. Data sets (Source).

No.	Category	Details of the POI types
1	Business	Amusement facilities, livestock and fisheries facilities
2	Education	Educational and welfare facilities
3	Health	Hospital
4	Commercial	Catering, large shopping plaza, shopping
5	House	Residential region
6	Industrial	Factories or warehouse facilities
7	Public Service	Government, public library, senior citizen center
8	Religion	Religion
9	Tourism & Culture	Accommodation, Scenic spot, Cultural, tourism and leisure facilities
10	Transportation	Transportation facilities

Table 2. Categories of POI data.

3.2. Results and verification

The IE of the POIs crossing each cell was calculated using the primary grid (1000 m), and if the value is greater than Th1 (mean), the cells are divided, and this process is repeated twice (that is, until the cell size is 250 m). The distribution of the IE after the iteration process is repeated twice is shown in Figure 4, which indicates that the distribution is similar. According to the mean IE values, the POI data for July 2021 were 0.384, 0.394, and 0.390, and the POI data for May 2023 were similar to that of July (0.382, 0.392, and 0.380).

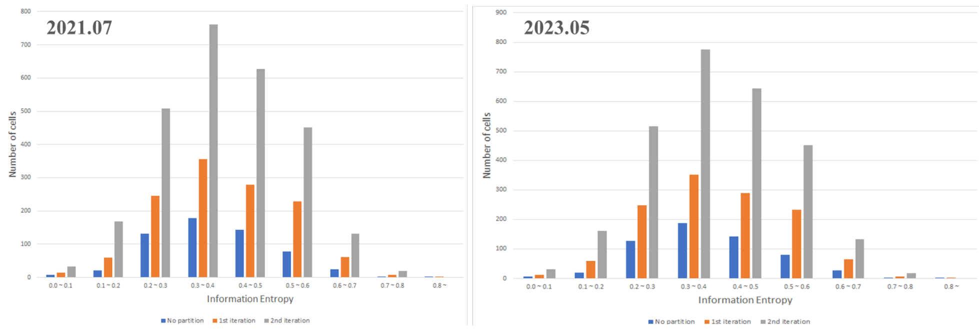


Figure 4. Distribution of the mixed degree.

However, the generated vector data cube was visualized as an array composed of geometry and features (Figure 2(b)), and time series analysis was performed (Figure 5). An increase (red) and a decrease (yellow) in the hierarchical level were observed. With a change in time, a higher level of the grid indicates an increase in the land use mixing degree of the cell, and a lower level indicates a decrease in the land use mixing degree.

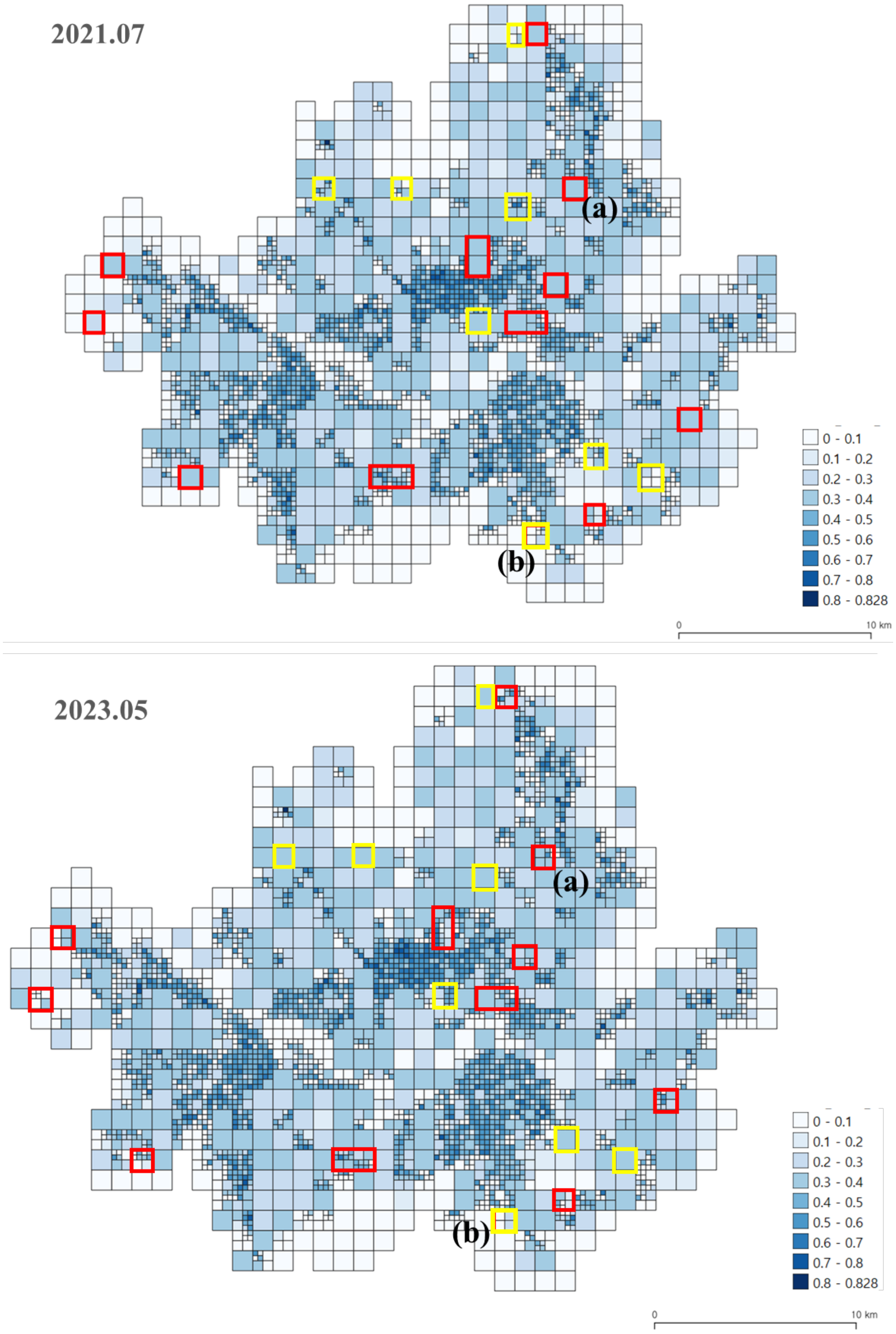
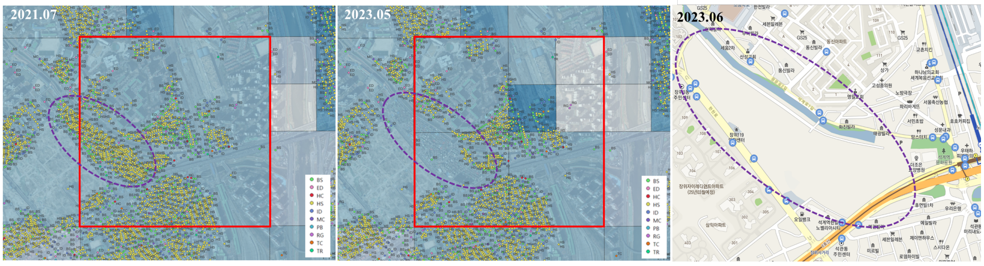
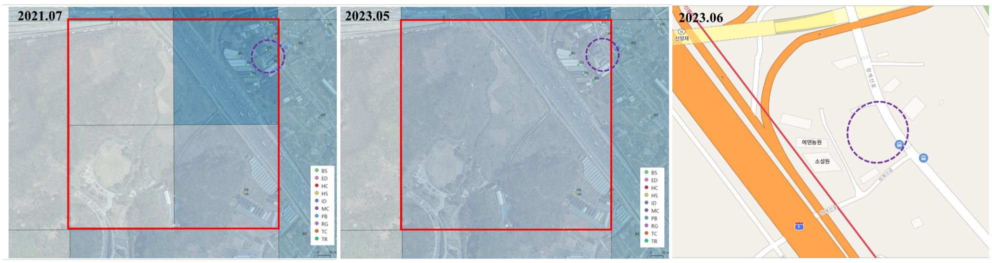


Figure 5. Vector data cube model for the time-series land-use mixing degree.

Therefore, a visual comparison was performed on the Kakao Map (https://map.kakao.com/), where a map updated in June 2023 was serviced for one of the regions corresponding to each case (Figure 6). Figure 6(a) shows an image of a cell with a higher level, which suggests that the land use mixing degree is high, as a large-scale housing complex (HS) was demolished for redevelopment (purple circle) and the POI type corresponding to the dwelling decreased. Figure 6 (b) shows am image of a cell with a lower level, and it suggests a decrease in the land use mixing degree, as an industry (ID) disappeared in 2021 (purple circle). This visual evaluation of some areas confirmed the significance of the proposed method.



(a)



(b)

Figure 6. Visible evaluation.

4. Conclusion

In this study, the time-series analysis of the land use mixing degree was performed, and the primary grid (1000 m) was divided using the IE of the POI for each cell to analyze the land use mixing degree. Furthermore, to manage the large amount of land use mixing, it was structured as a data cube discussed in EO. To this end, a vector data cube model that considers a hierarchical cell was proposed. Thereafter, a visual evaluation was performed after applying the proposed method to the POI data of Seoul, ROK, in 2021 and 2023, and the visual evaluation confirmed the significance of the proposed method. Similar to previous studies, future studies should analyze the land use mixing degree using multi-source data, and technologies for storing and managing vector data cubes should be developed.

Acknowledgement

This paper was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. 2023R1A2C2008146) and this work is supported by the Korea Agency for Infrastructure Technology Advancement(KAIA) grant R&D program of Digital Land Information Technology Development Program funded by the Ministry of Land, Infrastructure and Transport(MOLIT) (Grant RS-2022-00143804).

References

Baumann P, Rossi A P, Bell B, Clements O, Evans B, Hoenig H, Hogan P, Kakaletris G, Koltsida P, Mantovani S, Figuera R M, Merticariu V, Misev D, Pham H B, Siemen S, Wagemann J (2018) Fostering Cross-Disciplinary Earth Science Through Datacube Analytics. In: P.P. Mathieu and C. Aubrecht, eds. Earth observation open science and innovation. Switzerland: Springer Cham:91–199

Christian H E, Bull F C, Middleton N J, Knuiman M W, Divitini M L, Hopper p, Giles-Corti B (2011) How important is the land-use mix measure in understanding walking behavior? Resutls from the RESIDE study, International Journal of Behavioral Nutrition and Physical Activity 8(1): 55

Cui H, Wu L, Hu S, Lu R, Wang S (2020) Recognition of Urban Functions and Mixed Use Based on Residents' Movement and Topic Generation Model: The Case of Wuhan, China. Remote Sensing 12(18):2889. doi.org/10.3390/RS12182889

Gao F, Yue P, Cao Z, Zhao S, Shangguan B, Jiang L, Hu L, Fang Z, Liang Z (2022) A Multi-source Spatio-temporal Data Cube for Large-scale Geospatial analysis, International Journal of Geographical Information Science 36(9): 1853-1884, doi.org/10.1080/13658816.2022.2087222

Gao Q, Fu J, Yu Y, Tang X (2019) Identification of Urban Regions' Functions in Chengdu, China, based on vehicle trajectory data. PLoS ONE 14: e0215656.

Guo F, Li C, Cheng G, Chen C, Gan J (2015) Spatial-temporal Coupling Characteristics of Population Urbanization and Land Urbanization in Northeast China. Economic Geography 35:49–56

Jia X, Li J, Jia W (2016) Measure on New Urbanization Coordination Level and Compare Spatial Differences of Anhui Province. Economic Geography 36:80–86

Jing C, Zhang H, Xu S, Wang M, Zhuo F, Liu S (2022). A hierarchical Spatial Unit Partitioning Approach for Fine-grained Urban Functional Region Identification. Transactions in GIS 26: 2691–2715. doi.org/10.1111/tgis.12979

Kang Y, Wang Y, Xia Z, Chi J, Jiao M, Wei Z.W (2018) Identification and Classification of Wuhan Urban Districts based on POI. Journal of Geomatics 43:81–85

Long Y, Liu X, Featured G. (2013) How Mixed is Beijing, China? A Visual Exploration of Mixed Land Use. Environment and Planning A 45:2797–2798

Nativi S, Mazzetti P, Craglia M (2017) A View-based Model of Data-cube to Support Big Earth Data Systems Inter-operability. Big Earth Data 1(1–2):75–99: doi.org/10.1080/20964471.2017.1404232

Open Geospatial Consortium (2022) OGC Geodatacube Standard Working Group Charter, http://www.opengeospatial.org/legal/

Wagemann J, Clements O, Figuera RM, Rossi AP, Mantovani S (2018) Geospatial Web Services Pave New Ways for Server-based On-demand Access and Processing of Big Earth Data. International Journal of Digital Earth 11 (1):7–25

Wang Y, Gu Y, Dou M, Qiao M (2018). Using Spatial Semantics and Interactions to Identify Urban Functional Regions. ISPRS International Journal of Geo-Information 7(4):130. doi.org/10.3390/ijgi7040130

Wei C, Wang Z, Lan X, Zhang H, Fan M (2018) The Spatial-temporal Characteristics and Dilemmas of Sustainable Urbanization in China: A New Perspective Based on the Concept of Five-in-one. Sustainability 10:4733

Wu Q, Zhang L, Wu Z (2018) Identifying City Functional Areas Using Taxi Trajectory Data. J. Geom. Sci. Technol. 35:413–417, 424

Xia X, Lin K, Ding Y, Dong X, Sun H, Hu B (2021) Research on the Coupling Coordination Relationships between Urban Function Mixing Degree and Urbanization Development Level Based on Information Entropy. International Journal of Environmental Research and Public Health 18:242. doi.org/10.3390/ijerph18010242