SoftwareX 27 (2024) 101861

Contents lists available at ScienceDirect

SoftwareX

journal homepage: www.elsevier.com/locate/softx

Original software publication

landusemix: A Python package for calculating land use mix

Mehmet Ali Akyol [a,*], Sebnem Duzgun [b], Nazife Baykal [a]

[a] Middle East Technical University, Graduate School of Informatics, Ankara, Turkey
[b] Colorado School of Mines, Mining Engineering Department, Golden, CO, United States

ARTICLE INFO

Keywords:
Land use mix
Entropy index
Herfindahl–Hirschman index
Geographical information systems
Urban planning

ABSTRACT

Integrating different land uses within a geographic area is essential to urban planning and development. Accurate and fast land use mix (LUM) measurement is necessary for evaluating urban diversity and sustainability. In this paper, we present `landusemix`, a Python package developed to calculate LUM using two distinct indices: the Entropy Index and the Herfindahl–Hirschman Index (HHI). The `landusemix` package provides tools for GIS researchers and urban planners to measure the diversity and concentration of land use. Detailed descriptions of the methodologies employed and examples of practical usage are provided. Researchers can use this package to calculate LUM quickly and in bulk, and its results can be easily incorporated into further analysis.

Code metadata

Current code version	0.1.3
Permanent link to code/repository used for this code version	https://github.com/ElsevierSoftwareX/SOFTX-D-24-00298
Permanent link to Reproducible Capsule	https://doi.org/10.6084/m9.figshare.25864297.v1
Legal Code License	MIT License
Code versioning system used	Git
Software code languages, tools, and services used	Python
Compilation requirements, operating environments & dependencies	Python $\geq$ 3.6
If available Link to developer documentation/manual	https://github.com/makyol/landusemix/blob/main/README.md
Support email for questions	akyol.mehmet@metu.edu.tr and issues section in the GitHub repository

Software metadata

Current software version	0.1.3
Permanent link to executables of this version	https://pypi.org/project/landusemix/
Permanent link to Reproducible Capsule	https://doi.org/10.6084/m9.figshare.25864297.v1
Legal Software License	MIT License
Computing platforms/Operating Systems	OS Independent
Installation requirements & dependencies	Python $\geq$ 3.6 and packages of pandas, geopandas, and rasterio
If available, link to user manual - if formally published include a reference to the publication in the reference list	https://landusemix.readthedocs.io/
Support email for questions	akyol.mehmet@metu.edu.tr and issues section in the GitHub repository

1. Motivation and significance

Urban planning and development play critical roles in shaping the sustainability and livability of cities. One significant aspect of urban planning is integrating diverse land uses within a geographic area. A well-balanced land use mix (LUM) can enhance urban diversity, improve service accessibility, reduce transportation needs, and foster social interaction [1]. Consequently, accurately measuring LUM is essential for evaluating urban diversity and sustainability.

Urban areas are characterized by various land uses, including residential, commercial, industrial, and recreational areas. The integration and distribution of these different land uses significantly impact urban

* Corresponding author.
E-mail addresses: akyol.mehmet@metu.edu.tr (Mehmet Ali Akyol), duzgun@mines.edu (Sebnem Duzgun), baykal@metu.edu.tr (Nazife Baykal).

https://doi.org/10.1016/j.softx.2024.101861
Received 31 May 2024; Received in revised form 20 August 2024; Accepted 21 August 2024
Available online 26 August 2024

sustainability, quality of life, and overall urban functionality. Measuring LUM provides valuable insights into urban diversity and informs urban planning and policy-making [2]. Traditional methods of calculating LUM often involve manual processes or lack flexibility and scalability, presenting a need for more advanced tools.

The concept of LUM has been explored extensively in the literature, with several indices proposed to quantify it [3,4]. The Entropy Index and the Herfindahl–Hirschman Index (HHI) are the most commonly used [5]. They provide alternative methods to calculate LUM, capturing different aspects of land use concentration. The Entropy Index measures diversity by indicating how evenly land uses are distributed, while the HHI focuses on concentration, highlighting dominance or lack thereof. These indices are widely recognized and frequently used in land use studies, offering both complementary and comprehensive analysis. Previous studies have utilized these indices to address various urban planning challenges, such as evaluating land use diversity [6,7], monitoring the balance of commercial and residential areas [5,8], and assessing the impacts of urban development policies [9,10].

In recent years, integrating geospatial technologies with Python programming has significantly advanced the field, offering powerful tools for land use classification, suitability analysis, and change modeling. Previous tools and methods for calculating these indices have limitations in terms of usability and scalability. For instance, some GIS software includes functions for land use analysis, but these often require extensive manual input and are not explicitly tailored for LUM calculations. Additionally, the `landscapemetrics` is a R package offers comprehensive landscape metrics for spatial analysis [11]. Similarly, `pylandstats` provides various diversity indices for ecological studies [12]. However, the `landusemix` Python package focuses on LUM identification and gathers indices that help in this specific analysis. By integrating LUM identification indices, `landusemix` simplifies the calculation of LUM metrics, making it an essential tool for urban planners and GIS researchers seeking practical and targeted solutions.

To address this gap, we introduce `landusemix`, a Python package specifically designed to calculate LUM using the Entropy Index and the HHI. Offering both indices in the `landusemix` package allows researchers to choose the measure that best fits their specific analysis needs. This tool aims to provide GIS researchers, urban planners, and policymakers with a researcher-friendly, efficient, and flexible means for measuring the diversity and concentration of land use. It provides easy-to-understand documentation and integrates with familiar libraries like `pandas` [13,14], `geopandas` [15], and `rasterio` [16]. By leveraging these libraries, the package is designed to efficiently handle large datasets and perform bulk calculations without requiring heavy GIS applications, which are often less flexible for large-scale or repetitive LUM analysis. Additionally, it supports multiple data formats (CSV, Shapefiles, GeoJSON, raster) and offers a modular design for customization. The package streamlines the calculation of LUM indices, reducing the need for advanced GIS tools and enabling large-scale or repetitive analysis with less manual intervention, making it ideal for integration into broader research workflows. Accurately measuring LUM is critical for evaluating urban diversity, which influences various socio-economic and environmental factors [17]. The package facilitates the systematic analysis of land use patterns, providing precise and reproducible metrics that can be incorporated into broader urban studies. The package is designed to be installed and used straightforwardly, supporting input in land use area data and returning calculated indices for further analysis or visualization.

Researchers and developers can easily integrate `landusemix` into their existing workflows. An example usage scenario involves loading land use data, instantiating the LandUseMixIndices class with this data, and then calling the methods to calculate the Entropy Index and HHI. Detailed documentation and examples are provided to help users get started quickly.

The remainder of this article is structured as follows: Section 2 provides a detailed software description of the `landusemix` package, including its architecture and functionalities. Section 3 discusses the omissions and limitations. Section 4 presents illustrative examples to demonstrate the practical usage of the package. Section 5 discusses the impact of the package on urban planning and GIS research. Finally, Section 6 concludes with a summary of findings and proposes future research directions.

2. Software description

2.1. Overview

`landusemix` is a Python package designed to calculate LUM using the Entropy Index and the HHI. This package provides GIS researchers, urban planners, and policymakers tools to efficiently measure and analyze land use diversity and concentration. By leveraging Python's powerful data processing libraries, `landusemix` provides a practical and researcher-friendly solution for LUM calculations, allowing for efficient and scalable analysis without requiring extensive manual setup common in traditional GIS software.

2.2. Architecture

The `landusemix` package is organized into several key components. Fig. 1 below is an image illustrating the software architecture: Each component is designed to perform specific functions related to LUM calculations:

- **Core Module (`indices`)**: This module contains the main class and functions for calculating the Entropy Index and HHI.
- **Utilities Module (`utils`)**: Provides additional utility functions for data handling and preprocessing.
- **Data Module (`data`)**: Includes sample land use data for testing and experimentation.
- **Tests Module (`tests`)**: A suite of unit tests to ensure the correctness and reliability of the package.

2.3. Key features and functionalities

The `landusemix` package offers the following key features:

- **Entropy Index Calculation**: Computes the Entropy Index, a measure of land use diversity [18], using the formula:

  $$\mathrm{ENT} = -\frac{\left[\sum_{i=1}^{k} P_i \ln\left(P_i\right)\right]}{\ln(k)}$$

  where P_i represents the proportion of the total area occupied by the ith land use type, and k is the total number of different land use types. The Entropy Index ranges from 0 (no diversity) to 1 (maximum diversity), indicating the diversity of land use types [19].
- **HHI Calculation**: Computes the HHI, a measure of land use concentration, using the formula:

  $$\mathrm{HHI} = \sum_{i=1}^{k} \left(100 \times P_i\right)^2$$

  where P_i represents the proportion of the total area occupied by the ith land use type. The HHI ranges from 0 (many small equally-sized areas) to 10,000 (one single area), indicating the concentration of land use types [5].
- **Utility Functions**: Includes functions for data preprocessing, such as loading of different types of data formats such as CSV, Shapefiles, GeoJSON, and raster data.
- **Documentation and Examples**: Provides detailed documentation and practical usage examples to assist users in getting started quickly.

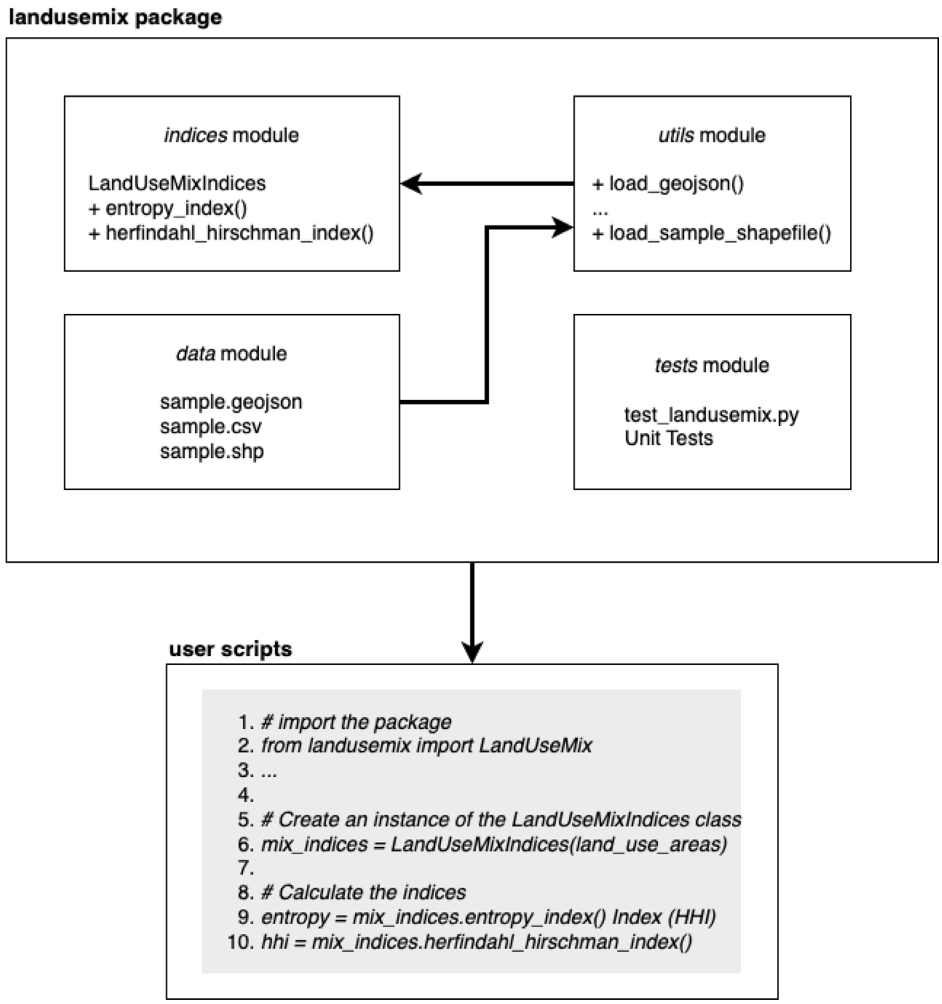

Fig. 1. Architecture of the landusemix Python package.

3. Omissions and limitations

While the `landusemix` package offers robust methods for calculating LUM using the Entropy Index and HHI, there are several limitations to consider:

- **Index Scope**: The package is limited to two indices (Entropy Index and HHI). Although these indices are widely used, other indices, such as the Dissimilarity Index [20], the Gini Coefficient, the Atkinson Index [21], or the Clustering Index [22], could provide additional insights into the LUM. These indices often require specific assumptions about spatial arrangement and social factors, which may not apply universally to all study areas or contexts. Therefore, they have been omitted from this package to maintain a broader applicability.
- **Spatial Context**: The current version of the package does not consider the spatial arrangement of land use types. The indices are based solely on the proportion of the land use areas without considering their spatial distribution. Considering the spatial distribution could provide valuable information for urban planning contexts. However, some other indices, such as the Atkinson Index, Clustering Index, and Dissimilarity Index, do consider the spatial context. Still, they have been omitted in this version to keep it generic enough for researchers.
- **Visualization Tools**: While the package provides accurate calculations, it lacks built-in visualization tools for presenting LUM results. Users need to rely on external libraries or software to visualize their findings.

Future versions of `landusemix` could incorporate several enhancements to address the current limitations by including additional indices and integrated visualizations.

4. Illustrative examples

4.1. Installation

You can use `pip` to install the `landusemix` package. The following command installs the package and its dependencies:

```
pip install landusemix
```

4.2. Example usage

Here, we provide an example of using the `landusemix` package to calculate the Entropy Index and HHI for a given set of land use areas.

4.2.1. Importing the package

First, we import the necessary modules from the package:

```
from landusemix import LandUseMixIndices
from landusemix.utils import *
```

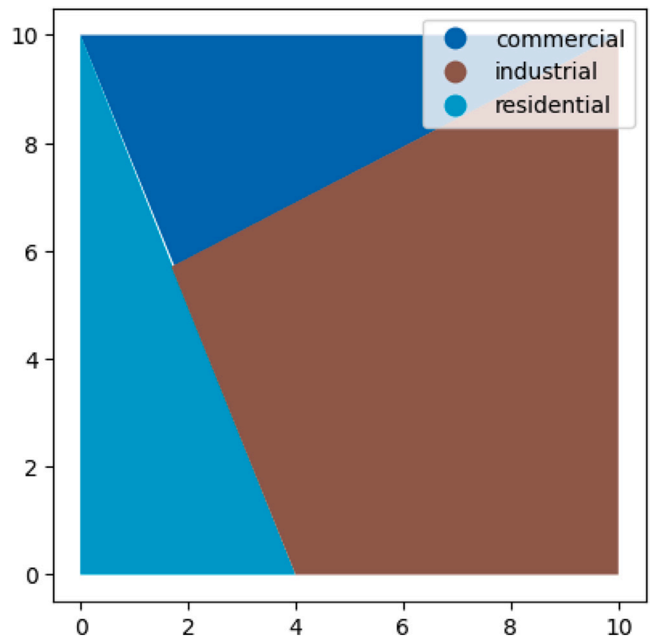

Fig. 2. Spatial representation of three different land use types (commercial, industrial, and residential) within a sample shapefile. The colors indicate different land use categories: commercial (blue), industrial (brown), and residential (cyan), showing how these areas are distributed and divided within the region.

4.2.2. Loading data

We can load the land use data from a dictionary or other compatible format. For this example, we define a dictionary representing different land use areas in square meters:

```
land_use_areas = {
  'residential': 5000,
  'commercial': 3000,
  'industrial': 2000,
}
```

You can also load data via the package's Utils functions. Fig. 2 shows the distribution of three different land use types — commercial, industrial, and residential — within a sample shapefile.

We can load the shapefile and calculate areas as such:

```
shp_gdf = load_shapefile('landusemix/data/shapefiles/multiple.shp')
shp_gdf['area'] = shp_gdf.geometry.area
land_use_areas = shp_gdf.groupby('use')['area'].sum().to_dict()
print(land_use_areas)
# {'commercial': 21.46, 'industrial': 58.67, 'residential': 20.0}
```

4.2.3. Calculating indices

Next, we create an instance of the `LandUseMixIndices` class and use its methods to calculate the Entropy Index and HHI:

```
# Create an instance of the LandUseMixIndices class
mix_indices = LandUseMixIndices(land_use_areas)

# Calculate the Entropy Index
entropy = mix_indices.entropy_index()
print(f"Entropy Index: {entropy}") # Entropy Index: 0.88

# Calculate the HHI
hhi = mix_indices.herfindahl_hirschman_index()
print(f"Herfindahl-Hirschman Index: {hhi}")
# Herfindahl-Hirschman Index: 4291.41
```

4.3. Results interpretation

The output of the above code provides the calculated Entropy Index and HHI for the given land use areas. The higher the Entropy Index, the more diverse the LUM; conversely, a higher HHI indicates a higher concentration of specific land use types.

The calculated values for the Entropy Index and HHI provide meaningful insights into land use diversity and concentration regarding LUM. For instance, an Entropy Index of 0.88 indicates a high diversity of land use types, suggesting an even distribution. Conversely, an HHI value of 500 indicates moderate concentration, with some land use types potentially dominating the area. Values closer to 0 for either index indicate lower diversity or higher concentration, respectively.

5. Impact

The `landusemix` package advances the tools available for urban planning and GIS research. This package facilitates a deeper understanding of urban diversity and its implications for sustainability and livability by providing a researcher-friendly and efficient means to calculate LUM indices.

5.1. Contributions to urban planning

Accurate and efficient measurement of LUM is essential for several reasons in urban planning:

- **Informing Policy Decisions**: The insights gained from analyzing the LUM can guide policy decisions to promote balanced urban development and enhance accessibility to essential services. These indices can be applied within specific boundaries or zones, using additional data, such as shapefiles or GeoJSON files, to generate localized insights. This approach has proven effective in real-world applications, as demonstrated in our recent work on identifying land use mix within fishnet grids using point-based geospatial data [23]. The HHI can identify areas with a significant concentration of specific land uses, which may indicate potential issues with service accessibility. For instance, high HHI values in residential areas might highlight a lack of nearby commercial or recreational facilities. Researchers can use these insights to make data-driven recommendations for zoning changes or new service allocations to improve accessibility.
- **Evaluating Urban Diversity**: By quantifying the diversity and concentration of land use, planners can identify areas requiring interventions to improve functionality and livability. The Entropy Index can help assess urban diversity by indicating how evenly different land uses are distributed within a geographic area. High entropy values suggest diverse land use, which is often associated with sustainable urban environments. Planners can use these insights to promote mixed-use development, enhancing overall urban sustainability.
- **Supporting Sustainable Development**: Understanding land use patterns and their impact on urban sprawl, transportation needs, and resource allocation helps promote sustainable urban growth. Both indices can provide insights into land use patterns that foster or hinder social interactions. Urban planners can design strategies to encourage community well-being by analyzing areas

with diverse or concentrated land uses. For instance, lower HHI and higher entropy values could support the creation of public spaces that facilitate social interactions and community cohesion.

The `landusemix` package empowers urban planners with reliable metrics that can be easily integrated into various planning processes, from initial assessments to long-term strategic planning.

5.2. Contributions to GIS research

The role of GIS in urban studies is critical, as it allows for the spatial analysis and visualization of land use patterns. The `landusemix` package enhances GIS research in the following ways:

- **Enhanced Data Analysis**: The package provides practical tools for the quantitative analysis of land use data, enabling large-scale and flexible analysis that can be easily integrated into further data mining and machine learning methods, offering advantages over traditional GIS tools.
- **Reproducible Research**: The package promotes reproducible research by offering standardized methods for calculating LUM, ensuring consistency and accuracy in urban studies.

GIS researchers can leverage the capabilities of `landusemix` to perform advanced analysis of land use diversity, contributing to a broader understanding of urban dynamics.

6. Conclusions

In this paper, we introduced `landusemix`, a Python package developed to calculate LUM using the Entropy Index and the HHI. Accurate measurement of LUM is essential for evaluating urban diversity and sustainability, and the `landusemix` package provides a reliable, researcher-friendly, and efficient tool for this purpose.

We discussed the motivation and significance of the package, highlighting its importance in urban planning and GIS research. By simplifying the calculation of key LUM indices, `landusemix` reduces the potential for human error and lessens the dependence on heavy GIS applications such as ArcGIS and QGIS. The package is designed to streamline and scale analysis, making it possible to efficiently handle larger datasets and perform repetitive calculations more easily, thanks to its integration with efficient libraries such as `pandas`, `geopandas`, and `rasterio` and allowing the integration the results for further analysis within the Python ecosystem.

We also discussed the broader impact of the package on urban planning and GIS research, emphasizing its contributions to policy decision-making, sustainable development, and comprehensive urban studies.

While `landusemix` offers robust features, there are still areas for improvement, such as incorporating additional indices, accounting for spatial context, and integrating visualization tools. These enhancements could provide a more comprehensive assessment of the LUM and cater to more advanced user needs.

In conclusion, the `landusemix` package is a valuable resource for researchers and practitioners in urban planning and GIS, providing essential tools for analyzing and understanding land use diversity and its implications.

CRediT authorship contribution statement

Mehmet Ali Akyol: Writing – review & editing, Writing – original draft, Validation, Software, Resources, Methodology, Conceptualization. **Sebnem Duzgun:** Writing – review & editing, Supervision, Investigation. **Nazife Baykal:** Writing – review & editing, Supervision, Investigation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

[1] Iannillo A, Fasolino I. Land-use mix and urban sustainability: Benefits and indicators analysis. Sustainability 2021;13(23):13460.
[2] Jiao J, Rollo J, Fu B. The hidden characteristics of land-use mix indices: An overview and validity analysis based on the land use in Melbourne, Australia. Sustainability 2021;13(4):1898.
[3] Manaugh K, Kreider T. What is mixed use? Presenting an interaction method for measuring land use mix. J Transp Land use 2013;6(1):63–72.
[4] Zhuo Y, Jing X, Wang X, Li G, Xu Z, Chen Y, Wang X. The Rise and Fall of Land Use Mix: Review and Prospects. Land 2022;11(12):2198.
[5] Song Y, Merlin L, Rodriguez D. Comparing measures of urban land use mix. Comput Environ Urban Syst 2013;42:1–13.
[6] Im HN, Choi CG. The hidden side of the entropy-based land-use mix index: Clarifying the relationship between pedestrian volume and land-use mix. Urban Stud 2019;56:1865–81. http://dx.doi.org/10.1177/0042098018763319.
[7] Jiao J, Rollo J, Fu B. The Hidden Characteristics of Land-Use Mix Indices: An Overview and Validity Analysis Based on the Land Use in Melbourne, Australia. Sustainability 2021. http://dx.doi.org/10.3390/SU13041898.
[8] Im HN, Choi CG. Measuring pedestrian volume by land use mix: Presenting a new entropy-based index by weighting walking generation units. Environ Plan B Urban Anal City Sci 2020;47:1219–36. http://dx.doi.org/10.1177/2399808318824112.
[9] Zhou D, Xu J, Lin Z. Conflict or coordination? Assessing land use multi-functionalization using production-living-ecology analysis.. Sci Total Environ 2017;577:136–47. http://dx.doi.org/10.1016/j.scitotenv.2016.10.143.
[10] Jehling M, Hecht R. Do land policies make a difference? A data-driven approach to trace effects on urban form in France and Germany. Environ Plan B Urban Anal City Sci 2021;49:114–30. http://dx.doi.org/10.1177/2399808321995818.
[11] Hesselbarth MHK, Sciaini M, With KA, Wiegand K, Nowosad J. landscapemetrics: an open-source R tool to calculate landscape metrics. Ecography 2019. URL https://api.semanticscholar.org/CorpusID:198257231.
[12] Bosch M. PyLandStats: An open-source Pythonic library to compute landscape metrics. PLoS ONE 2019;14. URL https://api.semanticscholar.org/CorpusID:199632776.
[13] The pandas development team. Pandas-dev/pandas: Pandas. 2020, http://dx.doi.org/10.5281/zenodo.3509134.
[14] McKinney W. Data structures for statistical computing in python. In: van der Walt S, Millman J, editors. Proceedings of the 9th python in science conference. 2010, p. 56–61. http://dx.doi.org/10.25080/Majora-92bf1922-00a.
[15] den Bossche JV, et al. geopandas/geopandas: v1.0.1. Zenodo; 2024, http://dx.doi.org/10.5281/zenodo.12625316.
[16] Gillies S, et al. Rasterio: geospatial raster I/O for Python programmers. Mapbox; 2013–, URL https://github.com/rasterio/rasterio.
[17] Christian HE, Bull FC, Middleton NJ, Knuiman MW, Divitini ML, Hooper P, Amarasinghe A, Giles-Corti B. How important is the land use mix measure in understanding walking behaviour? Results from the RESIDE study. Int J Behav Nutr Phys Activity 2011;8:1–12.
[18] Turner MG, Gardner RH, O'neill RV, O'Neill RV. Landscape ecology in theory and practice. vol. 401, Springer; 2001.
[19] Song Y, Knaap G-J. Measuring the effects of mixed land uses on housing values. Reg Sci Urban Econ 2004;34(6):663–80.
[20] Duncan OD, Duncan B. A methodological analysis of segregation indexes. Am Sociol Rev 1955;20(2):210–7.
[21] Atkinson AB, et al. On the measurement of inequality. J Econom Theory 1970;2(3):244–63.
[22] Massey DS, Denton NA. The dimensions of residential segregation. Soc Forces 1988;67(2):281–315.
[23] Akyol MA, Temizel TgT, Duzgun S, Baykal N. Identification of Land Use Mix Using Point-Based Geospatial Data in Urban Areas. Appl Sci 2024;14(16). http://dx.doi.org/10.3390/app14166871, URL https://www.mdpi.com/2076-3417/14/16/6871.