Flexurba: an open-source R package to flexibly reconstruct the Degree of Urbanisation classification.

Van Migerode Céline[*1,2], Poorthuis Ate[†2] and Derudder Ben[‡1,3]

[1]Public Governance Institute, KU Leuven, Belgium
[2]Division of Geography and Tourism, KU Leuven, Belgium
[3]Department of Geography, Ghent University, Belgium

GISRUK 2024

Summary

We propose *flexurba*, a library written in R with the first open reconstruction of the Degree of Urbanisation algorithm to classify cities, towns, and rural areas. The package offers enhanced flexibility and facilitates constructing alternative versions of the classification by customising the minimum population size required for a city, and more 'hidden' implementation details including contiguity requirements and smoothing rules. The package enables a broad range of analyses beyond the Degree of Urbanisation's original application, such as evaluating alternative urban delineations, sensitivity analyses and comparative research. *Flexurba*'s source code be explored here, and the documentation is included on this website.

KEYWORDS: Degree of Urbanisation, R package, urban delineations

1. Introduction

The Degree of Urbanisation (DEGURBA) is an operational definition of cities, towns, and rural areas that is widely used in policy reports and academic circles alike. It is developed to provide a globally consistent delineation of urban areas and enhance statistical comparability across nations (Dijkstra et al., 2021). Much effort has been devoted to making the workflow underlying DEGURBA's methodology as transparent and reproducible as possible: it is extensively documented in a manual (Eurostat, 2021) and a set of analytical tools is provided by the Global Human Settlement Layer (GHSL) to reproduce the classification. The current suite of tools employs a graphical user interface. This certainly enhances the ease of use, but also obscures the underlying code and restricts the user from adapting implementation details in the algorithm. Consequently, it complicates extending or amending the methodology to facilitate other, related uses, such as sensitivity analyses and/or comparative research. Open-source DEGURBA tools in popular programming languages such as R or Python would address these constraints and help further expand research with and on the Degree of Urbanisation.

Against this backdrop, we reconstructed the classification algorithm in R and openly published the code as the *flexurba* package: a **flex**ible reconstruction of the DEG**URBA** algorithm. The package seamlessly complements the already-rich set of existing tools and further improves the overall usability of the methodology. The contribution of the package is fourfold. First, the package encompasses the first open reconstruction of DEGURBA's algorithm in any programming language. In this way, we contribute to open science and respond to Boeing's (2020) recent call for open-source spatial research software. Second, as the package is written in R – a comprehensive and flexible environment increasingly employed in geospatial research – it facilitates incorporating DEGURBA in other common spatial analytical workflows. Third, the package provides flexibility to customise the parameters in the

[*] celine.vanmigerode@kuleuven.be
[†] ate.poorthuis@kuleuven.be
[‡] ben.derudder@kuleuven.be

algorithm and evaluate the consequences of certain implementation choices. The user can construct alternative versions of DEGURBA by changing the minimum population thresholds and more 'hidden' implementation details, such as the contiguity requirements and smoothing rules. Fourth, two different versions of the Degree of Urbanisation are included in the package: the version of GHSL Data Package 2022 (European Commission, 2022) and the version of GHSL Data Package 2023 (European Commission, 2023). Consequently, the package allows a comparison of differences between the versions and assessing the effect of changing implementation details.

2. The Degree of Urbanisation methodology

In the grid cell classification of DEGURBA, the cells of a 1 km² population grid are classified into three different categories based on the following rules, detailed in the GHSL Data Package 2023 (European Commission, 2023):

- **Urban centres** are clusters of cells (rook contiguity) with a minimum population density of 1,500 inhabitants per km² or a minimum built-up density threshold. In addition, the total population in these clusters should be at least 50,000. Gaps are filled, and edges are smoothed with a majority rule.
- **Urban clusters** are clusters of cells (queen contiguity) with a minimum population density of 300 inhabitants per km² and a minimum total population of 5,000 inhabitants. Cells that belong to urban centres are removed from urban clusters.
- **Rural grid cells** do not belong to an urban centre or cluster.

For more information, readers can consult Dijkstra et al. (2021), the methodological manual (Eurostat, 2021), or the GHSL Data Packages (European Commission, 2022; 2023).

3. The *flexurba* package

The grid cell classification is implemented through the function `classify_grid` in the *flexurba* package and offers customisable classification rules via its arguments. **Code example 1** illustrates the code for four different parameterisations applied to the case of Belgium. First, Belgium's pre-processed population grid, built-up area grid, and land grid are loaded in the object `data_belgium`. Afterwards, the grid cell classification is constructed (A) with standard parameters, (B) with decreased population thresholds, (C) with queen contiguity for urban centres, and (D) without the built-up area criterium. The resulting classifications are visualised in **Figure 1**. Decreasing the population thresholds or changing the contiguity rule for urban centres results in more and larger urban centres (e.g., Aalst – a mid-sized city – becomes an urban centre), whereas removing the built-up area rule decreases the coverage of urban centres (e.g., the port of Antwerp is no longer part of an urban centre). Minor variations in the parameters thus lead to differences in the classification and may have substantial implications for policymaking and research that draw on DEGURBA.

Apart from the grid cell classification, the *flexurba* package provides various other functionalities, including spatial units classification, pre-processing the necessary input data and plotting the results. In that way, researchers can conduct an entire workflow with DEGURBA, from acquiring the data to visualising the results. For an overview of all functions, we refer to the package documentation website.

```
library(flexurba)

data_belgium <- load_grid_data_belgium()

# (A) standard parameters
classification_1 <- classify_grid(data_belgium)

# (B) decreased population density and size thresholds
# for urban centres (UC) and urban clusters (UCL)
classification_2 <- classify_grid(data_belgium,
                                  parameters = list(UC_density_threshold  = 1000,
                                                    UC_size_threshold     = 30000,
                                                    UCL_density_threshold = 200,
                                                    UCL_size_threshold    = 3000))

# (C) queen contiguity for urban centres
classification_3 <- classify_grid(data_belgium,
                                  parameters = list(UC_contiguity_rule = 8))

# (D) without the built-up criterium for urban centres
classification_4 <- classify_grid(data_belgium,
                                  parameters = list(UC_built_criterium = FALSE))
```

Code Example 1: Grid classification for Belgium with 4 parameterisations

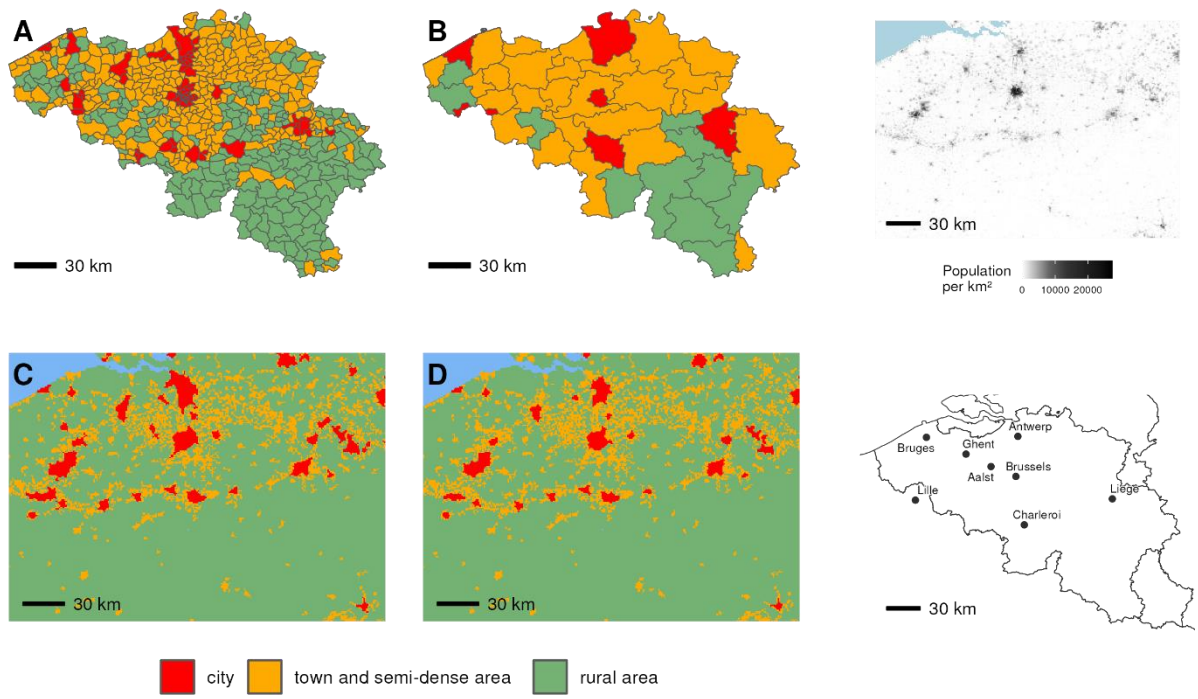

Figure 1: Result of the grid classifications constructed in Code Example 1

4. Potential applications

DEGURBA aims to provide a globally consistent delineation of urban areas (Dijkstra et al., 2021). However, it might be relevant for academic research to adapt the standard parameters in the algorithm, as different delineations may be more appropriate depending on the research question. For example, Henderson et al. (2021) investigate economic density in Sub-Saharan Africa and argue that the minimum density of 300 inhabitants per km² for urban clusters is too low. They propose a slightly increased threshold of 500 inhabitants and employ a population smoothing filter instead of using the standard majority rule to smooth edges of urban centres. Urban delineations with these alternative

classification rules can easily be constructed with the *flexurba* package.

Furthermore, the package allows for sensitivity analysis of minor implementation changes in DEGURBA's classification rules. The algorithm behind DEGURBA is often altered in the context of new GHSL Data Package releases. As an example, in Data Package 2022, DEGURBA included a fixed minimum built-up density threshold of 50% to account for the presence of office parks, shopping malls, and transport infrastructure. In Data Package 2023, the built-up density rule is slightly adapted: a minimum 'optimal' threshold is employed, instead of the minimum fixed threshold (European Commission, 2022; 2023). Based on the official definition of the 'optimal' built-up threshold in Data Package 2023, using the *flexurba* package, we calculated that the global 'optimal' threshold for the GHSL data of 2020 is 20% built-up density. Changing the built-up rule from a fixed to an 'optimal' threshold thus implied a decrease from 50% built-up density to 20% built-up density, which caused a substantial increase in the coverage of urban centres. The *flexurba* package facilitates uncovering the implications of such implementation changes and thus allows performing (potentially spatially-explicit) sensitivity analyses.

5. Conclusion

We present *flexurba*, an open software product built on well-established spatial frameworks in R. *Flexurba*'s source code can be freely explored and downloaded from this GitLab repository, and the documentation of the functions and several vignettes are included on this website. We hope to continue expanding and improving the package based on future requests, potential bug fixes, and contributions from the authors and the community. By sharing the first open reconstruction of the Degree of Urbanisation algorithm, we invite researchers to explore the possibilities of related urban delineations and repurpose the methodology from its original application to a broad range of related urban analyses.

6. Acknowledgements

The research for this paper was supported by patronage funding provided to KU Leuven for carrying out scientific research into urban processes, and partly by Internal Funds KU Leuven grant number STG/20/021. This work is part of the first work package in the PhD project 11P4224N, funded by Research Foundation Flanders (FWO). The paper utilises the open data products provided by the Global Human Settlement Layer.

References

Boeing, G (2020). The right tools for the job: The case for spatial science tool-building. *Transactions in GIS*, 24(5), 1299–1314.

Dijkstra, L, Florczyk, A J, Freire, S, Kemper, T, Melchiorri, M, Pesaresi, M, and Schiavina, M (2021). Applying the Degree of Urbanisation to the globe: A new harmonised definition reveals a different picture of global urbanisation. *Journal of Urban Economics*, 125, 103312.

European Commission (2023). *GHSL Data Package 2023* (JRC133256; Publications Office of the European Union).

European Commission (2022). *GHSL Data Package 2022* (JRC129516; Publications Office of the European Union).

Eurostat (2021). *Applying the Degree of Urbanisation—A methodological manual to define cities, towns and rural areas for international comparisons—2021 edition*.

Henderson, J V, Nigmatulina, D, and Kriticos, S (2021). Measuring urban economic density. *Journal of Urban Economics*, 125, 103188.

Biographies

Céline Van Migerode is PhD student in Urban and Regional Geography of the Research Foundation Flanders (FWO) at the KU Leuven Public Governance Institute and Division of Geography and Tourism. She is investigating disagreement and variation in the definition and measurement of urban concepts.

Ate Poorthuis is an Assistant Professor in Big Data and Human-Environment Systems at the KU Leuven Division of Geography and Tourism. His research explores the possibilities and limitations of big data, through quantitative analysis and visualization, to better understand how cities work.

Ben Derudder is a Professor in Urban Studies at the KU Leuven Public Governance Institute. His research adopts an explicitly spatial approach to studying urban and regional questions: city networks at the global scale, urban and regional polycentricity, the transformation of Chinese urban regions, and region-formation in Flanders.