Spatial data disaggregation application in Just Transition monitoring

Yiannis Karagiannis [1], Despoina Kanteler [2], Sofia Pavlidou [3], Konstantinia Nikolaidou [4] and Lefteris Topaloglou [5,*]

1 Department of Chemical Engineering, University of Western Macedonia, ZEP P.C. 50100, Kozani, Greece; i.karagiannis@uowm.gr
2 Department of Chemical Engineering, University of Western Macedonia, ZEP P.C. 50100, Kozani, Greece; drdcbs00001@uowm.gr
3 Department of Mechanical Engineering, University of Western Macedonia, ZEP P.C. 50100, Kozani, Greece; pavlidoutm@gmail.com
4 Department of Chemical Engineering, University of Western Macedonia, ZEP P.C. 50100, Kozani, Greece; k.nikolaidoy@uowm.gr
5 Department of Chemical Engineering, University of Western Macedonia, ZEP P.C. 50100, Kozani, Greece; etopaloglou@uowm.gr
* Correspondence: i.karagiannis@uowm.gr

Abstract

Background: In this paper we are going to explore the data disaggregation methodologies that can be used in monitoring Just Transition processes. Data disaggregation methodologies are essential for the success of the "leave no one behind" (LNOB) principle of the 2030 Agenda for Sustainable Development by identifying and addressing inequalities. The LNOB principle is put in place to ensure that development efforts benefit all members of society. A basic requirement for the success of this effort is to monitor the progress and identify which populations groups are left behind using detailed disaggregated data. These vulnerable population groups can be identified by using various dimensions, namely geographic location, income, sex, age, race, ethnicity, migratory status, disability. Analysing disaggregated data is important to understand the specific needs and conditions of these groups and to design effective policies as well as ensure the inclusiveness of development efforts (ADB, 2021).

Theoretical Framework: Data disaggregation is the process of breaking down broad datasets into finer, more specific informational units, of a categorical or spatial nature. It enables a more detailed analysis by shedding light on hidden trends that appear when examining subsets of the original data. By applying these methodologies policymakers, researchers, and organisations can understand how progress toward the SDGs achievement is developing at various hierarchical levels and subsets, (ADB, 2021). One case of specific importance is the spatial data disaggregation of spatial data to lower-level administrative units. Population disaggregation methodologies used in recent decades can be categorised as areal interpolation and dasymetric mapping, (Qiu, 2022).

Literature Review: Area-based areal interpolation methods are mainly areal weighting and pycnophylactic interpolation, (Comber & Zeng, 2019). Of course, there exists other area-based areal interpolation methods which are improvements of these methods. Areal interpolation, by using the spatial relationship of two sets of areas can transfer data from one set of objects, source zones, to another, target zones. Areal interpolation can be subcategorised to areal interpolation

without ancillary information and areal interpolation with ancillary information based on whether ancillary data is used. Another categorisation can be derived by the consistency of the statistical variable values during the transformation process from source to target zones: volume-preserving (e.g. area-based areal interpolation) and non-volume-preserving (e.g. point-based areal interpolation). The most fundamental method in area-based areal interpolation is areal weighting, also known as proportional reallocation. An improvement on simple area weighting is mask area weighting is because it uses a mask to define where, within the target zone, the source data should be allocated. Land cover can be used to identify populated areas and create the mask. (EEA, 2012). Pycnophylactic interpolation, which is an extension to area weighting method, produces a smooth and continuous population surface, by assuming that the population of an area tends to be similar to the population of nearby areas and iteratively applying a smooth function to the target zones' weighted average of its nearest neighbours, adjusted so that in each iteration the total population is constant.

Dasymetric mapping, was named and developed by Semenov-Tian-Shansky (1928) and subsequently popularized by Wright (1936). This method consists of subdividing the source zone into smaller areas that can reflect the spatial changes in the population based on ancillary data, (Huyen Do et al., 2015). We assume that the population distribution patterns in the small areas are similar or the same. In the final step, we apply areal interpolation to generate population distribution data at different scales. This method can be considered as an improvement of areal interpolation. Dasymetric mapping can be categorised into binary dasymetric, multi-class dasymetric, and intelligent dasymetric mapping. Binary dasymetric mapping uses ancillary data to divide a source zone into two sub-zones, usually populated and unpopulated areas. The unpopulated area is masked out, and all the population is distributed to the populated area. The binary method is widely discussed due to its simplicity. It is the subjective decision of the researcher to decide which area is populated and which one is not. In addition, according to the level of regional economic development, the source zone can also be divided into urban and rural areas. In multi-class dasymetric method we divide a source zone into several sub-zones, which can effectively improve the accuracy of population disaggregation. The sub-zones are divided according to local knowledge and then we estimate the population count (density) of each sub-zone. For example, we can use ach land cover class as a sub-zone and then regression analysis to determine the population density of each. A more complex modelling method is intelligent dasymetric that supports a variety of methods, such as empirical sampling, predefined population density statistics, regression and machine learning. It is worth mentioning that deep learning methods can extract highly abstract features, and their non-linear expression abilities are better than traditional machine learning methods in many fields, (Qiu, 2022).

Ancillary data used for disaggregation modelling can be land cover, nighttime lights, Infrastructures and Environmental factors. Modern data sources can be open data initiatives, online service providers and data collected through volunteer geographic information activities, (Huyen Do et al. 201).

Conclusions: This paper reviews and summarises the main approaches used in spatial data disaggregation. The amount of data that can support spatial data disaggregation is increasing. Precision and accuracy of this methodologies and the quality of ancillary data used should always be examined and the results should be evaluated and cross-checked with questionnaires and field surveys.

Extended Abstract References

ADB - Asian Development Bank. (2021). Practical guidebook on data disaggregation for the Sustainable Development Goals. https://dx.doi.org/10.22617/TIM210117-2.

Huyen Do, V., Thomas-Agnan, C., & Vanhems, A. (2015). Spatial reallocation of areal data another look at basic methods. Revue d'Économie Régionale & Urbaine, (1/2), 27-58.

Qiu, Yue, Xuesheng Zhao, Deqin Fan, Songnian Li, and Yijing Zhao. 2022. 'Disaggregating Population Data for Assessing Progress of SDGs: Methods and Applications'. International Journal of Digital Earth 15(1):2–29. doi: 10.1080/17538947.2021.2013553.

EEA - European Environment Agency. 2012. Technical Guidance on Provision of Spatial Data for MSFD Reporting. European Topic Centre - Onland, coastal, marine waters.

Comber A, Zeng W. Spatial interpolation using areal features: A review of methods and opportunities using new forms of data with coded illustrations. Geography Compass. 2019;13:e12465. https://doi.org/10.1111/gec3.12465