geographical analysis

Geographical Analysis (2022) **0**, n/a

Testing Transferability: Quantitative Evaluation of Labor Market Area Definition Methods in Three Contrasting Countries

José Manuel Casado-Díaz[1,2], Mike Coombes[3], and Lucas Martínez-Bernabéu[1]

[1]International Economics Institute, University of Alicante, Alicante, Spain, [2]Department of Applied Economic Analysis, University of Alicante, Alicante, Spain, [3]Centre for Urban and Regional Development Studies, Newcastle University, Newcastle upon Tyne, UK

Sub-national economic policies increasingly use labor market areas (LMAs) rather than administrative areas for analysis and implementation. How a set of LMAs was defined influences the results of such analyses, and so accurate policy delivery needs appropriately defined LMAs. Multinational bodies need comparable LMA definitions in many countries, calling for a definition method that is transferable across national boundaries. This article applies quantitative metrics to evaluate LMAs defined in three contrasting countries by three methods that represent the main methodological approaches. The deductive approach – based on a center and hinterland – is too inflexible to deal with differing geographical circumstances and cannot cope with statistical zones that are very small, or do not respect settlement structure. The alternative inductive methods tested define appropriate LMAs in each country, with the newer method performing slightly better in statistical terms. The article also exemplifies the usefulness of the metrics for comparisons of alternative regionalizations.

Introduction

This article evaluates exemplar methods of widely used approaches to defining labor market areas (LMAs) using well-founded quantitative analysis. The motivation comes from the continuing policy need for well-defined LMAs to provide the most accurate analysis of local economy conditions within a country (OECD, 2020). The article applies a set of quantitative metrics, derived from the LMA concept itself, to empirically evaluate LMAs defined in three countries by three selected methods.

The diffusion of geographic information systems, linked to increasingly available small area data, has enabled a growth of local-scale policy analyses (Dijkstra, Poelman, and Veneri, 2019).

Correspondence: Lucas Martínez-Bernabéu, International Economics Institute, University of Alicante, Alicante, Spain.
e-mail: lucas.martinez@ua.es

Submitted: May 9, 2022. Revised version accepted: October 18, 2022.

doi: 10.1111/gean.12353

Yet outside academia there is only a limited awareness that the results from analyzing zonal datasets partly depend upon the boundaries of the areas analyzed (e.g., Taubenböck et al., 2021). This is the modifiable areal unit problem (Openshaw and Taylor, 1981; Openshaw, 1984). One major review of this issue in the context of European spatial policy concluded that the fact that the results of an analysis depend upon the choice of zones analyzed can be "very disturbing for the decision maker" (Grasland and Madelin, 2006, p. 2). The problem can be mitigated by using areas whose boundaries were defined to be policy-appropriate. The appropriate areas for much economic and social policy are LMAs, with the result that many countries have been defining LMAs for some time (Cattan, 2002; OECD, 2020).

The value of LMAs as analytical units for policy is clearly illustrated by the LMA forming the basis of the metropolitan area definitions in the United States which a decade ago had already been "standard statistical area delineations for approximately 60 years" (Sunstein, 2010, p. 37246). The value of LMAs derives from them being functional regions because they are "composed of areas or locational entities which have more interaction or connection with each other than with outside areas" (Brown and Holmes, 1971, p. 57). These characteristics of cohesion and self-containment are the key advantages of LMAs over the "default" areas to use for policy, local or regional administrative areas, that are more susceptible to spatial spillovers and inconsistency in econometric analysis. Many administrative boundaries have been unchanged for decades, due to the need for stability in governance structures, so they cannot reflect the evolution of a country's economic and social geography (Forstall, Greene, and Pick, 2009).

Fowler and Jensen (2020) recently highlighted that there are several official sets of LMAs in the United States. This illustrates the diversity of LMA definition methods previously highlighted by several methodological reviews (e.g., Casado-Díaz and Coombes, 2011; Klapka and Halás, 2016). Scientific research into LMA definition methods continues but has not led to the consensus about best practice that has long been called for (e.g., Schubert et al., 1987). This scientific challenge – identifying demonstrably better ways to define LMAs – is the empirical motivation of this article.

At the same time, the benefit to policy analysis of using LMAs will depend on the "quality" of the definitions of those LMAs. This has been emphasized recently by the OECD when identifying "the need for meaningful geographies for analysis and policy … such as metropolitan areas, LMAs, daily urban systems or, more generally, functional areas … offering precise information on policy-relevant areas" OECD (2020, p. 12). Analyses such as those to identify areas most in need of policy support will depend on the analyzed LMAs consistently reflecting local economic geography. Consistency of definition, as well as appropriateness of boundaries, is essential for the "fairness" of such a ranking of places' need for policy support. OECD (2020) also recognizes that an international "best practice" LMA definition method applied in different countries would make possible meaningful cross-national analyses at the local scale. In practice, this means that such a method needs transferability: it must consistently define appropriate LMAs in countries with very different topographies, settlement structures, and available data. This article aims to identify an approach to defining LMAs, which can meet this challenge of transferability.

The article is organized as follows. Section 2 first identifies three principal categories of LMA definition method in research literature and/or official statistical practice, then selects appropriate exemplars of each method type and summarizes the key differences between them. Section 2 then highlights how the three countries selected for the empirical analyses, Spain, Sweden, and the United Kingdom, differ in ways that will test the transferability of the methods. The basic results of applying the three selected methods in the three contrasting countries are then outlined. Section 3 specifies the evaluation metrics, each of which represents one element of the LMA concept. It then reports all the metric values for each of the 3×3 sets of LMAs, highlighting substantial differences between the values for the LMAs defined by different methods. Section 4 discusses the results of the empirical analyses more thematically, drawing out the main implications for the selection of LMA definition methods. Section 5 provides the principal conclusions of these findings, indicates their policy implications (while acknowledging limitations of the empirical analyses) and outlines some opportunities for future research that could build on the material presented here.

Approaches to the definition of LMAs

The different approaches to define the boundaries of LMAs can be divided into those that are essentially inductive, and those termed deductive due to requiring that all LMAs conform to one predetermined spatial structure (van der Laan and Schalke, 2001). The latter category includes most of the earliest methods, in which the customary predetermined structure was of an urban center with its hinterland. The inductive approach, by contrast, is purely data-driven, and defines LMAs regardless of their morphology.

There are two main reasons why most newer regionalization methods – which are nearly all dependent on analyzing commuting flow data (cf., Obaco, Royuela, and Xavier, 2020) – are inductive methods. The first reason is that new method developments are responding to the increasingly diverse patterns in modern commuting. Polycentric urban systems have emerged alongside the "traditional" urban system of well-spaced cities with distinct hinterlands, with other trends including a growth of "hybrid" work modes due to partial homeworking. This increase in the variety of LMA morphologies is no problem for inductive methods, but challenges the inflexibility of the deductive approach's predetermined center and hinterland structure that was based upon mid-twentieth century patterns. The second reason is that massively reduced computational constraints mean there is now little need for deductive methods' reduction in data processing time due to them only analyzing flows that fit the centers and hinterland model (Lankford, 1969).

A third possible reason for the development of deductive methods previously was that the structure of an urban center and rural hinterland was effectively suggested by the zones often used for reporting commuting data. In much of continental Europe in particular, official datasets use zone boundaries based on local administrative areas, and these historically were *either* towns and cities *or* the "rural residual" (Champion and Hugo, 2004, p. 9). A set of data zones that were either urban or rural – but not a mix of the two – was readily analyzed within a framework presuming urban centers and rural hinterlands. By way of contrast, U.K. commuting data has since 1980 been reported for zones whose size constraint meant that they are often no more than fragments of urban areas, and this stimulated the development of inductive LMA definition methods, beginning with Smart (1974).

The continuing use of deductive methods makes it important for the empirical analyses in this article to include such a method. The metropolitan area definitions in the United States might have been the obvious option given their longevity and the fact that other countries have mirrored them, followed by attempts to generalize them on an international basis (e.g., Hall and Hay, 1980; Dijkstra, Poelman, and Veneri, 2019). However these are "metropolitan" LMAs and so are not complete mappings of the LMAs in any country that includes rural regions with no large city.

One set of well-used urban-centered LMAs covering a whole country are the official Swedish LMAs "lokala arbetsmarknader" (whose definition method will be referred to here as LAm). The LAm method is of particular interest because it has also been applied – and the LMAs it defines used for policy – in other countries (Denmark, Finland, and Norway). (A description of the LAm method is provided in Coombes et al., 2012, Annex 1, pp. 129–130). Its first step uses criteria based on commuting flow patterns, not urban size, to identify employment centers, thus allowing LMAs to be defined even in remote regions (SCB, 1992). All remaining zones are then allocated to centers with a hierarchical procedure based on commuting outflows. Due to being strictly hierarchical, the allocation process can be visualized in a single dendrogram. At the same time, a hierarchical process has the disadvantage of "locking-in" allocations that may be sub-optimal. (An example would be that area A may group with X because of marginally stronger links with X than either Y or neighbor Z, but if subsequently Y groups with Z then the combined links of A with the combined YZ may well be stronger than its links with X.) The successful use of the LAm method in several countries is the reason for its selection here as exemplar deductive approach to defining LMAs in the evaluation analyses below.

Turning to inductive methods, there is a long history of use for official statistics and policy of the boundaries delimited by the Coombes and Bond (2008) method that defines Travel-to-Work Areas (TTWAs), the official U.K. LMAs. Refined over a period of five decades, versions of the TTWA method are used in various countries (Casado-Díaz and Coombes, 2011). As well as being inductive rather than deductive, it also differs from the LAm by being less hierarchical: early area groupings may be changed later in the process so that groupings closer to the optimal (in terms of self-containment and cohesion) are possible. Only groupings that still do not meet the minimum requirements are reviewed, so this is a limited improvement. Thus although the method performs locally optimal choices in re-grouping the zones under review, the complexity of the problem makes it likely that some regionalizations with slightly better allocation of zones (in terms of self-containment and cohesion) will not be considered.

Some new regionalization methods have been developed specifically as global optimizers, rooted in taxonomic principles unrelated to the geography of LMAs. Such algorithms define LMAs by making a broader search of the solutions space and optimizing a statistic based on commuting flow data. One recent example is the GEA method (Martínez-Bernabeu, Flórez-Revuelta, and Casado-Díaz, 2012; Martínez-Bernabeu and Casado-Díaz, 2022) that was devised in Spain for defining LMAs in that country but has also proved its utility in defining Chilean LMAs (Casado-Díaz, Martínez-Bernabeu, and Rowe, 2017; Rowe, Casado-Díaz, and Martínez-Bernabeu, 2017) and so provides a suitable third method to test here. Its global optimization comes through replacing the deterministic procedure characteristic of both the TTWA and the LAm methods with a stochastic search, another innovation enabled by more rapid modern computing. However the fact that results of stochastic procedures are not replicable can be a problem for policy-makers, hence their very limited use in a policy context.

Table 1 summarizes the principal distinguishing features – as described above – of the three LMA definition methods that will be evaluated below. Table 1 does not list features these methods have in common, but by which they differ from some other regionalization methods that also fully cover a territory. Two such features should be mentioned briefly here. The first is that these methods are all agglomerative, unlike alternatives ranging from Monmonier (1973) to Farmer and Fotheringham (2012) in which a "top down" process subdivides a complete territory. The second common feature of these methods is that they have statistical criteria to determine when the agglomerative procedure stops. This differs from methods such as Intramax (Masser and Scheurwater, 1980) which continues its aggregation procedure until a single "region" includes

Table 1. Critical Characteristics of the Three Selected Exemplar LMA Definition Methods

Method	Logic?	Morphology?	Hierarchical?	Optimizing?	Policy use?
LAm	Deductive	Center-based	Fully	No	Yes
TTWA	Inductive	Unrestricted	Slightly	Partially (local)	Yes
GEA	Inductive	Unrestricted	No (stochastic)	Fully (global)	No

the whole territory, then having a second procedure to select a preferred set of interim groupings as its final results.

Each of the three methods to be evaluated has key parameters that determine the structure of the LMAs it defines. As a deductive method, LAm needs a criterion to identify which zones become the centers of LMAs: zones which are not centers are grouped with the center to which they send most commuters. This simple solo zone-to-zone flow criterion ignores the possibility that the non-center may send more commuters to a *combination* of zones which has already been produced by allocating other non-centers with a different center. A key feature of the LAm method is that its criteria are limited to the grouping procedure: there is no final set of criteria that test whether the defined LMAs satisfy appropriate statistical criteria, such as commuting self-containment[1] which is central to many LMA definition procedures. Whatever set of LMAs emerges from LAm's grouping process is the final set of boundaries. The assumption is that the criteria to identify centers and to guide the grouping procedure will generate well-defined LMAs. The general validity of this assumption will be tested below by applying the LAm method with its Swedish criteria to three contrasting countries.

This approach of testing the generalizability of a country's method and parameters is also followed by applying the official U.K. method and statistical criteria to all three countries studied here. The key criterion in the U.K. TTWA grouping procedure combines four commuting flow proportions in a formula developed from one initially devised by Smart (1974). The method continues grouping and re-grouping zones until all the LMAs meet statistical criteria that set minimum levels of both self-containment and working population size, with a trade-off between these parameters so the final LMAs can be appropriate in both metropolitan and remote rural areas (Coombes, 2010).

The third method applied to the three countries represents a departure from established approaches because it has a stochastic – rather than a deterministic – process that aims to maximize its global objective function (Martínez-Bernabeu, Flórez-Revuelta, and Casado-Díaz, 2012). The complexity of the optimization problem (viz. identifying the best set of LMAs for a given territory) makes a genetic algorithm appropriate. The area groupings emerge from an intensive trial-and-error process that gradually improves the quality of the set of LMAs in terms of the objective function. Given the focus here on comparing the results of different methods, the same minimum levels of self-containment and population size as are used in the TTWA analysis are also applied with this method. The difference with the GEA method is that it does not stop as soon as these criteria are met by all the potential LMAs, instead it continues testing alternative solutions to find the globally optimal definition of LMAs. This process involves a second set of parameters that controls the operation of the evolutionary algorithm by determining how intensely and efficiently the algorithm searches for better solutions before stopping and delivering its final results (cf., Li, Church, and Goodchild, 2014). These methodological parameters do not directly influence key characteristics of the LMAs such as their level of self-containment and size.

Characteristics of the three contrasting countries

Before any method could be generally recommended, it needs to produce useful results outside its country of origin, because most methods are developed in response not only to their origin country's geography, but also the nature of its commuting dataset. The three methods evaluated here are from countries with contrasting geographies: Sweden (LAm), United Kingdom (TTWA), Spain (GEA). Table 2 illustrates some differences between the countries, using data from their 2001 Censuses which provide closely comparable commuting datasets. The extensive remote areas of Sweden, and to a lesser extent Spain, are reflected in these countries' smaller populations occupying substantially larger territories than the United Kingdom. The methods to define LMAs analyze commuting flows between Census data zones. Table 2 shows that Spain and the United Kingdom have similar numbers of data zones despite their differing total populations, whereas the fewer zones in Sweden necessarily have a large average population.

Eurostat (2020) emphasizes that LMA definitions can be sensitive to the data zones analyzed. As in most countries, the data zones in Sweden and also Spain are administrative areas that often include one settlement of whatever size, resulting in a set of zones of widely differing population size (Table 3). The U.K. data zones are not administrative areas and in fact were defined to have very similarly sized populations, so that many split up large settlements while others cover large thinly populated areas. Table 3 shows that there are no extremely small zones in Sweden, where administrative area reform has been much more radical than in Spain. All the countries have a wider zone size range in terms of the number of jobs at workplaces than of working residents because workplaces are more clustered than are homes: this is magnified in the United Kingdom where the zone definitions restrict the population size range, leading to some including very

Table 2. Dimensions of the Three Countries Analyzed

2001 Data		Sweden	United Kingdom	Spain
Total territory	Working population (million)	4.09	26.62	14.73
	Land area (1,000 km^2)	410	247	502
	No. of Census commuting data zones	289	10,558	8,031
Zones	Mean no. of working residents (thousands)	14.1	2.5	1.8

Table 3. Dimensions of the Data Zones in the Three Countries Analyzed

2001 Data		Sweden	United Kingdom	Spain
Zones: no. of working residents	Minimum	1,033	237	0
	Median	6,731	2,032	153
	Maximum	374,121	17,725	1,225,956
Zones: no. of workplace jobs	Minimum	825	61	0
	Median	5,953	1,386	125
	Maximum	531,912	266,442	1,485,561
Zones: land area (km^2)	Minimum	8.72	0.13	0.03
	Median	676	5	35
	Maximum	19,371	3,321	1,750
% commuting that is intrazonal		70.6	23.8	68.6

many jobs in city centers like London. Table 3 also shows a median land area of only 5 km^2 for the U.K. zones and this inevitably results in many commuting flows crossing their boundaries. The proportion of U.K. commuting that is intrazonal is less than a quarter, whereas in the other countries over two-thirds of the workforce live and work in the same zone (Table 3). All of these contrasts in the zones of the three countries provide important tests of the transferability of the LMA definition methods between countries.

Outline results of the three exemplar methods in the three contrasting countries

Table 4 reveals that while the three methods produce[2] broadly comparable numbers of LMAs in the case of Sweden, the results are not similar in the other countries. The Swedish official method LAm defines one third fewer LMAs in its "home" country than do the other methods, even though in the other countries it defines around three times as many. Both the other methods apply statistical criteria to prevent their final results including LMAs that are inappropriate LMAs in terms of their self-containment or size, so their results in Sweden indicate that in its "home" country the LAm method is grouping some areas that can be appropriately left as separate LMAs. In other words, maximizing the number of plausible separate LMAs is not an explicit objective of the LAm method in the way that it is for TTWAs (Coombes, 2010). Yet when the LAm method was applied to data for Spain and the United Kingdom, it produced far more LMAs than did the other methods (Table 4), particularly single-zone LMAs. There is no simple contrast between the results of the other methods: GEA defined slightly fewer Swedish LMAs than the TTWA method but noticeably more than the latter in the other countries. The higher number of U.K. LMAs produced by GEA in the TTWA method's "home" country is very notable because an explicit aim of the TTWA method is to maximize the number of plausible separate LMAs it defines.

Table 5 confirms the similarity of the results of the three methods in Sweden, especially in the median size of their LMAs' working populations. All the methods define a similarly sized London LMA – the largest U.K. LMA – but the large number of U.K. LMAs defined by LAm causes their median size to be notably low (14,133). Yet this value appears relatively

Table 4. Number of Resulting LMAs (total and single-zone)

Method	Sweden		United Kingdom		Spain	
LAm	88	40	681	9	1,536	989
TTWA	126	53	218	0	492	45
GEA	120	58	265	0	583	54

Table 5. LMA Working Populations

	Sweden			United Kingdom			Spain		
Method	Min.m	Median	Max.m	Min.m	Median	Max.m	Min.m	Median	Max.m
LAm	1,253	14,537	1,082,322	464	14,133	3,619,455	1	262	2,447,627
TTWA	3,368	14,537	840,401	3,769	57,819	3,376,179	3,332	8,439	2,260,167
GEA	3,368	14,029	840,401	3,769	50,346	3,214,712	3,300	6,259	2,244,969

Table 6. LMA Land Area (km^2)

	Sweden			United Kingdom			Spain		
Method	Min.m	Median	Max.m	Min.m	Median	Max.m	Min.m	Median	Max.m
LAm	411	3,110	27,410	5	201	4,057	1	97	2,321
TTWA	139	1,825	33,561	144	882	5,272	13	658	8,138
GEA	139	1,786	36,239	58	707	5,061	13	623	6,864

unexceptional when compared to the median of the LAm's Spanish LMAs, 262, which indicates that its results include over 750 LMAs whose working populations are less than a tenth of the smallest LMA defined by either of the other methods. Table 6 reports the equivalent values for the land area size of the 3×3 sets of LMAs, with the LAm's median values in Spain (especially) and the United Kingdom again making its results exceptional and of doubtful plausibility. In terms of maxima and minima, Table 5 shows that in all countries LAm defined both the largest and the smallest LMAs in terms of working population, but in terms of land area its largest LMA is smaller than the equivalent LMAs defined by the inductive methods (Table 6).

Evaluation of the three methods in terms of their results

This section moves on from outline statistics on the three methods' results in the three countries to more formally evaluate the methods in terms of their success in defining LMAs in differing conditions. This evaluation requires relevant metrics, with each metric not only relating to the concept of the labor market (Fowler and Jensen, 2020), but also having been shown to discriminate appropriately between sets of LMAs. Martínez-Bernabeu, Coombes, and Casado-Díaz (2020) provides the basis for the selection of eight such metrics: first the relevant evaluation criteria (e.g., cohesion) were derived from the LMA concept itself, then a set of candidate metrics (e.g., minimum self-containment) were identified that reflect these criteria, and then the metrics were tested empirically to identify those providing appropriate assessments of many sets of LMAs. Table 7 shows that, for the four key criteria identified, eight candidate metrics were selected after the empirical testing (Martínez-Bernabeu et al. op. cit.) These metrics are applied below to evaluate LMAs defined in the three countries by the three methods and – on that basis – to evaluate the methods themselves. Each metric is described further below before its results are discussed.

Autonomy

Autonomy is the most widely used criterion of well-defined LMAs (e.g., van der Laan and Schalke, 2001), indicating how separate the LMAs are from each other in terms of commuting flow data. The key metrics are based on self-containment, with the simplest being the value of the LMA with the lowest self-containment. Table 8 reports this Minimum metric value for each of the 3×3 sets of LMAs, and these values show the impact of the difference in the number of LMAs produced by the LAm as against the other methods. As noted earlier, LAm defines fewer LMAs in its "home" country Sweden than the other methods, giving them the larger average size, which partly explains their higher minimum self-containment. The other methods' Minimum values in all countries reflect these methods' explicit self-containment requirement (66.6%). As noted earlier, LAm defined many more – and several much smaller – LMAs than the other methods in Spain and the United Kingdom. Table 8 shows that this results in minimum values in those countries (18.7%, 7.7%), indicating that some of these LMAs are far from reaching plausible levels of Autonomy.

Table 7. Metrics for Evaluating LMAs

Criterion	Metric
Autonomy	Minimum LMA self-containment: $\min_{M \in P} \frac{\sum_{i \in M} \sum_{j \in M} T_{ij}}{max\left(\sum_{i \in M} \sum_j T_{ij}, \sum_{i \in M} \sum_j T_{ji}\right)}$
	Median of LMAs' self-containments: $\operatorname{median}_{M \in P} \frac{\sum \sum_{j \in M} T_{ij}}{max\left(\sum_{i \in M} \sum_j T_{ij}, \sum_{i \in M} \sum_j T_{ji}\right)}$
	Global self-containment: $\frac{\sum_{M \in P} \sum_{i \in M} \sum_{j \in M} T_{ij}}{\sum_{M \in P} O_M}$
Homogeneity	1 – (Gini coefficient of LMAs' working population sizes): $1 - \frac{\sum_{M \in P} \sum_{N \in P} \lvert w_M - w_N \rvert}{2n(P)^2 \overline{w}}$
	1 – (Gini coefficient of LMAs' land area sizes): $1 - \frac{\sum_{M \in P} \sum_{N \in P} \lvert a_M - a_N \rvert}{2n(P)^2 \overline{a}}$
Balance	1 – (Gini coefficient of LMAs' job ratios): $1 - \frac{\sum_{M \in P} \sum_{N \in P} \left\lvert \frac{O_M}{D_M} - \frac{O_N}{D_N} \right\rvert}{2n(P)^2 \overline{O/D}}$
Cohesion	Interaction index: $\sum_{M \in P} \sum_{i \in M} \left(\frac{\left(\sum_{j \in M}(T_{ij}) - T_{ii}\right)^2}{\sum_{k \in P} T_{ik} \cdot \left(\sum_{j \in M} \sum_{k \in P} T_{kj} - \sum_{k \in P} T_{ki}\right)} + \frac{\left(\sum_{j \in M}(T_{ji}) - T_{ii}\right)^2}{\sum_{k \in P} T_{ki} \cdot \left(\sum_{j \in M} \sum_{k \in P} T_{jk} - \sum_{k \in P} T_{ik}\right)} \right)$ [n.b. less reliable in regionalizations with more single-zone LMAs]
	Number of LMAs [n.b. a proxy, not consistently reliable]

Source: Martínez-Bernabeu, Coombes, and Casado-Díaz (2020).

Table 8. Autonomy: Three Metrics Related to LMAs' Self-containment

	Sweden			United Kingdom			Spain		
Method	Min.m	Median	Global	Min.m	Median	Global	Min.m	Median	Global
LAm	70.7%	87.5%	93.0%	18.7%	59.0%	72.2%	7.7%	85.2%	94.4%
TTWA	68.0%	82.7%	87.4%	66.7%	76.4%	81.4%	68.0%	85.9%	90.7%
GEA	68.0%	84.4%	88.7%	66.7%	73.6%	78.6%	66.7%	85.2%	90.1%

A more holistic assessment is provided by also considering the Median self-containment of each set of LMAs. With this metric too the sheer number of LMAs can be influential, with the LAm's relatively few Swedish LMAs having a rather high Median whereas the value for its numerous U.K. LMAs' shows over half of them to be under 60% self-contained (Table 8). Due to the LAm setting no minimum size, its analysis of Spain leaves very many tiny remote areas ungrouped; most of these have few commuters in or out, resulting in a Median self-containment value similar to that of the far fewer LMAs defined by the other methods. The third metric of Autonomy is the Global level of self-containment of a set of LMAs: the proportion of a country's commuters who work within the same LMA as they live. Table 8 shows high values on all the metrics and in all countries for the LMAs defined by the methods that, unlike LAm, are inductive rather than deductive and do set minima for self-containment and size. Again, the relative numbers of LMAs is influential: GEA produced fewer Swedish LMAs than the TTWA method and so tends to have slightly higher Autonomy metric values, with the same relationship holding in the other countries where the TTWA method defined fewer LMAs and so tends to have the higher metric values.

Homogeneity

Homogeneity metrics reflect a preference for those sets of LMAs that vary less in size, in terms of either working population or land area (Franconi, Ichim, and D'Aló, 2017). Table 9 shows that the LAm's LMAs have the lowest values – greatest size variance – on both metrics, except in

Table 9. Homogeneity: Two Metrics Related to LMAs' Size Distribution

Method	Sweden		United Kingdom		Spain	
	Working population size	Land area size	Working population size	Land area size	Working population size	Land area size
LAm	0.2825	0.5285	0.3220	0.4289	0.0692	0.2869
TTWA	0.3807	0.4716	0.3894	0.6299	0.2891	0.4977
GEA	0.3602	0.4704	0.4000	0.5921	0.2701	0.5524

terms of land area in its "home" country of Sweden. The especially wide variation in its LMAs' population size is unsurprising given that its LMAs included both the largest and smallest LMA in each country (Table 5). The many single-zone Spanish LMAs result in LAm's exceptionally low population Homogeneity (0.0692). This metric value is further reduced by LAm defining very large LMAs around larger cities. LAm defines such large metropolitan LMAs not only because it limits the number of zones that can be LMA "centers," but also because of the way it groups non-center zones into the LMAs. This grouping uses a simple zone-to-zone flow criterion which ignores the possibility that more of a non-center zones' commuters may work in other non-center zones than in any center; the resulting metropolitan-size LMA may encompass a polycentric structure that "hides" several linked but relatively self-contained LMAs. These outcomes are traceable to the LAm method being developed to fit the situation in Sweden where the metropolis is moderately sized and most data zones encompass whole settlements *and* nearby areas; neither of these conditions is found in Spain or the United Kingdom. There are only slight differences in the metric values for the other methods' LMAs (Table 9). One consistent finding across all 3×3 sets of LMAs is that the land area metric values are higher than those for working population. The explanation is that people tend to avoid lengthy commuting trips, leading to local clusters of flows dominating the commuting patterns and most analyses producing LMAs whose area size reflects a "reasonable" commuting distance, even though the LMAs' population sizes do vary widely.

Balance

A defining characteristic of LMAs is that they link labor demand to labor supply by internalizing commuting flows, so there should be a reasonable balance between LMAs' working resident populations and their workplace job numbers. A "perfect" set of LMAs would all have a job ratio (the number of jobs divided by the size of the working population) equal to 1. An autonomy of 100% implies a job ratio of 1, but a job ratio of 1 does not ensure perfect autonomy. This metric shows again that LAm performs best in its "home" country of Sweden but noticeably less well than the other methods in the other two countries (Table 10). There is relatively little difference in this metric value for the LMAs produced by the other two methods; once more the method that defined fewer LMAs in a country having the higher metric value there (this is GEA in the Swedish case but TTWA in the other countries).

Cohesion

Table 11 presents the values for two metrics of Cohesion, with one simply the number of LMAs that each method defined in that country. A larger number of LMAs tends to mean fewer large LMAs that would inevitably include many zone pairs with few direct interactions between them.

Table 10. Balance: Metric Based on the Job Ratio

Method	Sweden	United Kingdom	Spain
LAm	0.9802	0.9129	0.9208
TTWA	0.9669	0.9667	0.9634
GEA	0.9705	0.9536	0.9623

Table 11. Cohesion: Two Metrics with Different Limitations

	Sweden		United Kingdom		Spain	
Method	No.	Interaction index	No.	Interaction index	No.	Interaction index
LAm	88	0.0676	681	0.0797	1536	0.0322
TTWA	126	0.0741	218	0.0641	492	0.0302
GEA	120	0.0770	265	0.0687	583	0.0324

Martínez-Bernabeu, Flórez-Revuelta, and Casado-Díaz (2012) identified the count of LMAs as only a crude metric of Cohesion, while acknowledging the seminal Goodman (1970) article's view that defining fewer larger LMAs to achieve higher levels of Autonomy tends to produce lower levels of Cohesion due to ignoring LMAs' "essentially local character" (p. 185). Table 11's other metric is an interaction index that measures the level of intra-LMA interaction between LMAs' constituent zones. This metric also has a limitation, because no value is obtainable for LMAs comprising a solitary zone.

The fact that LAm produces very many LMAs in both Spain and the United Kingdom has been identified above as problematic, which underlines the uncertain value of a count of LMAs as an indicator of the positive LMA characteristic of Cohesion. The limitation of the interaction index, the other metric of Cohesion, applies to sets of LMAs of which a large proportion comprise a single zone: this is critical in the case of LAm whose results in the United Kingdom and especially Spain include so many separate LMAs that single zone LMAs are very numerous. This is not an issue for the LMAs produced by the inductive methods. Table 11 shows that in all three countries the interaction index value for the GEA's LMAs is higher than those for the TTWA method defined. This is particularly significant in the case of Sweden where the TTWA method's LMAs were the more numerous. The way that the GEA repeatedly readjusts its results until reaching a more globally optimal solution has a clear benefit in terms of Cohesion.

Discussion

The metrics in the empirical analyses above were developed by Martínez-Bernabeu, Coombes, and Casado-Díaz (2020) where the TTWA and Intramax methods were used to define LMAs in the United States to test alternative possible evaluation metrics. These tests examined how the value of a metric varied as the aggregation process proceeds from its starting point, when every zone is separate, through to its completion when a single "region" incorporates every zone. However, the principal focus of the discussion was on sets of LMAs that provide a level of granularity (i.e., average size of region) which was similar to existing policy-relevant boundary sets, such as the Metropolitan Areas. Those results in Martínez-Bernabeu et al. (op. cit.) provide additional evidence on the possible relationship highlighted above between the values for certain

Table 12. Relationship Between Metric Values and Number of LMAs (Higher Levels of Granularity)

Metric	TTWAs (Martínez-Bernabeu et al.)	LAm: TTWA/GEA (excluding Spain)	TTWA: GEA (mainly Spain and United Kingdom)
Autonomy: minimum self-containment	Negative	Negative	Negative
Autonomy: median self-containment	Negative	Negative	Negative
Autonomy: global self-containment	Negative	Negative	Negative
Homogeneity: working population size	Negative	Uncertain	Uncertain
Homogeneity: land area size	Negative	Negative	Uncertain
Balance: distribution of job ratios	Negative	Negative	Negative
Cohesion: interaction index	Positive	Positive	Uncertain

metrics and the number of LMAs defined by a method. Any such relationship needs to be identified here before any interpretation of metric values as showing the superiority of method X over method Y. If the value of a metric clearly varies according to the granularity of a set of LMAs, then its values will only be very strong evidence of the superiority of method X if both sets of LMAs are similar in number. Nevertheless the indicators can be used to find the desired level of granularity, choosing the one with the preferred trade-off between the values on the main indicators. For example, if cohesion or spatial detail were deemed more important for a given programmatic purpose, the best regionalization will be the one that maximizes the cohesion indicators subject to some minimum levels of self-containment. If self-containment was more important, the one with maximum global autonomy subject to some maximum region size levels would be preferred.

Table 12 summarizes evidence on the relationship between each metric's values and the number of LMAs. A negative relationship exists when a metric tends to be lower if a set of LMAs has greater granularity (i.e., if there are more LMAs), with the metric value tending to rise as the number of LMAs falls due to the aggregation process. Table 12 first draws on evidence in the interpretation of the empirical testing by Martínez-Bernabeu et al. (op. cit.). Table 12's other columns draw on the empirical results presented above. The middle column compares the values of LAm's LMAs to those of the LMAs defined by the inductive methods (which tend to have similar values) *but* this comparison sets aside the problematic Spanish results from LAm. Table 12's final column considers the rather slight difference between the values for the sets of LMAs defined by the TTWA method and GEA; here it is the Swedish results that are less helpful in testing for a relationship with granularity because these methods produced very similar numbers of LMAs.

Table 12's evidence offers some support for the assumptions underpinning the Goodman (1970) warning about greater levels of aggregation: restraining the aggregation process to keep a high level of granularity does have a positive effect on the LMAs' level of Cohesion, but a negative effect on Autonomy. In broad terms, the evidence here is that aggregation – reducing granularity – tends to decrease Cohesion but increase not only Autonomy, but also Balance, with a similarly negative but less clear-cut relationship between granularity and Homogeneity.

The motivation for compiling the metric values was to provide evidence in an evaluation of the results of three different LMA definition methods in three different countries. The initial focus of this evaluation compares the deductive LAm method's LMAs to those defined by inductive methods. Evaluation starts with Autonomy metrics because this characteristic is core to the concept of the LMA. One of these metrics, minimum self-containment, reports the value for one LMA and thus mainly shows whether a self-containment minimum was part of the method. In its "home" country Sweden LAm defined one third fewer LMAs than did the other methods, so the LAm's LMAs' higher median and global self-containment metric values for LAm's Swedish LMAs is partly due to the inverse relationship between granularity and Autonomy. The most dramatic Autonomy metric value is the extremely low median self-containment of LAm's LMAs in the United Kingdom (Table 8) that shows the majority are under 60% self-contained, and this is a clear failing of the method.

The number of LAm-defined Swedish LMAs is lower than the number defined by both inductive methods even though the latter methods require all their LMAs to be sufficiently self-contained. LAm's center and hinterland method defines larger LMAs around the main cities and these include some outlying areas with reasonably high levels of Autonomy. The uneven size of the LAm-defined Swedish LMAs results in a far lower population Homogeneity metric value than the LMAs the inductive methods defined, despite the evidence of Martínez-Bernabeu et al. (op. cit.) for an inverse relationship between this metric and granularity. Comparing the results of the LAm to the inductive methods' LMAs thus reveals that even in its "home" country it failed to identify some potential LMAs with appropriate levels of Autonomy, and defined a set of LMAs with a low level of population size Homogeneity. More critical still is the evidence from the other countries. LAm's center and hinterland method, created for the Swedish data zones, produces problematic results in Spain and the United Kingdom where many data zones are small and/or fragments of settlements.

The other focus for the evaluation is the comparison of the inductive methods' LMAs. It is useful to first look at the values in Spain and the United Kingdom; here the LMAs defined by the TTWA method are markedly outnumbered by those defined by GEA, so the relationship between each metric and the level of granularity must be considered. Most of these relationships are negative (Table 12), so the lower granularity of the TTWA-defined LMAs can partly explain their higher Autonomy values of median and global self-containment, and the higher value for the metric of Balance. The relationships between granularity and the Homogeneity metrics was less clear. The relationship of granularity with the interaction index metric is generally positive and in both countries the GEA's LMAs have the higher interaction index value which accords with their higher granularity.

The inductive methods defined similar numbers of Swedish LMAs – only slightly fewer by GEA – allowing metric values to be interpreted with little concern over granularity. While the difference between these sets of LMAs' median and global metric values are not great, both show the GEA defined areas to have higher Autonomy. GEA's LMAs also have the higher value for Balance, although its Homogeneity values are marginally the lower. Perhaps of greatest interest is the higher interaction index value for GEA's LMAs that shows them to have greater Cohesion, despite the TTWA method's LMAs having the slightly higher level of granularity that otherwise might be seen as a crude indicator of a higher level of Cohesion.

Conclusions

Brandmueller et al. (2017) document how evidence-based territorial economic and social policy depends on analyzing data for appropriately defined areas, which increasingly means for

LMAs. One crucial feature of LMAs is that they cover a whole territory, unlike the recent OECD-EU definitions of major urban regions (Dijkstra, Poelman, and Veneri, 2019). There are alternative approaches to the definition of LMAs, so this article evaluated three representative definition methods. If a single method can produce suitable LMA boundaries in contrasting countries – that is, it has proven transferability – then the LMAs that it defines can provide the basis for meaningful policy-relevant *cross-national* analyses at a local scale.

Most recent methods for defining LMAs or other functional regions adopt an inductive approach, unlike the earlier deductive methods that assumed every defined region had the same structure (usually one urban center and its hinterland). One of the three methods evaluated here was the deductive center-based method to define the official Swedish LMAs. The other methods evaluated were inductive: the method that defines the official U.K. LMAs, plus a stochastic method developed by academics in Spain. The three “home” countries of these methods provide highly contrasting geographies and datasets to test method transferability (i.e., the extent to which they define appropriate LMAs across a diversity of countries). One surprising empirical finding was that the inductive methods defined more LMAs than the deductive method in the latter’s “home” country of Sweden. Rather more importantly, the deductive method produced implausibly large numbers of areas in the other countries. The deductive method readily identifies LMA “centers” among the large Swedish data zones – whose boundaries reflect local settlement patterns – but it also interpreted as “centers” many small Spanish data zones in remote regions. The deductive method was also unable to find plausible “centers” among the U.K. data zones, (whose small population size constraint means most of them are no more than neighborhoods).

Although the labor market concept provides the template for appropriate definitions of LMAs, empirical evaluation of different LMA definition methods’ results is rare. This article draws upon Martínez-Bernabeu, Coombes, and Casado-Díaz (2020) by applying their evaluation metrics to the three LMA definition methods’ results in three different countries. These metrics formalize a transferability test here because they assess whether a method can produce appropriate LMAs not only in its “home” country, but also where conditions are very different. Some preliminary methodological findings concerned relationships between these metrics, with more LMAs (and hence reduced LMA average size) associated with lower metric values related to Autonomy, which is core to the concept of the LMA. There is a similarly negative relationship between granularity and Balance, while the negative relationship is less clear-cut between granularity and Homogeneity. One positive relationship is between granularity and Cohesion: if a country is divided into more, smaller on average, LMAs, then those LMAs will tend to be more internally integrated.

Autonomy metrics are of primary interest in evaluating sets of LMAs, and hence the methods that produced them. The low minimum self-containment value for the LAm’s LMAs reflect the absence of a self-containment minimum in that method. One effect is a median self-containment value of under 60% for the U.K. LMAs defined by LAm, which is strong evidence that the method fails to consistently defined robust LMAs. At the same time in its “home” country of Sweden LAm fails to identify some of the robust small LMAs that are defined by both the inductive methods. LAm’s Swedish LMAs also have rather low population size Homogeneity. Most critical of all for the key concern here with transferability is the clear evidence from results in Spain and the United Kingdom that the LAm’s center and hinterland method creates problematic results in countries where data zones are small and/or fragments of settlements. A much more general problem for deductive methods is that they presume all LMAs have the same morphology, and this cannot be the basis for a harmonized method that can define robust LMAs across the huge diversity of the EU (Franconi, Ichim, and D'Aló, 2017), let alone elsewhere. Indeed the diversity *within* most countries is too great to fit within a single structure, leading the OECD (2020) to conclude that any country including any non-metropolitan areas will find a center and hinterland approach an inadequate basis for defining a full set of LMAs.

Turning to a comparison of the inductive methods, it is necessary to recall the relationships between most metrics and granularity, because the TTWA method defined notably fewer LMAs than did GEA in both Spain and the United Kingdom. The TTWA LMAs' lower granularity partly explains their higher values of Autonomy, and higher value for Balance. The interaction index metric values also conform to expectations: this relationship is generally positive and in both countries the GEA's LMAs have the higher metric value that accords with their greater level of granularity. However, the inductive methods defined similar numbers of Swedish LMAs, so these metric values can be interpreted with little concern over granularity.

The differences between the metric values for the Swedish LMAs defined by the inductive methods are small, but both show that the GEA defined areas have higher levels of Autonomy. Of greatest interest is the higher interaction index value for GEA's LMAs, which indicates their higher level of Cohesion (despite the slightly higher granularity of the TTWA method's LMAs). This is clear evidence that the stochastic nature of the GEA method helps it statistically outperform the TTWA method due to the latter's limited capacity for optimization (Watts, 2013).

The conclusions in this article are based on findings from empirical analyses of commuting data, focused on "sensitivity" to methods, and as such no allowance was made for any effects of errors in those datasets (cf., Foote, Kutzbach, and Vilhuber, 2021), or for the use of different datasets for the same country. Future work is, therefore, to investigate sensitivity to data issues. Another limitation to this evidence is its spatial coverage (three European countries). Even so this testing of the LMA definition methods' transferability has produced the unarguable finding that the deductive method LAm is unable to define robust sets of LMAs in either Spain or the United Kingdom, partly due to the data zones there being small and/or fragments of settlements. It is possible that the evaluation of the inductive methods is sensitive to the small number of countries analyzed, but the finding that the repeated optimization by the stochastic approach produces statistically superior results is likely to be generally applicable. At the same time, this form of analysis is not ideal in a policy context because of not being independently replicable.

This article suggests further openings in the already vibrant research field of functional region definition methods. One emerging challenge comes from a new emphasis on publishing official data for grid squares (United Nations, 2019). If this trend extends to the commuting datasets that are critical to LMA definition then, because grid squares are clearly fragments of settlements, inductive methods would be the only option. Further developments of the optimization-based methods such as GEA are possible, but the recent interest in optimizing "modularity" (Newman and Girvan, 2004) of functional regionalizations has been shown to offer no real benefits because its null model is not appropriate for spatially-constrained interaction networks (Martínez-Bernabeu and Casado-Díaz, 2021). Examining alternative null models could help assess the utility of modularity for functional regionalization. Most existing approaches optimize a single indicator, such as one representing Cohesion, but one alternative could be to maximize a synthetic index that includes several metrics, which each represent one of the critical criteria (e.g., Cohesion, Autonomy, and Homogeneity). This index could offer a flexibility to the policy-maker by them deciding how to weight each indicator to reflect how important the criteria it represents is in that specific policy context. At the same time, there could be a minimum threshold on any metric that reflects a criterion that is especially important (e.g., Autonomy).

The comparison of regionalizations from different methods in this article is a quantitative evaluation of alternative sets of LMAs. It shows how the metrics used could enable policy-makers make more informed decisions about the boundaries that are most appropriate for a specific purpose. Analysts carrying out regionalizations could use the metrics to guide the setting of their parameters so that the results better fit the specific purpose for which the defined boundaries will be used. To help disseminate the insights from this work among the broader policy-making community, one step forward could be to create open-source software that accepts as inputs the grouping of areas into LMAs, plus the relevant commuting data, to then output the indicator values for those LMAs.

Acknowledgements

Colin Wymer (Newcastle University) was, along with the authors, a key member of the research team that produced the nine sets of Labor Market Areas used in the analyses for this article. Funding for that research project was received from Eurostat (who facilitated access to the Census commuting datasets) via sub-contracts that were administered by DevStat 2011–2012. Funding for related Spanish analyses was via Project PID2020-114896RB-100/Agencia Estatal de Investigación, AEI/10.13039/501100011033, Spanish State Plan for Scientific and Technical Research, and Innovation 2017–2020, and Project AICO/2021/062, Generalitat Valenciana, Department of Innovation, Universities, Science and Digital Society.

Notes

1 Self-containment is a variable that considers either supply-side self-containment (the proportion of an area's employed population that works within the area) or demand-side self-containment (the proportion of jobs within an area that are filled by residents of that area); the key criterion in the TTWA method combines both indicators by assessing an area in terms of the lower of its supply-side and demand-side self-containment values.

2 Maps of the three methods' results in each country are provided in Supporting Information.

References

Brandmueller, T., G. Schäfer, P. Ekkehard, O. Müller, and V. Angelova-Tosheva. (2017). "Territorial Indicators for Policy Purposes: NUTS Regions and Beyond." *Regional Statistics* 7, 78–89.

Brown, L., and J. Holmes. (1971). "The Delimitation of Functional Regions, Nodal Regions, and Hierarchies by Functional Distance Approaches." *Journal of Regional Science* 11, 57–72.

Casado-Díaz, J. M., and M. Coombes. (2011). "The Delineation of 21st Century Local Labour Market Areas: A Critical Review and a Research Agenda." *Boletín de la Asociación de Geógrafos Espanoles* 57, 7–32 & 393–396.

Casado-Díaz, J. M., L. Martínez-Bernabeu, and F. Rowe. (2017). "An Evolutionary Approach to the Delineation of Labour Market Areas: An Empirical Application to Chile." *Spatial Economic Analysis* 12, 379–403.

Cattan, N. (2002). *Redefining Territories: The Functional Regions*. Paris: Organisation for Economic Co-operation and Development.

Champion, T., and G. Hugo. (2004). "Introduction: Moving Beyond the Urban-Rural Dichotomy." In *New Forms of Urbanization: Moving Beyond the Urban-Rural Dichotomy*, 3–24, edited by T. Champion and G. Hugo. Aldershot: Ashgate.

Coombes, M. (2010). "Defining Labour Market Areas by Analysing Commuting Data: Innovative Methods in the 2007 Review of Travel-to-Work Areas." In *Technologies for Migration and Population Analysis: Spatial Interaction Data Applications*, 135–51, edited by J. Stillwell, O. Duke-Williams, and A. Dennett. Hershey, PA: IGI Global.

Coombes, M., and S. Bond. (2008). *Travel-to-Work Areas: The 2007 Review*. London: Office for National Statistics.

Coombes, M., J. M. Casado-Díaz, L. Martínez-Bernabeu, and F. Carausu. (2012). "Study on Comparable Labour Market Areas – Final Research Report." Eurostat-Framework contract, (6001), 2008-001, available at https://www.istat.it/it/files/2014/12/ANNEX-I_Final-report_LMA.pdf

Dijkstra, L., H. Poelman, and P. Veneri. (2019). *The EU-OECD Definition of Functional Urban Areas*. Paris: Organisation for Economic Co-operation and Development.

Eurostat (2020). European Harmonised Labour Market Areas: Methodology on Functional Geographies with Potential Eurostat Statistical Working Paper, European Union, Luxembourg.

Farmer, C., and A. Fotheringham. (2012). "Network-Based Functional Regions." *Environment and Planning A* 43, 2723–41.

Foote, A., M. Kutzbach, and L. Vilhuber. (2021). Recalculating … : How Uncertainty in Local Labor Market Definitions Affects Empirical Findings. (Center for Economic Studies Working Paper 17-4R), Washington DC: U.S. Census Bureau.

Forstall, R., R. Greene, and J. Pick. (2009). "Which Are the Largest? Why Lists of Major Urban Areas Vary So Greatly." *Tijdshcrift voor Economische en Sociale Geografie* 100, 277–97.

Fowler, C., and L. Jensen. (2020). "Bridging the Gap Between Geographic Concept and the Data we Have: The Case of Labor Markets in the USA." *Environment and Planning A* 52, 1395–414.

Franconi, L., D. Ichim, and M. D'Aló. (2017). "Labour Market Areas for Territorial Policies: A European Approach." *Statistical Journal of the IAOS* 33, 585–91.

Goodman, J. (1970). "The Definition and Analysis of Local Labour Markets: Some Empirical Problems." *British Journal of Industrial Relations* 8, 179–96.

Grasland, C., and M. Madelin, eds. (2006). *The Modifiable Areas Unit Problem* 756. Luxembourg: ESPON (Final Report 3.4.3).

Hall, P., and D. Hay. (1980). *Growth Centres in the European Urban System*. London: Heinemann.

Klapka, P., and M. Halás. (2016). "Conceptualising Patterns of Spatial Flows: Five Decades of Advances in the Definition and Use of Functional Regions." *Moravian Geographical Reports* 24, 2–11.

Lankford, P. (1969). "Regionalization: Theory and Alternative Algorithms." *Geographical Analysis* 1, 196–212.

Li, W., R. Church, and M. Goodchild. (2014). "The p-Compact-Regions Problem." *Geographical Analysis* 46, 250–73.

Martínez-Bernabeu, L., and J. M. Casado-Díaz. (2021). "Standard Modularity Is Unsuitable for Functional Regionalization of Spatial Interaction Data." *Papers in Regional Science* 100(5), 1323–30. https://doi.org/10.1111/pirs.12617

Martínez-Bernabeu, L., and J. M. Casado-Díaz. (2022). "Automatic Delimitation of Labour Market Areas Based on Multi-Criteria Optimisation: The Case of Spain 2011." *Environment and Planning B: Urban Analytics and City Science* 49(2), 654–70. https://doi.org/10.1177/23998083211021104

Martínez-Bernabeu, L., M. Coombes, and J. M. Casado-Díaz. (2020). "Functional Regions for Policy: A Statistical 'Toolbox' Providing Evidence for Decisions Between Alternative Geographies." *Applied Spatial Analysis and Policy* 13, 739–58. https://doi.org/10.1007/s12061-019-09326-2

Martínez-Bernabeu, L., F. Flórez-Revuelta, and J. M. Casado-Díaz. (2012). "Grouping Genetic Operators for the Delineation of Functional Areas Based on Spatial Interaction." *Expert Systems with Applications* 39, 6754–66.

Masser, I., and J. Scheurwater. (1980). "Functional Regionalisation of Spatial Interaction Data: An Evaluation of some Suggested Strategies." *Environment and Planning A* 12, 1357–82.

Monmonier, M. (1973). "Maximum-Difference Barriers: An Alternative Numerical Regionalization Method." *Geographical Analysis* 5, 245–61.

Newman, M., and M. Girvan. (2004). "Finding and Evaluating Community Structure in Networks." *Physical Review E* 69, 026113.

Obaco, M., V. Royuela, and V. Xavier. (2020). "Identifying Functional Urban Areas in Ecuador Using a Varying Travel Time Approach." *Geographical Analysis* 52, 107–24.

OECD. (2020). *Delineating Functional Areas in All Territories OECD Territorial Reviews*. Paris: OECD Publishing.

Openshaw, S. (1984). *The Modifiable Areal Unit Problem*. (CATMOG 38). Norwich: Geobooks.

Openshaw, S., and P. Taylor. (1981). “The Modifiable Areal Unit Problem.” In *Quantitiative Geography*, 60–70, edited by N. Wrigley and R. J. Bennett. Henley-on-Thames: Routledge and Kegan Paul.
Rowe, F., J. M. Casado-Díaz, and L. Martínez-Bernabeu. (2017). “Functional Labour Market Areas for Chile.” *REGION* 4(3), R7–9. https://doi.org/10.18335/region.v4i3.199
SCB. (1992). “Local Labour Markets and Employment Regions, A New Regional Division of Labour Market Areas in Sweden Grounded on Commuting Statistics. (Labour Information 1991:7, Technical Report).” Stockholm: Statistics Sweden.
Schubert, U., S. Gerking, A. Isserman, and C. Taylor. (1987). “Regional Labour Market Modelling: A State of the Art Review.” In *Regional Labour Markets*, 53–94, edited by M. M. Fischer and P. Nijkamp. Amsterdam: North-Holland.
Smart, M. W. (1974). “Labour Market Areas: Uses and Definition.” *Progress in Planning* 2, 238–353.
Sunstein, C. (2010). “2010 Standards for Defining Metropolitan and Micropolitan Statistical Areas.” *Federal Register* 75, 37246–52.
Taubenböck, H., M. Reiter, F. Dosch, T. Leichtle, M. Weigand, and M. Wurm. (2021). “Which City Is Greener? A Multi-Dimensional Deconstruction of City Rankings.” *Computers, Environment and Urban Systems* 89, 101687.
United Nations. (2019). *The Global Statistical Geospatial Framework*. New York: UN Department of Economic and Social Affairs.
van der Laan, L., and R. Schalke. (2001). “Reality Versus Policy: The Delineation and Testing of Local Labour Market and Spatial Policy Areas.” *European Planning Studies* 9, 201–21.
Watts, M. (2013). “Assessing Different Spatial Grouping Algorithms: An Application to the Design of Australia’s New Statistical Geography.” *Spatial Economic Analysis* 8, 92–112.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher’s web site.