Functional Regions for Policy: a Statistical 'Toolbox' Providing Evidence for Decisions between Alternative Geographies

Lucas Martínez-Bernabeu[1] · Mike Coombes[2] 🆔 · José Manuel Casado-Díaz[3]

Abstract

Labour market areas and other functional regions (FRs) are increasingly used within research and policy, but how FRs are best defined is an unresolved issue. This is important because the policy impacts, or the research results, will differ depending on the specific FR boundaries used. As a result of this sensitivity (termed the Modifiable Areal Unit Problem), quantitative metrics are needed so that differing sets of FR boundaries can be evaluated. To meet this need the paper firstly reviews the concept and use of labour market areas – the form of FRs most widely used in policy – to identify relevant criteria for evaluating any regionalisation comprising a set of FRs. Next a range of potential measurable indicators for each of the criteria is defined. These candidate indicators are then exemplified by applying them to a huge number of alternative sets of FRs. From this empirical evidence a short-list of preferred indicators is identified, creating a statistical 'toolbox' for evaluating sets of FRs. The paper ends by first sketching possible processes within which applying the indicators can help policy-makers with a decision over the appropriate set of FRs for a specific policy, before finally outlining some potential future research developments.

Keywords Functional regions · Policy boundaries · Evaluation indicators · Spatial interaction · Commuting self-containment · Labour market areas

✉ Mike Coombes
mike.coombes@newcastle.ac.uk

✉ José Manuel Casado-Díaz
jmcasado@ua.es

[1] International Economics Institute, University of Alicante, Alicante, Spain

[2] Centre for Urban and Regional Development Studies, Newcastle University, Newcastle upon Tyne, United Kingdom

[3] International Economics Institute and Department of Applied Economic Analysis (Aulario I), University of Alicante, Carretera de San Vicente del Raspeig s/n, 03690 San Vicente del Raspeig –, Alicante, Spain

⚫ Springer

Introduction

Functional regions such as labour market areas (LMAs) have become increasingly valued by economic policy-makers because they can internalise the effects of policy interventions due to being 'composed of areas … which have more interaction or connection with each other than with outside areas' (Brown and Holmes 1971, p. 57). These attributes of local cohesiveness and relative separateness mean that for much sub-national economic policy 'LMAs are potentially the most appropriate spatial units' (Brandmueller et al. 2017). Administrative areas rarely possess these attributes because their boundary definitions are constrained by history and indeed are 'often more or less arbitrary' (Eurostat 2017, p.4). Functional region boundaries are updatable to reflect changing patterns of social and economic relations by, in most cases, analysing the most recent data on commuting flows. The advantages of policy relevance and practical updatability prompted the USA to devise official LMA definitions over 50 years ago (Adams et al. 1999), followed by numerous European countries subsequently (Cattan 2002).

The importance for policy-makers of using the most appropriate set of areas stems from the Modifiable Areal Unit Problem (Openshaw 1984), the phenomenon of different results being produced by the same form of analysis of the same base data when the set of areas used for the analysis is changed, whether in terms of scale or basic zonation. Differences between the results produced using administrative areas and those using FRs are illustrated in studies for Eurostat (Coombes et al. 2012) and for OECD (Veneri 2013). Other evidence of the issue in policy-related contexts includes van der Laan and Schalke (2001), Séguin et al. (2012), and also Gutiérrez Posada et al. (2017). A related finding on policy implementation by Cheshire and Magrini (2008) is that cities in Europe grew faster if the areas over which their policies were delivered were recognizably FRs.

There is a long history of research on methods to define FRs: recent contributions include Flórez-Revuelta et al. (2008), Fusco and Caglioni (2011), Farmer and Fotheringham (2012), Koo (2012), Martínez-Bernabeu et al. (2012) and Kim et al. (2013). Schubert et al. (1987) called for a consensus on best practice in regionalisation methods but this remains elusive. Rather than enter that highly technical debate, this paper aims to identify appropriate measures for evaluating alternative sets of LMAs (cf. Cörvers et al. 2009). Creating such a 'toolbox' of statistical indicators for analysing commuting flow data echoes the listing of migration pattern indicators by Stillwell et al. (2014). Here the aim is to provide appropriate indicators for policy-makers who need to identify which set of FRs in a territory is most relevant to a given policy concern. Several 'candidate' sets of FRs might have been produced by alternative methods, or by changing parameters of the same method to define alternative sets of FRs (eg. 2 regionalisations of a territory, with one constituting a 'lower-tier' with more FRs which are therefore smaller on average).

To summarise, the objective of this paper is to identify relevant quantitative indicators for evaluating alternative sets of FRs. These statistical indicators could be used by a policy-maker in conjunction with other evidence, such as local knowledge, to select the most appropriate set of FRs for a particular policy. The approach taken is as follows. Section 2 reviews the concept and policy uses of LMAs – the form of FRs most widely used in policy – to identify the most relevant criteria for evaluating any regionalisation comprising a set of FRs. Section 3 specifies potential measurable indicators for each of these criteria. Section 4

exemplifies these indicators by applying them to a huge number of alternative sets of FRs (produced by two different methods applied to commuting data from the 2000 USA Census), and this empirical evidence leads to a short-list of preferred indicators. Section 5 then sketches possible processes within which applying the indicators could help policy-makers to identify an appropriate set of FRs, and outlines potential future research developments based on the findings of this paper.

Principles for the Evaluation of LMAs

The LMA concept derives from economic principles in which a market area is the locality where demand for, and supply of, a commodity such as labour meet and fix a price which applies over that area (Miron 2010). Although in the case of LMAs workers in different types of job are paid different rates, Goodman (1970) argued that LMAs which reflect the job search areas of most of the labour force are definable by analysing commuting by the total workforce. He set out two key requirements for well-defined LMAs: firstly that relatively few commuters cross their boundaries, and secondly that the pattern of commuting flows in each LMA is cohesive, where the ideal would be that it forms a single 'cluster' of interacting areas. These requirements need to be balanced against each other in practice when defining LMAs because the first requirement, a high level of self-containment, creates a "danger of seeking external perfection at the expense of the essentially local character of the market area" (Goodman 1970, p. 185, original emphasis). This 'danger' has grown strongly with increased longer-distance commuting reducing the separateness of neighbouring LMAs. In fact the commuting patterns in 1960s Britain already made 'perfect' LMA self-containment unlikely due to the "substantial overlap between some labour market areas" whilst the other "criterion of an internally-unified structure is also difficult to satisfy" (Johnson et al. 1974, p.35). Although it may be theoretically interesting to define overlapping LMAs, such definitions would not be relevant in this paper with its focus on LMAs defined for policy: very few policy geographies include overlapping areas because this could lead to conflicting policies being applied there.

In a review of methods to define LMAs by grouping basic territorial units (TUs), EUROSTAT and Coombes (1992) identified principles for policy-relevant LMA definitions and emphasised the Goodman (1970) principle of LMAs internalising most flows – for which they used the term *autonomy* – but excluded his other criterion, that of internal *cohesion*. The review focus was on real-world policy applications and genuinely internally cohesive LMAs cannot be defined where commuting patterns are very diffuse. The aim of this paper is to find quantitative indicators covering *all* key aspects of LMAs, so it seeks evaluation metrics of both these key criteria, while recognizing that in some applications the measures will be of the relative failure to define cohesive LMAs.

There are additional requirements for LMA definitions which are relevant in certain policy applications. LMA definitions may be preferred if they restrain the size difference between the largest LMA and all others, something that could be particularly relevant in countries whose largest city is very dominant in terms of size: EUROSTAT and Coombes (1992) termed this *homogeneity* (ie. limiting the LMA size range in a

territory). Other research on LMAs provides a range of possible additional LMA evaluation criteria. Persyn and Torfs (2010) and also Landré and Håkansson (2013) emphasise the **balance** between labour demand and supply in LMAs; this issue also relates to policy concerns about spatial mismatches between where people seeking work live and where there is available work (Houston 2005).

It is important here to note some potential forms of LMA evaluation which are *not* relevant to this paper. All four evaluation criteria specified above focus on the interaction data patterns analysed to define LMAs, whereas the analyses of Farmer and Fotheringham (2012) assess the effect of *changing* that dataset. Another approach that is not relevant here is that of Papps and Newell (2002), where it is the definition procedure which is varied. In such approaches the stability of a set of results is measured, but they do not generate quantitative evaluation criteria of the type sought here. Some studies compare the results from using different regionalisations for the types of analyses which frequently use LMAs – examples include income analyses (e.g. Johnson 1995; Barkley et al. 1995), and multiple variable analyses (Cörvers et al. 2009) – while others like Baumann et al. (1983) and Maza and Villaverde (2011) compare the results from analyses using FRs rather than administrative areas. Findings from these studies depend on the scale of the areas analysed along with the statistical techniques and data used. The aims of this paper require its evaluation indicators to only use the data used when sets of LMAs were defined by grouping TUs: commuting flows and population counts. This review has identified four potentially quantifiable LMA evaluation criteria.

(1) **autonomy**: maximise the LMAs' internalisation of commuting flows
(2) **homogeneity**: minimise the LMAs' size range
(3) **balance**: equalise the size of LMAs' labour demand and supply
(4) **cohesion**: maximise the interaction intensity within LMAs

Potential Quantifiable Indicators to Evaluate Sets of LMAs

This section of the paper specifies potential quantifiable indicators for each of the four LMA evaluation indicators identified above, and uses a consistent notation. TUs, the areas such as municipalities for which the relevant datasets are available, are denoted i, j and number N in total. A functional regionalisation – a set of FRs such as LMAs – covering a territory without overlaps is a partition P of that territory, and within it M, X and Y refer to individual FRs (each of which is a single TU or group of TUs). The commuting flow matrix comprises inter-TU flows such as T_{ij} (ie. the number of workers living in i and working in j); intra-TU flows T_{ii} are included on the same basis. Such matrices are analysed when defining LMAs and so also provide the relevant data input for the evaluator indicators discussed below.

As a generalisation, $T_{XY} = \sum_{i \in X} \sum_{j \in Y} T_{ij}$ is the number of residents in X working in Y, $O_i = \sum_j T_{ij}$ is the total number of working residents in i, and thus $D_i = \sum_j T_{ji}$ is the total number of jobs in i. Similarly, $O_X = \sum_{i \in X} \sum_j T_{ij}$ and $D_X = \sum_{i \in X} \sum_j T_{ji}$ are respectively the total number of working residents and jobs in X.

Autonomy

Measuring autonomy highlights the commuting flows whose origin and destination are both within the boundary of a single FR. The count of this internal flow is divisible either by the working population of the FR (to measure supply-side self-containment) or local job numbers (demand-side self-containment). The measure of self-containment for FR M is then the lower of the supply-side O_M and the demand-side D_M values as in Coombes et al. (1986):

$$A_M = \frac{T_{MM}}{max(O_M, D_M)} \tag{1}$$

Autonomy measures for the whole regionalisation are then derived from these values for its constituent FRs: the median of FR self-containment values (A1) is one possible evaluation indicator, as is the minimum FR self-containment (A2).

An alternative overall value, termed global self-containment (A3), is the percentage of all commuting flows in the territory which cross no FR boundary in the regionalisation:

$$A3(P) = \frac{\sum_{M \in P} T_{MM}}{\sum_{M \in P} O_M} * 100 \tag{2}$$

Several recent approaches to identifying separable interaction clusters – the basic notion underlying the criterion of autonomy – use the measurement 'modularity quality' created by Newman and Girvan (2004) for community detection within social networks. Fortunato and Barthelemy (2007) and also Lancichinetti and Fortunato (2011) found that FR definition methods that maximise modularity quality Q tend not to identify either very large or very small communities. More recently however, Kropp and Schwengler (2016) asserted that maximising Q is preferable to maximising self-containment, and so Q will here be assessed as a potential candidate evaluation indicator.

In the present context, AQ is calculated by comparing the proportion of all commuting flows that are within FRs against the hypothetical value of that proportion if the distribution of flows was uniform. Thus values of AQ higher than 0 indicate higher than expected modularity. For a weighted directed matrix such as commuting flow data, Leicht and Newman (2008) formulate the index:

$$AQ(P) = \sum_{M \in P} \left(\frac{T_{MM}}{O_P} - \frac{O_M D_M}{O_P^2} \right) \tag{3}$$

It is noteworthy that in the regionalisation context, an alternative way of expressing this same measure is:

$$AQ(P) = \sum_{M \in P} Z(M, M) \tag{4}$$

where Z is the interaction index developed for regionalisation analyses many years ago by Hirst (1977):

$$Z(X,Y) = Z(Y,X) = \frac{1}{O_P}\left(T_{XY} - \frac{O_X D_Y}{O_P} + T_{YX} - \frac{O_Y D_X}{O_P}\right) \qquad (5)$$

Homogeneity

For policy users of LMAs, the preference for a more homogeneous LMA size range has often centred on whether the largest FR is very much larger than all the others. Pereira (1997) responded to this concern by setting a maximum land area size for FRs, while for Trutzel and Brandmüller (2017) population size differences are the key concern. How far any regionalisation includes very large FRs is measurable by the size of the FR at the ninth decile of the FR size range (ie. the size of the FR which is larger than 90% of all FRs). Rather more holistic measures of size homogeneity H are derivable from the Gini coefficient that measures inequality across the size distribution. By subtracting the Gini coefficient from 1, the measure is higher for regionalisations with more equally sized FRs. Applying this measure to the population distribution of FRs produces indicator HO1, whereas the indicator using the ninth decile measure is HO2. Equivalent measures of FR size variation in terms of land area, HL1 and also HL2, respond to the concerns about accessibility that prompted Eckey et al. (2007) to set an internal FR travel time limit, and the limit on land area size set by Soares et al. (2017).

Balance

The balance between labour supply and demand in FRs is of interest to policy-makers because balanced labour markets are associated with a lower prevalence of long-distance commuting and its attendant disadvantages (Melo et al. 2012). The labour market balance (B) of an individual FR M is shown by the ratio between the number of jobs at local workplaces and its number of employed residents, which is termed the job ratio:

$$B_M = \frac{O_M}{D_M} \qquad (6)$$

Regionalisations consisting of balanced labour markets mostly have FRs with job ratios close to the value of 1, so a set of FRs is preferred if all its FR job ratios are close to unity. As a result, an appropriate indicator B1 for a regionalisation is derivable as 1 minus the Gini coefficient of the job ratios of its FRs. A simpler alternative indicator B2 is obtained as the ninth decile FR job ratio, which will highlight those FRs with most in-commuters and thus the largest cities in most cases.

Cohesion

The concern of Goodman (1970) that larger FRs may lack cohesion (C) stems from the probability that a large conurbation, for example, includes many TUs with strong

interactions with only a few of the other TUs in that FR. A regionalisation with fewer FRs will have larger FRs on average, so the number of FRs (CN) is a crude proxy cohesion indicator of cohesion. It is only a crude proxy because it is not always true that regionalisations with fewer FRs have lower cohesion levels: for example in the early stages of agglomeration procedures, TU groupings often internalise large inter-TU flows and so are creating FRs with higher cohesion levels.

Cohesion is directly indicated by the proportion of inter-TU commuters which are internal to FRs. ISTAT (2015) suggest as global indicator the percentage of inter-TU commuters that do not cross the boundaries of the FR to which they are assigned, but the highest value for this would be the case of a single FR covering the whole territory and that is not maximum cohesion. This problem is addressed by more complex approaches using an interaction index. A widely used interaction index (S) is attributed to Smart (1974), although in fact Ball (1980) clarifies that it was subsequently improved before its use in the computerised definition of the UK's official definitions of LMAs (Coombes et al. 1986):

$$S(X, Y) = \frac{T_{X,Y}^2}{O_X D_Y} + \frac{T_{Y,X}^2}{O_Y D_X} \tag{7}$$

Flórez-Revuelta et al. (2008) define a global index CS which sums the Smart interaction index between each TU and the rest of its FR:

$$CS = \sum_{X \in P} \sum_{i \in X} S(i, X \setminus i) \tag{8}$$

Casado-Díaz et al. (2017) found that maximising Eq. 8 can favour regionalisations with misallocated TUs. This potential problem is avoided in a revised version (R) of the Smart interaction index in which the magnitude it measures is transformed back to the scale of ratios (given that the original index is in fact the sum of two products of ratios):

$$CR = \sum_{X \in P} \sum_{i \in X} \sqrt{\frac{S(i, X \setminus i)}{2}} \tag{9}$$

It is inevitable that the value of CS, or of CR, for a single-TU FR will be zero, because there is no "rest or its FR" for the single TU to interact with. This zero input to the regionalisation's overall index creates a potential bias towards regionalisations without single-TU FRs. In practice many single-TU FRs remain isolated due to their low interaction with other FRs, so other regionalisations which group them into larger FRs may well have lower scores on these indices due to the effect on the interaction index of increased denominator size resulting from a merger.

Proposed Set of Evaluation Indicators

The 13 indicators emerging from the preceding discussion can now be listed, as in Table 1. Whilst in many applications the key criteria are likely to be autonomy and cohesion, the list also includes candidate quantifiable indicators for the two other criteria identified in the review of the labour market area concept. The indicators in

Table 1 need to be exemplified to assess their value in practice, and this is done in the next section of the paper.

Exemplification of Potential Indicators

This section of the paper empirically assesses the 13 candidate LMA evaluation indicators. These analyses are applied to a very large number of functional regionalisations which have been produced by applying 2 established methods to commuting data on the USA. Both methods are agglomerative so the results can be ordered by the count of FRs in each separate regionalisation. The charts below use as their horizontal axes the number of FRs in the regionalisation, and it will be remembered that this metric also acts as indicator CN. On the far left of each chart every TU is a separate FR, while on the far right a single FR includes all TUs. Candidate indicators are presented as curves which reveal how their values vary through the aggregation process. It is important to state here that these indicators are potentially applicable to any regionalisation, whether or not it was produced by an agglomerative method. The purpose of these analyses is not to evaluate the two regionalisation methods used, although the indicators could be valuable for that purpose (for example there might be a strong difference in a homogeneity indicator for the two regionalisations with the same number of FRs produced by the different methods).

The purpose of the empirical analyses is to identify the indicators which could help policy-makers select preferred sets of FRs from a range of alternatives. There are several ways in which an indicator can be seen to be useful for that purpose. One way could indeed be for the indicator to show that its value is consistently higher for the regionalisations from one method rather than the other. Another way is for the values of the indicator to show a clear 'peak' at a certain number of FRs: this would show a policy-maker that in terms of the criterion the indicator measures (eg. autonomy), the

Table 1 Indicators to be evaluated

Criterion	Indicator	
Autonomy	A1	median self-containment
	A2	minimum self-containment
	A3	global self-containment
	AQ	global modularity
Homogeneity	HO1	1 minus the Gini coefficient of working population size
	HO2	9th decile working population size
	HL1	1 minus the Gini coefficient of land area size
	HL2	9th decile land area size
Balance	B1	1 minus the Gini coefficient of job ratios
	B2	9th decile job ratio
Cohesion	CR	global interaction index R
	CS	global interaction index S
	CN	number of TU/FRs (a crude proxy: the horizontal axis on all charts)

set of FRs with the peak value is preferable to regionalisations produced by the same method with either more or fewer FRs.

Another way that an indicator can appear to be valuable would be if its values show a distinct 'step-change' in values at a point in the agglomerative process. This feature may be the least probable of the three, because results from aggregative methods usually change marginally step-by-step. Smooth trends to the indicator values are to be expected, so where any step-change is observable it will be highlighted. Evidence from other LMA definitions in the USA such as Fitzsimmons and Ratcliffe (2004) and the more recent Fowler et al. (2016) suggests that the results of most interest for policy purposes will be for regionalisations in the range of 1500–100 FRs.

LMAs Defined to Exemplify the Indicators

Census 2000 commuting flows[1] for the 49 continental states of the USA provide the analysis database, not only for both the regionalisation analyses but also for the subsequent indicator analyses. There are two different data resolutions available, here the dataset used has Counties[2] as the TUs. Table 2 reveals the substantial variations in key characteristics of the 3141 Counties: the significance of these variations is that at earlier stages of agglomeration many of the larger TUs may persist as single-TU FRs.

One regionalisation method used within academic analyses (eg. Landré 2012) is the Intramax procedure of Masser and Scheurwater (1980), a clustering algorithm implementable with FlowMap (De Jong and van der Vaart 2010). Intramax is a rigidly hierarchical process, iteratively selecting areas to merge based on their level of inter-action. Intramax does not produce a single regionalisation: it generates one set of FRs for each number between 1 and the total number of TUs, and so is usually followed by the selection of a single 'final' regionalisation based on some quality metric (as with the USA's Commuting Zones defined by Killian et al. 1993). This final step was not needed for this study because the aim was to output 3000+ regionalisations and for each of these to include a different number of FRs (thus covering the range from each TU being a single-TU FR, down to just one single FR including all TUs). The application of Intramax here used the Hirst (1977) interaction index.

Another set of 3000(+) regionalisations was defined[3] using the TTWA method from Coombes and Bond (2008). This method is non-hierarchical and designed to end when all FRs meet set criteria, but to provide comparable results to those from Intramax it was necessary to set criteria so that the procedure continued until every TU is in a single FR, reporting the 'state-of-play' after each step in the agglomeration process. Unlike in the TTWA method normally, the application here had no 'trade-off' between FR working population size and autonomy, measured by self-containment. It was necessary for the method to continue until all TUs were in a single FR and so no

---

[1] The commuting data for U.S., from the U.S. 2000 Census, is publicly available at U.S. Census Bureau web site: http://www.census.gov/population/www/cen2000/commuting/), as are the area figures (http://www.census.gov/geo/maps-data/data/gazetteer2000.html).

[2] In the alternative dataset the counties in the 6 states of New England are replaced by the smaller minor civil divisions; all the basic results reported here are robust to changing to that dataset (results obtainable from the authors on request).

[3] The two regionalisation methods implemented here ran in similarly short time on modern high power computing facilities.

Table 2 Characteristics of TUs (USA 2000)

	Mean	Median	Maximum	Minimum	*National*
Working population (,000)	40.81	10.52	3854.11	0.04	*128,169*
Land area (square kms)	3129	1687	382,910	5.10	*9,818,050*
Self-containment (%)	65.00	67.18	98.98	8.64	*100.00*
Job ratio	0.90	0.90	3.92	0.20	*1.00*

self-containment 'cap' was applied (nb. the same change was not made to the size measure). This produced the required 3000(+) sets of FRs for the evaluation of candidate indicators (nb. most of these 3000(+) regionalisations could not have been output by the standard TTWA method due to them including FRs that fail the set criteria for TTWAs).

The potential relevance of the indicators evaluated below is not limited to LMAs defined by these two sample methods: they could be applied to any set of non-overlapping FRs. Earlier it was noted that, because cohesion is expected to decline as the aggregation process creates larger FRs, the number of FRs in a regionalisation may provide a crude proxy indicator of cohesion. In every chart below the number of FRs provides the X axis and so there is no need for a separate chart of the FR count as proxy cohesion indicator CN.

Exemplifying Indicators of Autonomy

The values of any autonomy indicator are expected to rise as agglomeration proceeds because there are fewer boundaries for commuters to cross. Figure 1 shows the values
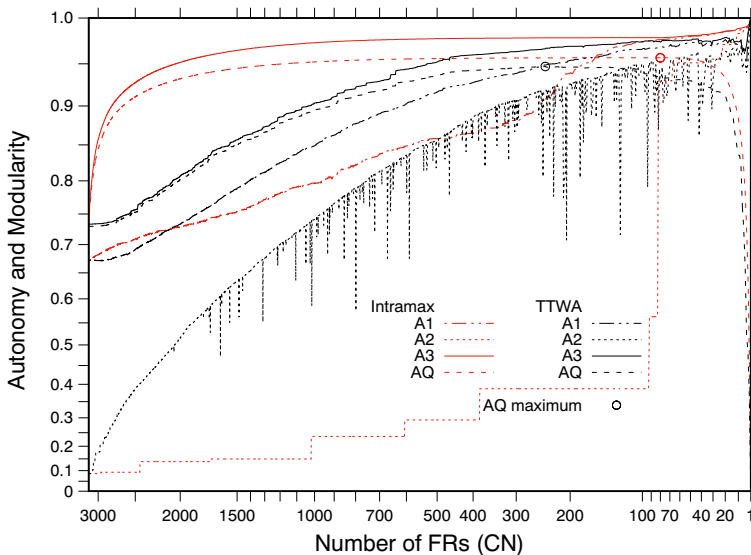
Fig. 1 Candidate indicators of autonomy

for each candidate indicator of autonomy, including the related modularity indicator, for each of the 3000+ regionalisations from the 2 alternative methods.

Figure 1 shows the values of all candidate indicators initially increasing as aggregation progresses with either method. As stated earlier, an indicator can prove valuable *either* by revealing strongly contrasting values for the regionalisations from the two methods *or* by finding either a notable peak or step-change in the values for regionalisations with differing numbers of FRs produced by the same method. Figure 1 reveals no peak or step-change in either set of global self-containment A3 values: this is unsurprising, because the individual steps in agglomerative processes only marginally change a set of FRs, and thus only slightly alter 'global' statistical indicators. As a result, the paper will from this point only mention a step-change or notable peak which *is* evident in results for an indicator, with their expected absence being left as the unstated default.

The major difference in global self-containment A3 values for the FRs from the two methods is an important feature of this indicator. This difference stems from the contrasting focus of the methods in their early aggregations: Intramax tends to prioritise very populous TUs/FRs and so its initial groupings internalise large flows which strongly increases global self-containment, whereas the TTWA method prioritises grouping TUs/FRs with very low self-containment which tend to be small and so grouping them has less impact on global self-containment.

There is an even greater contrast between the values for FRs from the 2 methods for the minimum self-containment indicator A2, despite the sporadic appearance of much lower A2 values for FRs from the TTWA method (nb. these are due to the modification made here to force outputs after each iteration of the algorithm which over-rules the normal restriction on potential outputs by the method). The A2 indicator shows that in the range of 1500–100 FRs of most interest there remain some Intramax FRs with very low self-containment values, a major difference to the equivalent TTWA FRs (Fig. 1). Differences between the values for the FRs from the 2 methods are also notable for median self-containment indicator A1: in this case they are higher for the Intramax FRs in the earlier stages of aggregation but lower in later stages. Thus both these indicators offer useful information above and beyond that from the global self-containment indicator A3.

Figure 1 also shows a distinct difference in the modularity values AQ for the FRs from the two methods. Both curves have a 'peak' but for the TTWA results this is indistinct, and for Intramax it is outwith the critical range of 1500–100 FRs. Most notably though, Fig. 1 reveals that through most of the aggregation process, and particularly across the critical range of 1500–100 FRs, the modularity curves follow extraordinarily similar trends to those of global self-containment A3. This suggests that any LMA evaluation process only needs one of the two indicators, and global self-containment is by far the more readily understood. The modularity curves also have the notable and counter-intuitive feature of declining to zero at the end of the aggregation process: this is because the autonomy value of a single-FR regionalisation must be 100% which means that the 'observed minus expected' basis of the modularity measure produces a value of zero.

Exemplifying Indicators of Homogeneity

Figure 2 shows curves for the two candidate indicators of FR population size homogeneity: the ninth decile value HO2 and the more comprehensive indicator HO1 which
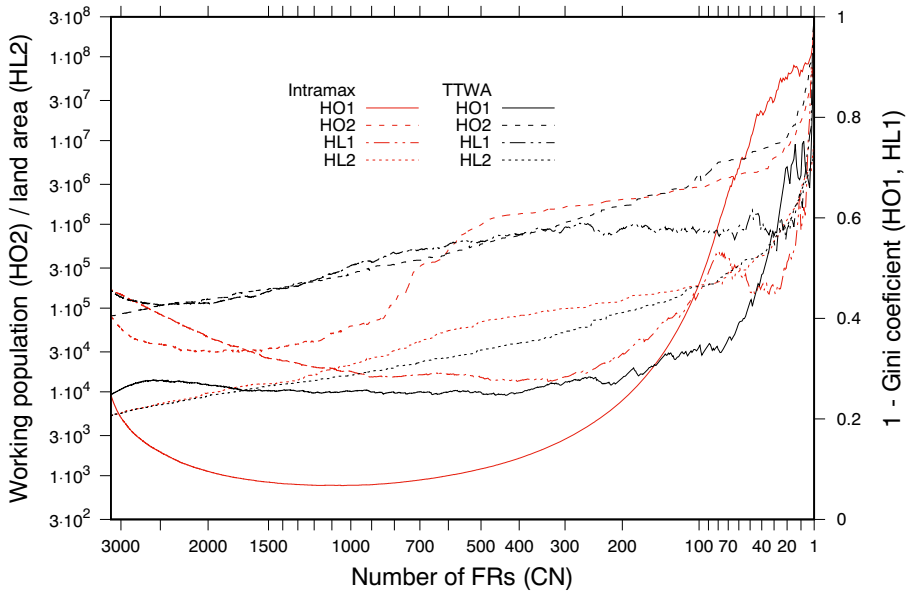
Fig. 2 Candidate indicators of homogeneity

has the Gini coefficient as its basis. Both indicators show contrasting values for the regionalisations from the 2 methods, mainly due to the tendency of the Intramax method to initially group TUs with large populations and so produce far more unequally sized FRs than those the TTWA method defines. One consequence of Intramax initially grouping more populous TUs is that the small TUs become a larger proportion of the FR size distribution so that the indicators decline as the largest FRs grow larger. The decile-based HO2 indicator for the Intramax results shows a trend with 2 step-changes with one an upward deflection and the other downward: neither is readily interpretable, suggesting that indicator HO1 is the more reliable measure of homogeneity.

Figure 2 also shows equivalent indicators (HL1, HL2) of land area size homogeneity. Indicator HL2 based on the Gini coefficient shows the more strongly contrasting values for the regionalisations from the 2 methods. The ninth decile indicator HL1 offers little additional intelligence beyond that provided by HL2.

Exemplifying Indicators of Balance

Figure 3 shows the two candidate indicators based on FR job ratios. In most actual regionalisations the median job ratio will be below 1, because there are usually several commuter 'exporting' TU/FRs for each job surplus TU/FR. The ninth decile indicator B1 highlights job importing FRs because they usually include the largest cities. Yet it is the B2 indicator based on the Gini coefficient which more consistently shows the difference between the 2 sets of results that is due to the Intramax agglomeration needing to proceed much further before creating the preferred outcome of mostly balanced labour markets. This empirical
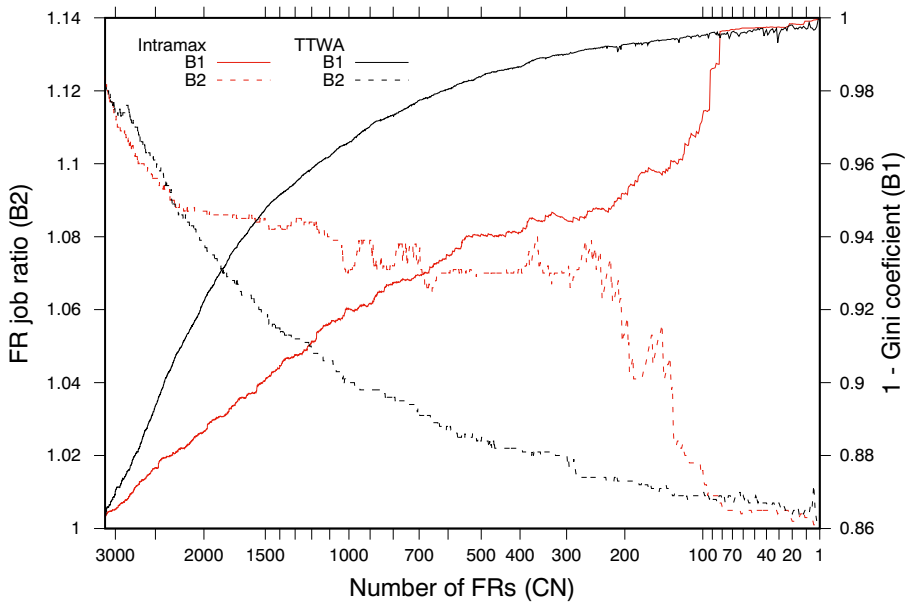
Fig. 3 Candidate indicators of balance

advantage of the B2 indicator is reinforced by the fact that its basis in a Gini coefficient makes it a far more holistic measure.

Exemplifying Indicators of Cohesion

Both versions of cohesion indicator based on an interaction index show the expected initial increase and subsequent decline of cohesion through the agglomeration processes of both FR definition methods (Fig. 4). With both indicators, Intramax regionalisation values also show a step-change, but these occur very close to the end of the agglomeration process. The indicators have similar results when applied to the TTWA regionalisations but the two Intramax curves are notably different to each other: CR rather counter-intuitively suggests that the level of cohesion barely changes across the range of 1500–300 FRs that is probably of the most interest. This evidence in favour of indicator CS runs contrary to that in a different context where *CR* was found superior (Casado-Díaz et al. 2017).

As mentioned earlier, there is a methodological issue with these indicators when applied to single-TU FRs. Consequently their greatest value is in analysing regionalisations with a high average number of TUs in each FR, such as partitions of the USA in which the 3000+ TUs have been grouped into 1500 or fewer FRs. It was also mentioned earlier that using the number of FRs as crude proxy measure of cohesion CN is dubious in the earlier stages of agglomeration because many early TU groupings increase cohesion levels, whereas further groupings of FRs which are already large will tend to lower their cohesion. Thus no measure of cohesion is entirely reliable for regionalisations in which many of the TUs were large and self-contained enough to constitute single-TU FRs.
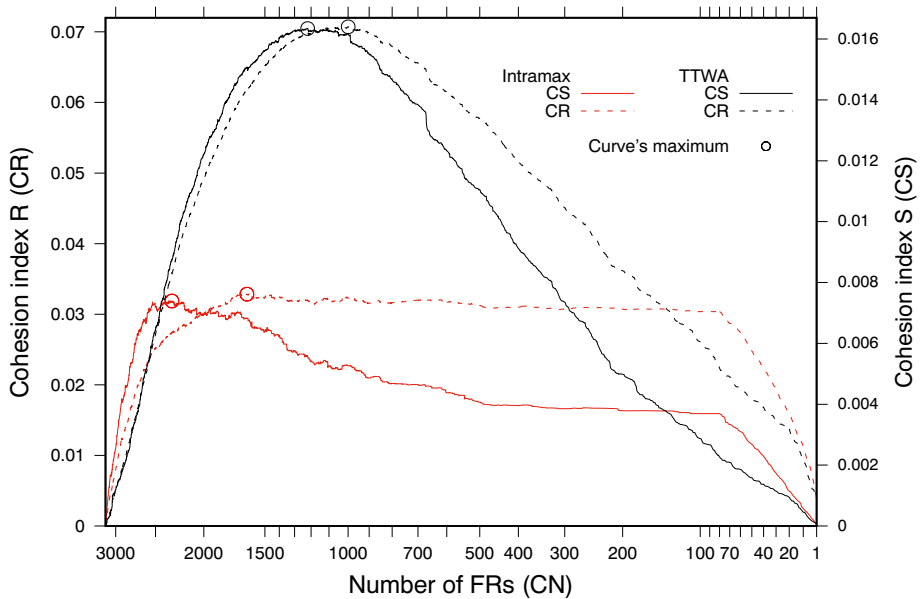
Fig. 4 Candidate indicators of cohesion

Summary of Findings from the Exemplifications

The exemplifications found informative indicators for each of the evaluation criteria. All of the candidate indicators of the key criterion autonomy were found to be informative, but because the modularity indicator AQ closely duplicates the more interpretable global self-containment indicator A3, the modularity indicator is considered superfluous. The other key criterion for Goodman (1970) was cohesion, and here the exemplifications favoured the indicator CS, although its values may be less reliable for regionalisations with numerous single-TU FRs. For the other two criteria, labour market balance and size homogeneity, indicators based upon the Gini coefficient proved the most informative.

Table 3 Summary of the preferred indicators

Criterion	Indicator		Usage
Autonomy	A2	Minimum self-containment	Threshold
	A1	Median self-containment	Maximise
	A3	Global self-containment	Maximise
Homogeneity	HO1	1 minus Gini coefficient of working population size	Maximise
	HL1	1 minus Gini coefficient of land area size	Maximise
Balance	B1	1 minus Gini coefficient of job ratios	Maximise
Cohesion	CS*	Global interaction index S	Maximise
	CN	(Number of TU/FRs: a crude proxy)	(Maximise)

*measure may be less reliable for regionalisations with numerous single-TU FRs

Table 3 lists the 8 indicators identified here as the most appropriate way to evaluate regionalisations against the 4 key criteria, and shows in its right-hand column how each indicator would be implemented. Most of the indicators are readily interpreted, with higher value regionalisations preferred to those with lower values. Indicator A2 is different because it provides a binary 'test' for regionalisations, rejecting any set of FRs in which one or more has a self-containment value below a set threshold. (In some contexts the priority is not the avoidance of small FRs but prevention of very large ones: in such cases the set 'threshold' size would fail all regionalisations which include any FR above that size, although if this value is set too low then no set of FRs will meet the requirement in a territory including a major metropolitan area.)

Next Steps

This paper responds to the widespread use of LMAs and other FRs in policy implementation and research, as with the renewed HM Government (2018) requirement that the boundaries of agencies charged with sub-national economic development should be better aligned with "functional economic areas that are conducive towards the development of strategy, policy and interventions" (p.7). There may be 'real world' consequences if policies are misdirected due to using inappropriately defined boundaries, because these can distort policy analyses (Grasland and Madelin 2006). The basic requirements for European statistical regions include the principle that "[o]bjective criteria for the definition of regions are necessary" (European Parliament and Council of the European Union 2003, preamble statement 7). Adopting that principle here has led the paper to identify objective quantifiable indicators, interpreting appropriate criteria, to evaluate alternative sets of FRs. The next step is to outline how the indicators could be used to support decisions by policy-makers.

Applying the Indicators to Evaluate Regionalisations

The paper derived from the concept of the LMA four potentially measurable evaluation criteria for regionalisations – autonomy, cohesion, homogeneity and balance – but for any particular policy it is likely that only one or two of these will be critically important. Thus it will be necessary to review the objectives of the policy concerned to assess the relevance of the four evaluation criteria. If only one criterion is important then implementation will be straightforward unless – as with autonomy (Table 3) – there are several potentially valuable indicators of that single criterion. If more than one indicator is expected to contribute valuable information to the evaluation, the next step is to find a framework within which the indicators can be combined. Evaluations will have the 'transparency' needed in policy contexts if this is a simple framework, rather than a 'black box' procedure combining a large number of indicators within a multi-criteria synthetic index (Nardo et al. 2008). Transparency will of course also be facilitated by using as few indicators as possible.

A hypothetical example can illustrate how a policy-relevant evaluation could use selected indicators. Sandford (2019) documents how English housing policy has since 2006 required all local authorities to plan for housing within boundaries[4] of housing

---

[4] New policy changes somewhat shifted "the focus away from housing market areas" as a set of defined FRs (Department for Communities and Local Government 2017, p.23).

market areas (HMAs), but how this particular form of FR boundary should be defined was not prescribed. Research funded by a government agency used advanced regionalisation methods to define several sets of HMAs, including the Strategic HMAs within a two-tier set of areas, and also a set of Single-tier HMAs (Jones et al. 2010). Sandford (2019, p.16) uses the term "closure" (self-containment) when identifying autonomy as the key criterion for HMAs, with this criterion relevant to both migration and commuting flows. Jones et al. (2010) set different minimum levels of migration and commuting self-containment when defining the above two sets of HMAs, so the minimum self-containment A2 indicator is of potential relevance here. Of even more value perhaps is the median autonomy indicator A1 because this is a holistic evaluation of the autonomy levels in each set of FRs. Table 4 reports these values for the above two sets of HMAs.

Values for the indicator A2 (Table 4) are largely the result of the different minimum self-containment thresholds which were set for the two regionalisations. The values for A1, being medians for the whole set of FRs, are less determined by this factor and hence are more informative. In terms of migration the Strategic HMAs actually have the slightly higher A2 value despite having had a lower minimum threshold. On the basis of these autonomy indicators alone, the only circumstance in which a policy-maker could prefer the Single-tier HMAs would be if they had an over-riding concern that every HMA should have a migration self-containment value well above 50%. Table 4 in fact also provides the key reason for the Strategic HMAs generally having higher autonomy values than those of the Single-tier HMAs: there are nearly 20% more of the latter than the former, so the Strategic HMAs have fewer boundaries for commuting or migration flows to cross. Recalling now that the number of FRs was cited here as a proxy indicator of cohesion (CN), these indicator values could tip the balance in favour of the Single-tier HMAs for any policy-maker whose priorities not only led to a preference for higher FR autonomy values but also to the seeking the extra local detail and probable higher cohesion level which a more numerous set of FRs could provide.

In the situation where more than one indicator is relevant, and they suggest that different regionalisations are preferable, transparency in the decision-making framework can be aided by visualisation of the 'trade-off' between the indicators. Referring back to the USA exemplifications presented earlier, across the crucial range of 1500–300 FRs it was seen that the autonomy indicator A3 had rising values (Fig. 1), whereas the cohesion indicator CS had declining values (Fig. 4). An assessment of the sets of FRs produced by the TTWA algorithm could favour a regionalisation of about 500 FRs because at this point in the aggregation the CS value has yet to decline far from its peak and, although the values of A3

Table 4 Appropriate evaluation indicators of alternative sets of English HMAs

	Commuting flows		Migration flows		Number [CN]
	A2 minimum self-containment	A1 median self-containment	A2 minimum self-containment	A1 median self-containment	
Strategic HMAs	77.51	85.01	52.40	67.66	105
Single-tier HMAs	72.58	82.06	56.57	67.38	124

are still rising, they are levelling off. Such a commentary on the visualisable evidence of the trade-off could be presented to the policy-maker, who might then make the final selection from a number of broadly similar sets of FRs after seeing their boundaries in a part of the territory which is particularly 'sensitive' for their policy context.

Concluding Discussion

In this paper four potentially measurable evaluation criteria for functional regionalisations have been derived from the concept of labour market areas, the form of FR most widely used in policy. Candidate indicators were identified for each criterion, and these were then exemplified by applying them to over 6000 regionalisations of the USA produced by two LMA definition methods, providing innovative information. New insights gained included the fact that a modularity Q indicator adds no useful information that is not available from the more interpretable measure of global self-containment. The results from the two methods proved different enough to help assess which indicators were informative. Table 3 lists the preferred indicators emerging from this assessment. The paper finally outlined ways to use the indicators to enable policy-makers to identify sets of FRs that would be more appropriate for a specific policy. It is argued that with the analytical support of spatial analysis professionals, policy-makers could readily understand how the indicators reflect their policy concerns and then choose the appropriate regionalisation, perhaps after a 'trade-off' between different priorities aided by visualisation of graphics and selected boundaries.

Possible future research building on this paper includes testing the identified indicators in territories that are very different to the USA. The indicators also have the potential to advance the long-standing debates over regionalisation methods: by applying them to regionalisations produced by different methods in contrasting territories, they could show which method produced FRs with higher levels of autonomy, for example. Some of the indicators might also provide better optimisation metrics than those that currently 'drive' certain methods, including those in Fusco and Caglioni (2011) or in Martínez-Bernabeu et al. (2012). Yet the most valuable next step may be simply to apply the approach described here to more 'real-world' cases of policy needing a set of FRs, thereby demonstrating empirically the benefit of quantitative indicators for decision support.

References

Adams, J., Van Drasek, B., & Phillips, E. (1999). Metropolitan area definition in the United States. *Urban Geography, 20*, 695–726.

Ball, R. (1980). The use and definition of travel-to-work areas in Great Britain: Some problems. *Regional Studies, 14*, 125–139.

Barkley, D., Henry, M., Bao, S., & Brooks, K. (1995). How functional are economic areas? Tests for intra-regional spatial association using spatial data analysis. *Papers in Regional Science, 74*, 297–316.

Baumann, J., Fischer, M., & Schubert, U. (1983). A multiregional labour supply model for Austria: The effects of different regionalisations in multiregional labour market modelling. *Papers in Regional Science, 52*, 53–83.

Brandmueller, T., Schäfer, G., Ekkehard, P., Müller, O., & Angelova-Tosheva, V. (2017). Territorial indicators for policy purposes: NUTS regions and beyond. *Regional Statistics, 7*, 78–89.

Brown, L., & Holmes, J. (1971). The delimitation of functional regions, nodal regions and hierarchies by functional distance approaches. *Journal of Regional Science, 11*, 57–72.

Casado-Díaz, J., Martínez-Bernabeu, L., & Rowe, F. (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.

Cheshire, P., & Magrini, S. (2008). Urban growth drivers in a Europe of sticky people and implicit boundaries. *Journal of Economic Geography, 9*, 85–115.

Coombes, M., & Bond, S. (2008). *Travel-to-work areas: The 2007 review*. London: Office for National Statistics.

Coombes, M., Green, A., & Openshaw, S. (1986). An efficient algorithm to generate official statistical reporting areas: The case of the 1984 travel-to-work areas revision in Britain. *Journal of the Operational Research Society, 37*, 943–953.

Coombes, M., Casado-Díaz, J., Martínez-Bernabeu, L., & Carausu, F. (2012). *Study on comparable labour market areas*. Luxembourg: Eurostat.

Cörvers, F., Hensen, M., & Bongaerts, D. (2009). Delimitation and coherence of functional and administrative regions. *Regional Studies, 43*, 19–31.

De Jong, T., & van der Vaart, N. (2010). *Manual Flowmap*. Utrecht: Utrecht University.

Department for Communities and Local Government. (2017). *Planning for the right homes in the right places: Consultation proposals*. London: Department for Communities and Local Government.

Eckey, H., Schwengler, B., & Türck, M. (2007). *Vergleich von Deutschen Arbeitsmarktregionen*. Nuremburg: Institut für Arbeitsmarkt und Berufsforschung.

European Parliament, & Council of the European Union. (2003). On the establishment of a common classification of territorial units for statistics. *Official Journal of the European Union, Regulation*. EC) 1059/2003.

EUROSTAT. (2017). *European regional and urban statistics reference guide*. Luxembourg: European Union.

EUROSTAT, & Coombes. M. (1992). *Étude sur les Zones d'Emploi*. Luxembourg: Office for Official Publications of the European Communities.

Farmer, C., & Fotheringham, A. (2012). Network-based functional regions. *Environment and Planning A, 43*, 2723–2741.

Fitzsimmons, J., & Ratcliffe, M. (2004). Reflections on the review of metropolitan area standards in the United States, 1990-2000. In A. Champion & G. Hugo (Eds.), *New forms of urbanization: Beyond the urban-rural dichotomy* (pp. 307–324). Aldershot: Ashgate.

Flórez-Revuelta, F., Casado-Díaz, J., & Martínez-Bernabeu, L. (2008). An evolutionary approach to the delineation of functional areas based on travel-to-work flows. *International Journal of Automation and Computing, 5*(1), 10–21.

Fortunato, S., & Barthelemy, M. (2007). Resolution limit in community detection. *Proceedings of the National Academy of Sciences, 104*, 36–41.

Fowler, C., Rhubart, D., & Jensen, L. (2016). Reassessing and revising commuting zones for 2010: History, assessment, and updates for US 'labor-sheds' 1990-2010. *Population Research and Policy Review, 35*, 263–286.

Fusco, G., & Caglioni, M. (2011). Hierarchical clustering through spatial interaction data: The case of commuting flows in South-Eastern France. In B. Murgante, O. Gervasi, A. Iglesias, D. Taniar, & B. Apduhan (Eds.), *Computational science and its applications - ICCSA 2011* (pp. 135–151). Berlin: Springer.

Goodman, J. (1970). The definition and analysis of local labour markets: Some empirical problems. *British Journal of Industrial Relations, 8*, 179–196.

Grasland, C., & Madelin, M., Eds. (2006). The modifiable areas unit problem. Luxembourg: ESPON (Final Report 3.4.3).

Gutiérrez Posada, D., Rubiera Morollón, F., & Viñuela, A. (2017). The determinants of local employment growth in Spain. *Applied Spatial Analysis & Policy, 11*, 511–533.

Hirst, M. (1977). Hierarchical aggregation procedures for interaction data: A comment. *Environment and Planning A, 9*, 99–103.

HM Government. (2018). *Strengthened local economic partnerships*. London: Ministry of Housing, Communities and Local Government.

Houston, D. (2005). Employability, skills mismatch and spatial mismatch in metropolitan labour markets. *Urban Studies, 42*, 221–243.

ISTAT. (2015). *La Nuova Geografia del Sistema Locali*. Rome: Istituto Nazionale di Statistica.

Johnson, K. (1995). Redefinition of the BEA economic areas. *Survey of Current Business, 75*, 75–81.

Johnson, J., Salt, J., Wood, P., Flowerdew, R., Hadjifotiou, N., & Robinson, H. (1974). *Housing and the migration of labour in England and Wales*. Farnborough: Saxon House.

Jones, C., Coombes, M., & Wong, C. (2010). *Geography of housing market areas: Final report*. London: Communities and Local Government.

Killian, M., Tolbert, C., Singelmann, J., & Desaran, F. (1993). Mapping social and economic space: The delineation of local labor markets in the United States. In J. Singlemann & F. Deseran (Eds.), *Inequalities in labor market areas* (pp. 69–79). Boulder: Westview Press.

Kim, H., Chun, Y., & Kim, K. (2013). Delimitation of functional regions using a p-regions problem approach. *International Regional Science Review, 38*, 235–263.

Koo, H. (2012). Functional regionalization in the Seoul metropolitan area using improved hierarchical aggregation methods. *Journal of the Korean Cartographic Association, 12*, 25–35.

Kropp, P., & Schwengler, B. (2016). Three-step method for delineating functional labour market regions. *Regional Studies, 50*, 429–445.

Lancichinetti, A., & Fortunato, S. (2011). Limits of modularity maximization in community detection. *Physical Review E, 84*, 066122.

Landré, M. (2012). Geoprocessing journey-to-work data: Delineating commuting regions in Dalarna, Sweden. *International Journal of Geo-Information, 1*, 294–314.

Landré, M., & Håkansson, J. (2013). Rule versus interaction function: Evaluating regional aggregations of commuting flows in Sweden. *European Journal of Transport and Infrastructure Research, 13*, 1–19.

Leicht, E., & Newman, M. (2008). Community structure in directed networks. *Physical Review Letters, 100*, 118703.

Martínez-Bernabeu, L., Flórez-Revuelta, F., & Casado-Díaz, J. (2012). Grouping genetic operators for the delineation of functional areas based on spatial interaction. *Expert Systems with Applications, 39*, 6754–6766.

Masser, I., & Scheurwater, J. (1980). Functional regionalisation of spatial interaction data: An evaluation of some suggested strategies. *Environment and Planning A, 12*, 1357–1382.

Maza, A., & Villaverde, J. (2011). EU regional convergence and policy: Does the concept of region matter? *Journal of Policy Modeling, 33*, 889–900.

Melo, P., Graham, D., & Noland, R. (2012). The effect of labour market spatial structure on commuting in England and Wales). *Journal of Economic Geography, 12*, 717–737.

Miron, J. (2010). *The geography of competition: Firms, prices and localization*. New York: Springer.

Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffmann, A., & Giovannini, E. (2008). *Handbook on constructing composite indicators: Methodology and user guidance*. Paris: OECD.

Newman, M., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E, 69*, 026113.

Openshaw, S. (1984). *The modifiable areal unit problem (CATMOG 38)*. Norwich: Geobooks.

Papps, K., & Newell, J. (2002). *Identifying functional labour market areas in New Zealand: A reconnaissance study using travel-to-work data*. Bonn: Forschungsinstitut zur Zukunft der Arbeit (IZA).

Pereira, A. (1997). As bacias de emprego em portugal continental. *Revista de Estatística, 4*, 19–41.

Persyn, D., & Torfs, W. (2010). *Functional labor Markets in Belgium: Evolution over time*. Leuven: Katholieke Universiteit.

Sandford, M. (2019). *Where do you draw the line? Local administrative boundaries in England*. London: House of Commons Library.

Schubert, U., Gerking, S., Isserman, A., & Taylor, C. (1987). Regional labour market modelling: A state of the art review. In M. M. Fischer & P. Nijkamp (Eds.), *Regional labour markets: Analytical contributions and cross-National Comparisons* (pp. 53–94). Amsterdam: Elsevier.

Séguin, A.-M., Apparicio, P., & Riva, M. (2012). The impact of geographical scale in identifying areas as possible sites for area-based interventions to tackle poverty: The case of Montréal. *Applied Spatial Analysis & Policy, 5*, 231–251.

Smart, M. (1974). Labour market areas: Uses and definition. *Progress in Planning, 2*, 238–353.

Soares, E., Figueiredo, R., & Vala, F. (2017). Defining labour market areas and its relevance from a statistical perspective: The Portuguese case. *Statistical Journal of the IAOS, 33*, 615–625.

Stillwell, J., Daras, K., Bell, M., & Lomax, N. (2014). The IMAGE studio: A tool for internal migration analysis and modelling. *Applied Spatial Analysis & Policy, 7*, 5–23.

Trutzel, K., & Brandmüller, T. (2017). Addressing significant gaps in the information requested for evidence-based sub-national policy making: Using new methods and new data sources. *Statistical Journal of the IAOS, 33*, 581–584.

van der Laan, L., & Schalke, R. (2001). Reality versus policy: The delineation and testing of local labour market and spatial policy areas. *European Planning Studies, 9*, 201–221.

Veneri, P. (2013). *On city size distribution: evidence from OECD functional urban areas (regional development working paper 2013/27)*. Paris: OECD.

Publisher's Note   Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.