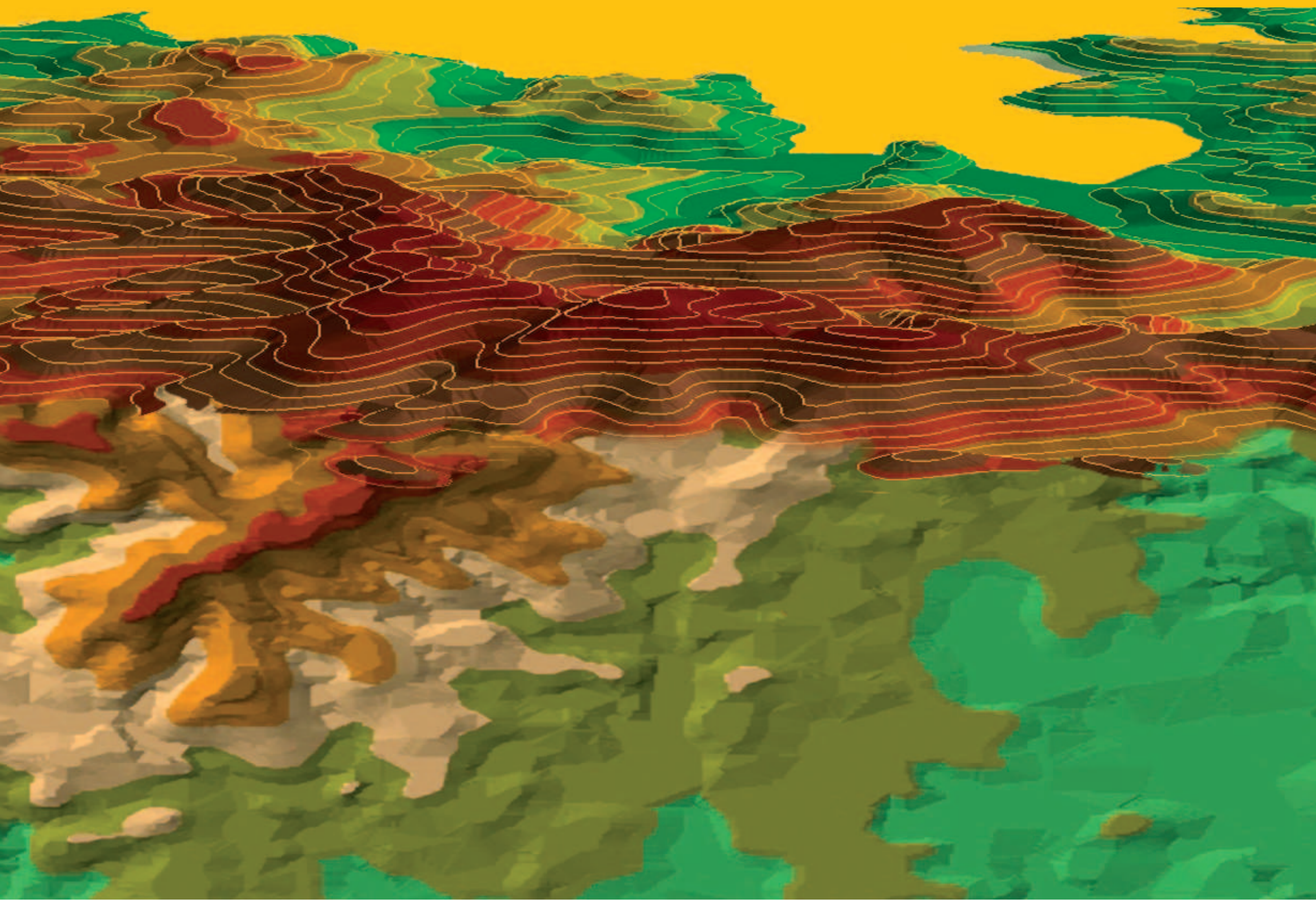
Vol. 24/2016 No. 2

MORAVIAN GEOGRAPHICAL REPORTS

ÚGN



MORAVIAN GEOGRAPHICAL REPORTS

PRICE

280 CZK (excluding VAT) per copy plus the postage
800 CZK (excluding VAT) per volume (four numbers per year) plus the postage

PUBLISHER

The Czech Academy of Sciences
Institute of Geonics, v. v. i.
Identification number: 68145535

MAILING ADDRESS

MGR, Institute of Geonics of the CAS, v. v. i.
Department of Environmental Geography
Drobného 28, 602 00 Brno, Czech Republic
(fax) 420 545 422 710
(e-mail) mgr@geonika.cz
(home page) http://www.geonika.cz

Brno, June 30, 2016

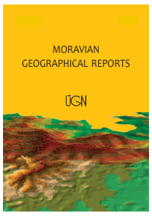
PRINT

NOVPRESS s.r.o., nám. Republiky 15, 614 00 Brno

ISSN 1210-8812 (Print)

ISSN 2199-6202 (Online)

Articles:



MORAVIAN GEOGRAPHICAL REPORTS

Institute of Geonics, The Czech Academy of Sciences

journal homepage: **http://www.geonika.cz/mgr.html**

doi: **10.1515/mgr-2016-0006**

DE GRUYTER OPEN

Conceptualising patterns of spatial flows: Five decades of advances in the definition and use of functional regions

Pavel KLAPKA [a]*, Marián HALÁS [a]

Abstract

Some fifty years in the development of ideas about the definition and use of functional regions are elaborated in this article, as an introduction to this Special Issue of the Moravian Geographical Reports. The conceptual basis for functional regions is discussed, initially in relation to region-organising interactions and their behavioural foundations. This paper presents an approach to functional regions which presumes that such regions objectively exist and that they are based on more or less tangible processes (however, a different view of regions is also briefly described). A typology of functional regions is presented and the development of methods for finding a definition of functional regions is discussed, as well as a typology for these methods. The final part of this article stresses the importance of functional regions in geographical research, and introduces some emerging new prospects in the study of functional regions.

Key words: *spatial flows; functional region; functional regionalisation; regional taxonomy; geographical thought*

Article history: Received 8 January 2016; Accepted 6 May 2016; Published 30 June 2016

1. Introduction

There is a long tradition in geographic research distinguishing between two basic types of regions: formal and functional regions (Robinson, 1953; Nystuen and Dacey, 1961; Haggett, 1965; Grigg, 1967; Abler et al., 1972; Symanski and Newman, 1973). These types differ in the character of their region-organising criteria: formal regions are based on scalar or vertical data; functional regions are based on vector or horizontal data (see Fig. 1). This division of data is based on their spatial characteristics. However, it should be noted that all geographical data have also their temporal dimension: they can be either instant (referring to one point in time, such as date of census) or periodical (recorded for a certain period). Apart from this division of data one should be aware that another role of time in geographical research regards the temporal evolution of geographic information.

Getting back to a spatial view of geographical data, scalar data are related to the concept of a site (Ullman, 1980), where importance is given to the vertical or static nature of this data, even though their distribution can vary over space. Very often this verticality is rather symbolic in human geography, where spatial distributions of various criteria overlay one another. In physical geography this verticality is more tangible: the character of the bedrock and the climate influence soil type, hydrological regime and the character of the vegetation at a particular site. In contrast vector data connect two sites, origin and destination, and thus are related to the concept of a situation (Ullman, 1980), where importance is given to the horizontal or movement nature of these data. These differences are mirrored in their distinct forms of spatial organisation and the inner structure of their respective types of regions. Papers in this Special Issue concentrate specifically on functional regions and functional regional taxonomy.

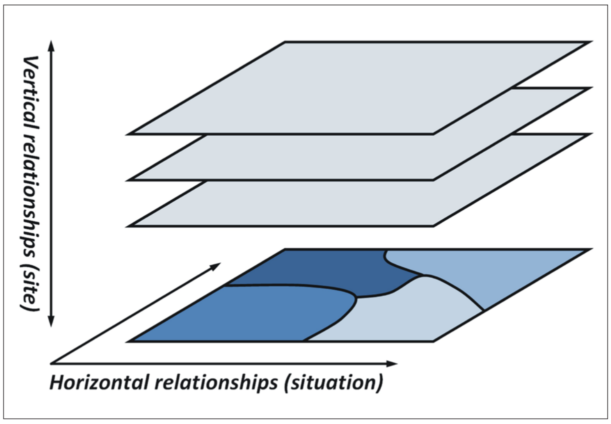


Fig. 1: Foundations of formal and functional regions
Source: Klapka et al. (2013a)

[a] Department of Geography, Faculty of Science, Palacký University Olomouc, Czech Republic (*corresponding author: P. Klapka, e-mail: *pavel.klapka@upol.cz*)

In 1967 the proceedings of the 4th general meeting of the Commission on Methods of Economic Regionalisation of the International Geographical Union, held in Brno, Czechoslovakia (September 7–12, 1965), were published by the Czechoslovak Academy of Sciences as the book "Economic Regionalisation", and edited by Miroslav Macka (1967). The book includes chapters by renowned geographers such as Brian Berry, Torsten Hägerstrand, Kazimierz Dziewoński, Hans Bobek and others. Fifty years have passed and the issue of functional regions and their definition has experienced rapid development, particularly in the methods of delineation of functional regions. This Special Issue of the Moravian Geographical Reports resumes the topic of functional regionalisation raised five decades ago and revives its importance, particularly in Central Europe, taking into account current knowledge and developments in this field.

The significance of reviving interest in the issue of functional regions and methods for their delineation is in accord with what we call "the second quantitative turn" in human geography. Regions objectively exist in reality no matter if individual perceptions and aggregated individual perceptions can make the concept of region somewhat blurry, both in a spatial and a cognitive sense. In this respect, the strong assumption that the boundaries of regions can be identified in space is not irrelevant at all. This is not in opposition to views which see regions as more or less temporary social constructions (see for instance, Murphy, 1991; Taylor, 1991; Terlouw, 2001). Even in this respect, Paasi (1991) sees, as part of their social construction (Terlouw, 2001), four shapes of regions (territorial, symbolic, institutional and functional), some of them being closer to the concept of a region as an objectively existing reality.

Given that the objective existence of regions is accepted or taken for granted, it is only logical that objective methods for their identification and definition should be applied, and that the objective methods lean towards quantitative approaches. This does not mean that objective regions are eternal entities. Quantitative approaches certainly examine the appearance, evolution, pulsation and demise of regions in time, but they concentrate on more tangible foundations for their existence than postmodern approaches. It is the quantitative approach that is discussed further in this paper. After decades of challenging the objective virtues of regions and concentrating on their social construction foundations, inspired for instance by the works of Bhaskar (1998, first published in 1979) and Giddens (1984), the importance of quantitative approaches to the identification and definition of regions is once again being acknowledged (for general consideration and personal confession see Johnston, 2008; Haggett, 2008; for the field of regional taxonomy see e.g. Casado-Díaz and Coombes, 2011; Farmer and Fotheringham, 2011). The behavioural foundations of regions, however, as seen by Giddens (1984) for instance, are briefly discussed in the following section.

2. The essence of spatial flows and the context of human behaviours

The existence of horizontal spatial flows is conditioned by the fact that planet Earth, as the subject matter for the discipline of Geography, is significantly non-homogeneous in both its physical geographical and human geographical traits, and this condition forms the basis for various kinds of spatial polarity. Usually this polarity has a tendency to precipitate spatial flows. Within the scope of physical geography these flows behave according to physical laws, and are manifested in the forms of wind streams and water flows. Wilson (1969) used the term "social physics" as an analogy between physical phenomena and social interaction. Within the scope of human geography these flows are induced by various manifestations of human behaviours, and this will be of interest in the following paragraphs.

In human geography, spatial flows (spatial interactions) have the character of aggregated individual horizontal flows, mobilities and contacts of persons, goods, finances and information. These attributes have their bases in the accomplishments and satisfaction of human needs, demands, purposes, or "projects", as they are called in time-geographical terminology (see e.g. Lenntorp, 1976; Pred, 1977; Timmermans, et al., 2002). In this respect, Golledge and Stimpson (1997) distinguish between two aspects of human behaviours: spatial behaviours and behaviours in space. The former concept refers to real movements in physical space; the latter concept comprises decision-making processes that underlie the actual spatial flows. They are goal- or "project"-oriented, to once again borrow from time-geographical terminology. Even though quantitative geography preferably works with the manifestations of spatial behaviours, the underlying processes should also be borne in mind. Within quantitative geography the role of spatial behaviour and perceptions is mostly reflected in the studies of movements, particularly related to the accessibility and shopping behaviour (see e.g. Blommestein, et al., 1980; more recently Kwan, 1998; Haynes, et al., 2003; Kwan, et al., 2003; Dijst, et al., 2008 to name just a few studies).

Spatial flows can be considered as reflections of both intentional and unintentional behaviours by individuals and society as an aggregation of individuals. An individual has to consider two moments, when speaking of spatial behaviours/behaviours in space, in order to satisfy psychological, social and economic needs (inter alia): the advantage of location and gaining maximum benefit from it, and the principle of least effort (Zipf, 1949) and optimising movements. Again, the heterogeneity of geographical space plays its role and the generally underlying geographical trait, distance, more precisely the relative distance, assumes crucial significance (for this factor, see Tobler, 1970 and his "first rule of geography"; Morrill, 1974 and his theory of the spatial organisation of society based, besides other factors, on the maximisation of spatial interaction; Abler et al., 1972 on the importance of relative space; and Ullmann, 1980 on "geography as the discipline of distance").

The aggregation of individual spatial behaviours and behaviours in space produces distinct spatial patterns, as some aspects of such behaviours gain importance over others, based on the hierarchy of needs and capability, coupling and authority constraints in spatio-temporal behaviours (see Lenntorp, 1976, Pred, 1977). Examples of situations that concern most of society are the residence-workplace relationships, residence-school relationships, shopping trips and leisure trips. These types of trips are largely responsible for the formation of spatial interaction patterns. Attention should also be paid to the temporal aspects of the above-mentioned situations, not only from the point of view of their evolution over time, but particularly from the point of view of their rhythm or period. Thus movements with a daily period of repetition are important for the purposes discussed in this paper. Both spatial flows and rhythms of human behaviours form recurring and regular behavioural patterns, which result in the spatial organisation of society, into the geographical organisation of space (see the classic works of Haggett, 1965; Abler et al., 1972; Morril, 1974). The conceptualisation of these patterns is discussed in the following section.

3. Functional regions

The first indications of interest in regions as organisational structures based on functional linkages occurred in the late pre-quantitative geography era (Platt, 1928; Brush, 1953; Robinson, 1953; Ullman, 1980, based on the ideas put forward by him in the first half of the 1950s). The greater attention paid to the patterns of spatial flows is related to the rise of quantitative geography that started in the late 1950s/early 1960s. Nystuen and Dacey (1961), Haggett (1965) and later Brown and Holmes (1971), used the term nodal region. In this instance, the spatial flows are oriented towards a core, a centre or a node, and the concept was inspired by the works of Dickinson (1930), Christaller (1933), Lösch (1940) and Isard (1956). All these approaches stressed the orientation of interaction movements towards a centre, a node.

During the same decade the use of another term occurred: functional region, but its lucid and rigorous definition was not stated, and sometimes it was used interchangeably with nodal region. As examples, consider: Philbrick (1957: 302) speaks of "areal functional organisation", then operates with the concept of nodality; Dziewoński (1967) quotes both terms as individual, Berry (1968) speaks of the organisational and functional aspect of these regions, though he considers nodality and polarisation to be their basis; Abler et al. (1972) stress functional relationships within functional regions; Morrill (1970) says in the first edition of his book that nodal region is a better term for functional region, while Grigg (1967) suggested that the term functional region be preferred to the term nodal region, although he saw them as near synonyms. In contrast, Brown and Holmes (1971) differentiated between the two terms. The ambiguity between the terms nodal region and functional region was discussed later, for instance by Symanski and Newman (1973). Unequivocal definitions for functional regions were put forward by Johnston and Rossiter (1981), who, in the case of planning regions conceived as functional regions, omitted the notion of the necessary orientation of interaction movements towards a node.

Based on our own earlier work (Klapka et al., 2013a; Halás et al., 2015) and inspiration drawn from Goodman (1970), Smart (1974), Bezák (2000), Karlsson and Olsson (2006), Farmer and Fotheringham (2011), we venture to put forward a simple and general concept of a functional region for theoretical discussion. In this respect, a functional region is seen as an organisational structure based on patterns of any relevant horizontal spatial relationships (e.g. vectors, interactions, movements, flows, etc.), and the concepts of functional regional autonomy and self-containment of a region that can be expressed by two interlinked principles: the principle of external separation and the principle of internal cohesiveness. This means that spatial relationships (their number, intensity) are maximised within a functional region and minimised across its borders, which ensure a high degree of functional regional autonomy (self-containment) for each respective functional region. Such a concept of a functional region asks for a simple identification criterion of a minimum of 50% of incident spatial relationships to occur within a region (nevertheless, in practice the percentage is usually higher) – see Fig. 2.

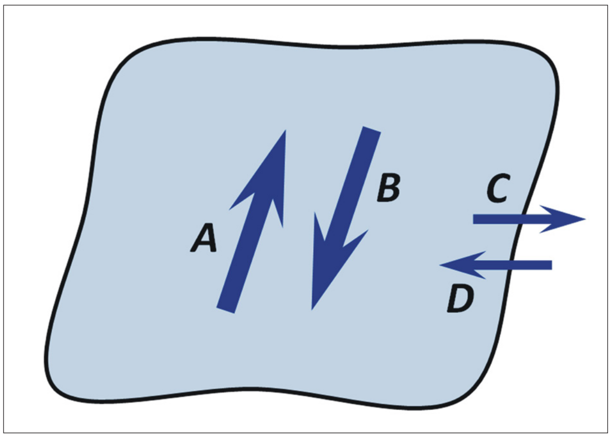


Fig. 2: Self-containment of a functional region
Source: Klapka, et al. (2013a)

Mathematically the self-containment of a region is expressed by $(A + B) > k(C + D)$; $k \geq 1$ where A and B are inner flows, C and D are cross border flows, and k is a coefficient setting the level of self-containment.

A finer classification of functional regions can be based on two criteria: inner structure and the character of region-organising relationships. As for the former criterion, functional regions reflect a so-called situational context (Ullman, 1980), when number, direction and intensity of horizontal spatial relationships vary across the space. This type of an organisational unity infers that such regions usually have a complex and non-homogeneous inner structure (unlike formal regions). In this respect, at least five theoretical models of functional regions can be distinguished (Klapka, et al., 2013a), see Fig. 3:

1. a functional region with a random pattern of inner spatial relationships;
2. a functional region with an oriented ordered pattern of inner spatial relationships, characterised by prevailing directions of flows;
3. a functional region with a channelled ordered pattern of inner spatial relationships, characterised by a concentration of flows into communication channels;
4. a functional region with a circular ordered pattern of inner spatial relationships, characterised by circulating flows; and
5. a functional region with a nodal ordered pattern of inner spatial relationships, characterised by a direction of flows towards a core (node).

Even though nodal regions are the most frequent instance of a functional region, the remaining instances in Figure 3 are not mere theoretical constructions, but they objectively exist. For instance, migration flows along short distances within a functional region at a local level, and family visits within a particular city zone have rather random patterns (see e.g. an earlier comment by Greer-Wooten and Gilmour, 1972). Some of the types can be determined by physical constraints and barriers such as a mountain range, a coastline, a huge river, etc. An important role in the spatial distribution of interactions is played by the shape and location of barriers. The less frequent type of circular flow can occur in this respect, for instance around a large lake which is a tourist attraction.

According to the second classification criterion, the character of a region-organising relationship, various types of movements (travel-to-work, travel-to-school, travel-

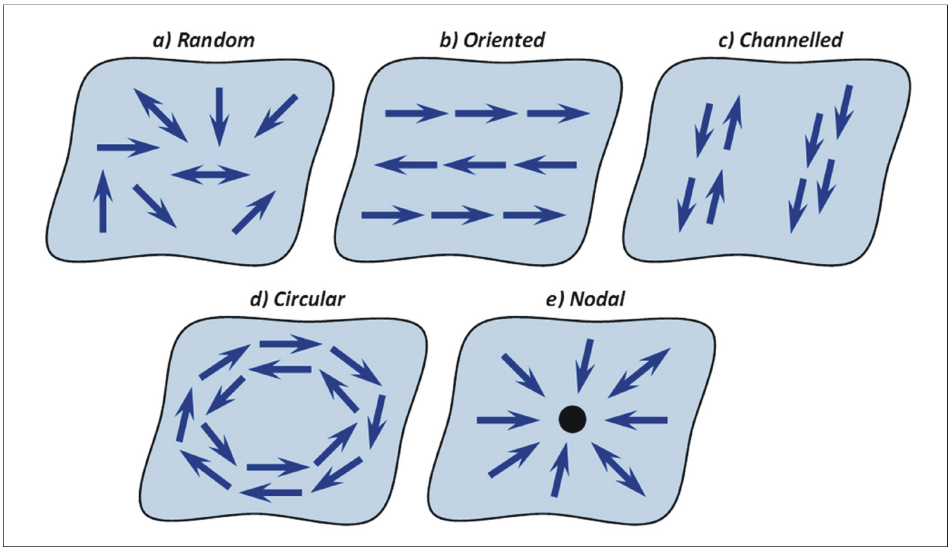


Fig. 3: Functional regions according to their inner structure
Source: adjusted according to Klapka, et al. (2013a)

to-services, leisure travels, etc.), their rhythm (e.g. daily, weekly) and types of core (e.g. urban, polycentric) are used to distinguish between functional regions. Thus functional regions based on travel-to-work flows are referred to as local labour market areas (originally discussed for example, by Goodman, 1970; Smart, 1974), or travel-to-work areas (originally discussed for example, by Ball, 1980; Coombes and Openshaw, 1982). Functional regions based on flows directly to an urban core are referred to as functional urban regions (originally discussed for example, by Berry, 1973; Hall and Hay, 1980), or, in cases where the flows have a daily rhythm as daily urban systems (originally discussed for example, by Berry, 1973; Hall, 1974; Coombes et al., 1979). All of these types of regions, as evidenced from the relevant literature, can be considered special instances of a general functional region.

Figure 4 shows a graphical expression of the most frequent types of specific functional regions, where the basic criterion, self-containment, is supplemented by further optional characteristics. Thus the functional urban region (Fig. 4a: FUR) needs to be organised around an urban core, the daily urban system (Fig. 4b: DUS) needs to be defined by daily movements and rhythms, and the local labour market area (Fig. 4c: LLMA) needs to be based on the interaction between workplace and residence.

4. Functional regional taxonomy

4.1. Conceptual framework

Functional regions are products of a functional regional taxonomy. A functional regional taxonomy is understood to be a set of approaches, methods and techniques used for the identification and definition of functional regions, which are usually based on the analysis of spatial relationships (interaction, movement, flows) between defined segments of geographic space. As such it is a part of a traditional and wide-ranging branch of human geography, i.e. spatial analysis and quantitative geography (see for example, Coombes, 2000).

A functional regional taxonomy has to take into account three crucial limitations. The first relates to the problem of the identification of geographical objects and the relevant hierarchical level needed for decisions concerning the choice of spatial zones to act as building blocks for a functional regional taxonomy. This can be called the principle of a basic spatial unit. The second limitation is expressed by the continuous character of geographic space and the distance separating basic spatial units. Again the first law of geography (see above reference to Tobler, 1970) is important in this respect. This limitation has advantageous effects in the formation of separated functional regions without the need to include information on the spatial neighbourhood of basic spatial units. The third limitation relates to the so-called modifiable areal unit problem (MAUP), which is addressed for instance by Openshaw (1984), Fotheringham and Wong (1991) and Unwin (1996).

A functional regional taxonomy is in fact an inherent part of MAUP (Baumann, et al., 1983; Cörvers, et al., 2009; Mitchell and Watts, 2010), as any effort to produce larger areas (regions) from a set of arbitrary and modifiable objects (basic spatial units) faces a considerable degree of spatial uncertainty and spatial bias. MAUP consists of two interlinked questions: how many larger areas should there be?; and, which means of amalgamating geographical objects

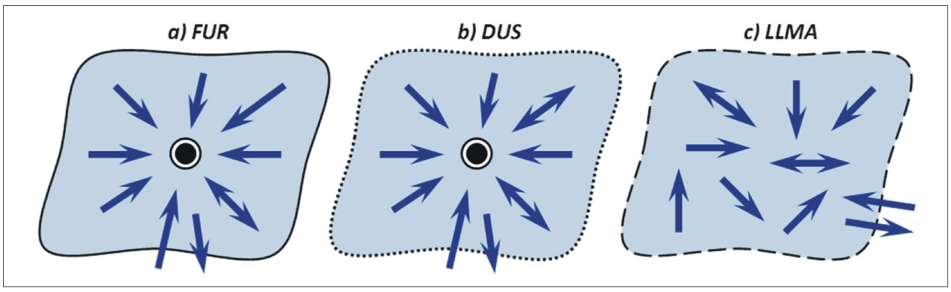


Fig. 4: Functional regions according to optional characteristics. Source: adjusted according to Klapka, et al. (2013a)

into larger areas should be used? In both cases, there are an almost infinite number of choices. The first question is known as the scale problem; the second question is known as the aggregation or zoning problem.

Two conceptual prerequisites frame the scope of a functional regional taxonomy: it is goal-oriented, and it has an exploratory rather than confirmatory character. The first prerequisite demands that there is a rigorous objective stating what is to be reached by a functional regional taxonomy and why. The second prerequisite holds that the results of a functional regional taxonomy are not known in advance. It should also be observed that there is no sole correct methodology for the analysis of interaction data, and that different approaches, methods and techniques can provide considerably different results (Van der Laan and Schalke, 2001; Klapka, et al., 2014).

Generally, the identification and definition of functional regions can be achieved through one or both of two interlinked tasks that can be viewed as two perspectives on the same problem. The first task is to search for similarities in spatial relationships across geographical space. The similarity is expressed by the intensity of spatial relationships, when higher intensity implies higher similarity, i.e. significant linkages between geographical objects. The second approach (see e.g. Coombes, 2000) is a quest for boundaries across geographical space. Retaining the argumentation of this paper based on the concepts of spatial non-homogeneity and spatial interaction, the boundaries can be conceived as areas where few or no spatial relationships occur.

Finally, there is the issue of the contiguity of functional regions, i.e. the difference between typological and individual functional regional taxonomy. The procedures leading to the definition of individual functional regions should theoretically comprise a contiguity constraint. Some methods, such as the Intramax (see below), comprise such a constraint. In contrast, the majority of approaches do not include this constraint, since it is the character of a space, the role of distance and behavioural constraints that produce contiguous typological functional regions, and these can be considered to be individual functional regions. To sum up, the taxonomic similarity of basic spatial units is closely related to the spatial proximity of these units.

4.2 Typology of approaches

Existing typologies of approaches leading to the definition of functional regions have been presented by Coombes (2000), Van der Laan and Schalke (2001), Casado-Díaz and Coombes (2011), Farmer and Fotheringham (2011), and Klapka et al. (2014), among others. Some of the terms in Table 1 have been used by the above-mentioned works, but sometimes with different meanings (see for example, the definition of hierarchical methods by Casado-Díaz and Coombes, 2011). In this paper, an attempt to provide a more detailed classification of functional regional taxonomic tasks is put forward in order to add to this field of study. All approaches come from the analysis of an interaction matrix, storing the information on contacts (i.e. flows and linkages) between basic spatial units.

Four criteria, each allowing two possibilities, are used in order to classify methodological approaches to the identification of functional regions (i.e. regional classes). The criteria are ranked in descending order from more generic to more specific (see Tab. 1).

The first criterion is based on the direction of a regional class formation, when either basic spatial units are grouped into larger clusters, or the set of basic spatial units is divided into smaller subsets. The second criterion distinguishes between methods that follow a general clustering principle, where a regional class is formed in one stage, and methods that are comprised of several stages, all of which can have various objectives. The basic difference is that, in the former case, once two basic spatial units (or clusters) are amalgamated, they can never be dissolved, while in the latter case the final clusters are formed after all stages are completed and all the rules fulfilled, and it is possible that a proto-cluster can be dissolved during the procedure. The third criterion distinguishes between tasks when the number of final clusters is known in advance (non-hierarchical methods) or it is not (hierarchical methods)[1]. Finally, the fourth criterion distinguishes between cases where the interaction matrix is interpreted as a graph, and where it is conceived as a numerical expression of the dissimilarity of respective basic spatial units.

Theoretically, each approach should be classifiable within each criterion. It must be admitted that some approaches do not exist, for logical reasons, or they are not used for practical reasons (they do not provide geographically acceptable results or are too demanding for computer processing). A survey of the literature shows that several selected approaches have been favoured so far. Graph-oriented methods occurred first (e.g. Nystuen and Dacey, 1961; Slater, 1976; Holmes and Haggett, 1977) as their application is quite simple, without the need for robust computation. In principle, they use solid or floating thresholds in order to identify significant flows,

Criterion	Approach
1. Direction of a regional class formation	*Agglomerative*
	Divisive
2. Character of a class-forming procedure	*Clustering*
	Rule-based (multistage)
3. Form of a regional class formation	*Hierarchical*
	Non-hierarchical
4. Form of an interaction matrix analysis	*Graph-oriented*
	Numerical

Tab. 1: Classification of approaches to a functional regional taxonomy. Source: authors' design

[1] This criterion should not be confused with a result of functional regional taxonomy, which can be hierarchical (more layers of usually nested functional regions) and non-hierarchical (only one layer of functional regions). In this case the terms hierarchical and non-hierarchical refer not to the result, but to the form of construction of a regional layer.

which occur on an oriented graph. The significant flows can be based for instance on primary linkage, minimum directionality linkage, salient linkage, or hierarchical linkage (Holmes and Haggett, 1977). These methods often produce unsatisfactory results, yielding disordered regional patterns which need subjective intervention in order to acquire contiguous and separated regional classes. Their strength lies in the simple preliminary analysis of a regional system. These methods have been used relatively recently, however, for example by Van der Laan and Schalke (2001), Karlsson and Olsson (2006), Drobne et al. (2010), Halás et al. (2010), and Farmer and Fotheringham (2011).

The second group of methods that has been widely used is based on numerical and clustering approaches. These methods involve the application of general cluster analysis on spatial problems using various linkage measures. Brown and Holmes (1971) used the functional distance approach based on mean first passage time (MFPT), where the interaction between two basic spatial units is taken as the measure of similarity in taxonomic space. Keane (1978) and, relatively recently, Cörvers et al. (2009), have also applied the functional distance method. According to the number of citations, the most successful approach in this group is the Intramax procedure. It was designed by Masser and Brown (1975) and refined by Masser and Scheurwater (1978), in reaction to comments made by Hirst (1977). This method builds upon Ward's (1963) hierarchical clustering procedure, which is adjusted for interaction data. The Intramax approach was applied by Nel et al. (2008), Drobne and Bogataj (2012), and Landré (2012) relatively recently.

The third group of methods, the so-called rule-based procedures, is comprised of the approaches most widely used today. The origins of the rule-based approaches can be found in the work of Smart (1974). His basic idea was further developed in great detail into a complex regionalisation algorithm designed by the Centre for Urban and Regional Development Studies (CURDS) in Newcastle, UK. Up to the present, three variants of the CURDS algorithm have been developed (Coombes et al. 1982, 1986; Coombes and Bond, 2008; Coombes, 2010). The principle of these methods comes from the definition of a set of rules that are applied in several stages and determine the results of the analyses. The rules are often used iteratively in order to reach or approximate an optimal solution. Minor adjustments to the algorithms regarding the constraint function were proposed by Casado-Díaz (2000) and Halás et al. (2015). Apart from the above-mentioned works concerning the territory of Great Britain, multistage methods were applied in a number of mainly European countries: Italy (Sforzi, 1997), Slovakia (Bezák, 2000; Halás et al., 2014), Spain (Casado-Díaz, 2000), New Zealand (Papps and Newell, 2002; Newell and Perry, 2005), Australia (Watts, 2004), Belgium (Persyn and Torfs, 2011), Poland (Gruchociak, 2012), and the Czech Republic (Klapka et al., 2014). There are also other types of rule-based approaches: for example, a graph theoretical multistage approach, differing from the CURDS algorithm, has been proposed by Kropp and Schwengler (2016).

Even this smaller number of three groups of methods of functional regional taxonomy that were put into practice gave rise to discussions concerning two points: a comparison of the methods, and criticism of the methods. The insufficiencies of the graph theoretical methods have been mentioned already. The numerical and clustering approaches were criticised for being too heuristic (Ball, 1980; Coombes and Openshaw 1982). In contrast, the multistage methods faced criticism for being subjective in a certain way and using pre-defined arbitrary criteria (e.g. Mitchell and Watts, 2010). Halás et al. (2015) proposed a procedure to mitigate the effects of arbitrary choice in the CURDS algorithms.

Despite the criticisms, however, the results of different methods were often compared. Masser and Scheurwater (1980) compared the functional distance method, the graph theoretical method iterative proportional fitting procedure (IPFP), and the Intramax method. Fischer et al. (1993) compared the IPFP and Intramax procedures. Watts (2009, 2013) made comparisons between the results of the CURDS algorithm and the Intramax method. Drobne et al. (2010) compared some more sophisticated graph theoretical methods with the Intramax approach. Klapka et al. (2013b) compared the results of the CURDS algorithm with simpler graph theoretical methods based on the primary linkage. Landré and Håkansson (2013) compared the results of the Intramax with graph theoretical methods. Most works cited in this paragraph reach the conclusion that aggregation procedures such as Intramax and the CURDS algorithm produce more correct results than graph theoretical methods and matrix transformation methods (e.g. IPFP). The question of whether to use hierarchical aggregation (Intramax) or rule-based aggregation (the CURDS algorithm) seems to remain open: it depends in part on the objectives of the research, but also, paradoxically, on subjective factors due to the preferences of the researcher.

5. Importance of functional regions and discovering future prospects

The problem of the definition of functional regions has a wide range of implications for the development of both geographical theory and practice. As far as the practical point of view is concerned, it has long been acknowledged by Haggett (1965) and Dziewoński (1967) that functional regions can serve better as a geographical tool for normative use than administrative regions. Functional regions have a vital role in fields such as spatial planning, regional economics, statistical geography, transport geography, etc.; effectively, in all cases where there is a need for some kind of spatial units with internal geographical logic in order to reduce possible spatial bias. It is considered that administrative, political and some statistical divisions do not necessarily reflect existing geographical realities, and that they may manifest a significant degree of inefficiency (see for instance, Coombes, 2010; Casado-Díaz and Coombes, 2011; Farmer and Fotheringham, 2011).

The theoretical implications appear to be more inspiring for geographers to acknowledge the importance of the study of functional regions. Such implications are more complex and mutually conditioned and unfolded. Most of them are grounded in spatial uncertainty or in spatial bias, which is a specific manifestation of the role and property of most geographic (spatial) characteristics. In this argument, spatial uncertainty is occasioned by the continuous character of geographic space and its measurable elements. In the most general sense, the above-mentioned MAUP emerges again. Every effort to define a system of functional regions faces questions concerning the number and composition of regional classes. These questions are complicated to solve and that is the challenge for future research. The solutions to MAUP, however, are strongly dependent on the research objectives used in defining functional regions. Possible directions of research can be aimed at the analysis of spatial associations, spatial distributions and spatial variability, or any combination of such phenomena.

In a more specific sense, spatial uncertainty unfolds from a more probabilistic rather than a deterministic concept of functional regions, which is again a reflection of the continuous character of geographic space. This approach concedes that the level of belonging of geographic objects (in this case, basic spatial units) to a regional class need not be an unambiguous inference. This idea comes from fuzzy set theory, where the membership of an object in a set can vary from zero to absolute validity. Coombes (2000) was among the first to suggest a more fuzzy approach to the definition of functional regions. Feng (2009) presented a method for the assessment of fuzziness in regional systems through the adoption of a so-called membership function. This function was later improved by Watts (2009, 2013). The adoption of the fuzzy set approach to functional regions opens up the possibility of identifying overlapping functional regions, even though their existence is usually ruled out by the principle that a basic spatial unit should belong to just one regional class. The issue of overlapping functional regions is discussed by Killer and Axhausen (2011), and the issue of the fuzziness of regional systems is discussed further by Feria et al. (2015). The fuzzy set approach, taken more generally, can be a cornerstone for the identification of so-called pulsating functional regions, which can vary in space and time and, turning full circle, are again more general consequences of spatial uncertainty.

Another future prospect in the study of the definition of functional regions is linked to the preceding paragraph and it is grounded in the search for an efficient way to reduce the risk of spatial uncertainty in the results of functional regional taxonomy. It is a characteristic of most procedures defining functional regions that after they reach a solution, according to set parameters, they terminate. These procedures are based on the so-called greedy algorithms. There might be a better solution with regard to the total self-containment or fuzziness of a regional system, however, which has not been identified by any particular method. This clearly requires a refinement of the distribution of basic spatial units into regional classes.

An emerging field of study in this respect is the application of evolutionary or genetic algorithms (clustering techniques) drawing inspiration from biology and genetics. Pioneering works in functional regional taxonomy were put forward by Flórez-Revuelta et al. (2008), Martínez-Bernabeu et al. (2012), and their approach was modified by Alonso et al. (2015). These heuristic methods generate a number of variant solutions (generations) to the large regionalisation problem. They are based on evolutionary principles, such as selection, mutation and crossover, and optimise a so-called fitness function.

Another way to tackle the issue of refining existing solutions to functional regionalisation is the use of non-hierarchical clustering strategies. Regarding this, new prospects are offered by the application of soft clustering methods, such as the fuzzy c means (FCM) algorithm, which is a soft variant of the frequently used non-hierarchical k means algorithm (for general definition and use, see for example, Bezdek et al., 1984, and Yang, 1993). As in the preceding case of algorithms based on evolutionary computation methods, a crucial role in these methods is played by a so-called objective function, which means that it should be very well defined and designed. The principle of the methods is grounded in the search for a global optimum through maximisation (or minimisation – it depends on the logic of the clustering algorithm) of the objective function. In a functional regional taxonomy this means that basic spatial units are iteratively reallocated between regional classes until there is no improvement in the value of the objective function, and all predefined criteria, such as self-containment and size, are met.

6. Special Issue on functional regional taxonomy

The present theoretical paper has introduced this Special Issue of the Moravian Geographical Reports on functional regions and functional regionalisation. It covers a wider spectrum of related problems in five papers, presenting specific contributions to the field of methodology and use of functional regions. Drobne and Lakner discuss the use of different objective functions in hierarchical aggregation procedures for the definition of functional regions. Martínez-Bernabeu and Casado-Díaz propose a methodology, based on evolutionary computation, to overcome the insufficiencies caused by limitations exposed by basic spatial units in the construction of functional regions. Erlebach, Tomáš and Tonev present the results of three methods to define functional meso-regions of the Czech Republic. Klapka, Halás, Netrdová and Nosek discuss the suitability of functional micro-regions for spatial analysis and present a comparison of functional regions and administrative regions in this respect. Olsson presents a spatial interaction modelling approach, particularly the issue of accessibility measures based on the use of distance-friction parameters, to the identification of functional regions.

Acknowledgement

The paper has been supported by the project of the Czech Science Foundation: "Spatial uncertainty and fuzzy regional systems: identification, analysis and implications of pulsating functional regions" (16-13502S). The authors are grateful for this support.

References:

ABLER, R., ADAMS, J. S., GOULD, P. (1972): Spatial organization: The geographer's view of the world. London, Prentice Hall.

ALONSO, M. P., BEAMONTE, A., GARGALLO, P., SALVADOR, M. (2015): Local labour markets delineation: an approach based on evolutionary algorithms and classification methods. Journal of Applied Statistics, 42(5): 1043–1063.

BALL, R. M. (1980): The use and definition of travel-to-work areas in Great Britain: some problems. Regional Studies, 14(2): 125–139.

BAUMANN, J. H., FISCHER, M. 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(1): 53–83.

BERRY, B. J. L. (1968): A synthesis of formal and functional regions using a general field theory of spatial behaviour. In: Berry, B. J. L, Marble, D. F. [eds.]: Spatial Analysis (pp. 419–428). Englewood Cliffs, Prentice-Hall.

BERRY, B. J. L. (1973): Growth centres in the American urban system. Cambridge, Mass., Ballinger.

BEZÁK, A. (2000): Funkčné mestské regióny na Slovensku. Geographia Slovaca 15, Bratislava, Geografický ústav SAV.

BEZDEK, J. C., EHRLICH, R., FULL, W. (1984): FCM: The fuzzy c-means clustering algorithm. Computers & Geosciences, 10(2): 191–203.

BHASKAR, R. (1998): The Possibility of Naturalism: A Philosophical Critique of the Contemporary Human Sciences, Third edition. London and New York, Routledge.

BLOMMESTEIN, H., NIJKAMP, P., VAN VEENENDAAL, W. (1980): Shopping Perceptions and Preferences: A Multidimensional Attractiveness Analysis of Consumer and Entrepreneurial Attitudes. Economic Geography, 56(2): 155–174.

BROWN, L. A., HOLMES, J. (1971): The delimitation of functional regions, nodal regions, and hierarchies by functional distance approaches. Journal of Regional Science, 11(1): 57–72.

BRUSH, J. E. (1953): The hierarchy of central places in Wisconsin. Geographical Review, 53(3): 380–402.

CASADO-DÍAZ, J. M. (2000): Local labour market areas in Spain: a case study. Regional Studies, 34(9): 843–856.

CASADO-DÍAZ, J. M., COOMBES, M. (2011): The delineation of 21st century local labour markets areas: a critical review and a research agenda. Boletín de la Asociación de Geógrafos Españoles, 57: 7–32.

CHRISTALLER, W. (1933): Die zentralen Orte in Suddeutschland. Jena, Gustav Fischer.

COOMBES, M. (2010): Defining labour market areas by analysing commuting data: innovative methods in the 2007 review of Travel-to-Work Areas. In: Stillwell, J., et al. [eds.]: Technologies for migration and commuting analysis: spatial interaction data applications (pp. 227–241). Hershey, IGI Global.

COOMBES, M. G. (2000): Defining locality boundaries with synthetic data. Environment and Planning A, 32(8): 1499–1518.

COOMBES, M. G., BOND, S. (2008): Travel-to-Work Areas: the 2007 review. London, Office for National Statistics.

COOMBES, M. G., DIXON, J. S., GODDARD, J. B., OPENSHAW, S., TAYLOR, P. J. (1982): Functional regions for the population census of Great Britain. In: Herbert, D. T., Johnston, R. J. [eds.]: Geography and the Urban Environment. Progress in Research and Applications 5 (pp. 63–112). Chichester, John Wiley and Sons Ltd.

COOMBES, M. G., DIXON, J. S., GODDARD, J. B., OPENSHAW, S., TAYLOR, P. J. (1979): Daily urban systems in Britain: from theory to practice. Environment and Planning A, 11(5): 565–574.

COOMBES, M. G., GREEN, A. E., OPENSHAW, S. (1986): An Efficient Algorithm to Generate Official Statistical Reporting Areas: The Case of the 1984 Travel-to-Work Areas Revision in Britain. The Journal of the Operational Research Society, 37(10): 943–953.

COOMBES, M. G., OPENSHAW, S. (1982): The use and definition of travel-to-work areas in Great Britain: some comments. Regional Studies, 16(2): 141–149.

CÖRVERS, F., HENSEN, M., BONGAERTS, D. (2009): Delimitation and coherence of functional and administrative regions. Regional Studies, 43(1): 19–31.

DICKINSON, R. E. (1930): The regional functions and zones of influence of Leeds and Bradford. Geography, 15(7): 548–557.

DIJST, M., FARAG, S., SCHWANEN, T. (2008): A comparative study of attitude theory and other theoretical models for understanding travel behaviour. Environment and Planning A, 40(4): 831–847.

DROBNE, S., BOGATAJ, M. (2012): A method to define the number of functional regions: An application to NUTS 2 and NUTS 3 levels in Slovenia. Geodetski vestnik, 56(1): 105–150.

DROBNE, S., KONJAR, M., LISEC, A. (2010): Razmejitev funkcionalnih regij Slovenije na podlagi analize trga dela. Geodetski vestnik, 54(3): 481–500.

DZIEWOŃSKI, K. (1967): Concepts and terms in the field of economic regionalization. In: Macka, M. [ed.]: Economic regionalization (pp. 25–36). Praha, Academia.

FARMER, C. J. Q., FOTHERINGHAM, A. S. (2011): Network-based functional regions. Environment and Planning A, 43(11): 2723–2741.

FENG, Z. (2009): Fuzziness of Travel to Work Areas. Regional Studies, 43(5): 707–720.

FERIA, J. M., CASADO-DÍAZ, J. M., MARTÍNEZ-BERNABÉU, L. (2015): Inside the metropolis: the articulation of Spanish metropolitan areas into local labor markets. Urban Geography, 36(7): 1018–1041.

FISCHER, M. M., ESSLETZBICHLER, J., GASSLER, H., TRICHTL, G. (1993): Telephone communication patterns in Austria: a comparison of the IPFP-based graph-theoretic and the Intramax approaches. Geographical Analysis, 25(3): 224–233.

FLÓREZ-REVUELTA, F., CASADO-DÍAZ, J. M., 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.

FOTHERINGHAM, A. S., WONG, D. W. S. (1991): The modifiable areal unit problem in multivariate statistical analysis. Environment and Planning A, 23(7): 1025–1044.

GIDDENS, A. (1984): The Constitution of Society. Outline of the Theory of Structuration. Cambridge, Polity.

GOLLEDGE, R. G., STIMPSON, R. J. (1997): Spatial behavior: a geographical perspective. New York–London, The Guildford Press.

GOODMAN, J. F. B. (1970): The definition and analysis of local labour markets: some empirical problems. British Journal of Industrial Relations, 8(2): 179–196.

GREER-WOOTTEN, B., GILMOUR, G. M. (1972): Distance and Directional Bias in Migration Patterns in Depreciating Metropolitan Areas. Geographical Analysis, 4(1): 92–97.

GRIGG, D. (1967): Regions, models, and classes. In: Chorley, R. J., Haggett, P. [eds.]: Models in Geography (pp. 461–509). London, Methuen.

GRUCHOCIAK, H. (2012): Delimitacja lokalnych rynków pracy v Polsce, Przegląd statystyczny. Numer specjalny, 2: 277–297.

HAGGETT, P. (1965): Locational Analysis in Human Geography. London, Arnold.

HAGGETT, P. (2008): The spirit of quantitative geography. Geographical Analysis, 40(3): 226–228.

HALÁS, M., KLADIVO, P., ŠIMÁČEK, P., MINTÁLOVÁ, T. (2010): Delimitation of micro-regions in the Czech Republic by nodal relations. Moravian Geographical Reports, 18(2): 16–22.

HALÁS, M., KLAPKA, P., BEDNÁŘ, M., TONEV, P. (2015): An alternative definition and use for the constraint function for rule-based methods of functional regionalisation. Environment and Planning A, 47(5): 1175–1191.

HALÁS, M., KLAPKA, P., BLEHA, B., BEDNÁŘ, M. (2014): Funkčné regióny na Slovensku podľa denných tokov do zamestnania. Geografický časopis, 66(2): 89–114.

HALL, P. (1974): The containment of urban England. Geographical Journal, 140(3): 386–408.

HALL, P., HAY, D. (1980): Growth centres in the European urban system. London, Heinemann Educational Books.

HAYNES, R., LOVETT, A., SÜNNENBERG, G. (2003): Potential accessibility, travel time, and consumer choice: geographical variations in general medical practice registrations in eastern England. Environment and Planning A, 35(10): 1733–1750.

HIRST, M. A. (1977): Hierarchical aggregation procedures for interaction data: a comment. Environment and Planning A, 9(1): 99–103.

HOLMES, J. H., HAGGETT, P. (1977): Graph theory interpretation of flow matrices: a note on maximization procedures for identifying significant links. Geographical Analysis, 9(4): 388–399.

ISARD, W. (1956): Regional science, the concept of region, and regional structure. Papers in Regional Science, 2(1): 13–26.

JOHNSTON, R. (2008): Quantitative Human Geography: Are We Turning Full Circle? Geographical Analysis, 40(3): 332–335.

JOHNSTON, R. J., ROSSITER, D. J. (1981): An approach to the delimitation of planning regions. Applied Geography 1(1): 55–69.

KARLSSON, C., OLSSON, M. (2006): The identification of functional regions: theory, methods, and applications. The Annals of Regional Science, 40(1): 1–18.

KEANE, M. J. (1978): A functional distance approach to regionalization. Regional Studies, 12(3): 379–386.

KILLER, V., AXHAUSEN, K. W. (2011): Understanding overlapping functional commuting regions with confidence ellipses and social network methods. Working Papers Traffic and Spatial Planning, 714, Zurich, IVT, ETH Zurich.

KLAPKA, P., HALÁS, M., ERLEBACH, M., TONEV, P., BEDNÁŘ, M. (2014): A multistage agglomerative approach for defining functional regions of the Czech Republic: the use of 2001 commuting data. Moravian Geographical Reports, 22(4): 2–13.

KLAPKA, P., HALÁS, M., TONEV, P. (2013a): Functional regions: concept and types. In 16th International Colloquium on Regional Sciences, Conference Proceedings (pp. 94–101). Brno, Masaryk University.

KLAPKA, P., HALÁS, M., TONEV, P., BEDNÁŘ, M. (2013b): Functional regions of the Czech Republic: comparison of simple and advanced methods of regional taxonomy. Acta Universitatis Palackianae Olomucensis, Facultas Rerum Naturalium, Geographica, 44(1): 45–57.

KROPP, P., SCHWENGLER, B. (2016): Three-Step Method for Delineating Functional Labour Market Regions. Regional Studies, 50(3): 429–445.

KWAN, M. P. (1998): Space-time and integral measures of individual accessibility: a comparative analysis using a point-based framework. Geographical analysis, 30(3): 191–216.

KWAN, M. P., MURRAY, A. T., O'KELLY, M. E., TIEFELSDORF, M. (2003): Recent advances in accessibility research: Representation, methodology and applications. Journal of Geographical Systems 5(1): 129–138.

LANDRÉ, M. (2012): Geoprocessing journey-to-work data: delineating commuting regions in Dalarna, Sweden. ISPRS International Journal of Geo-Information, 1(3): 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): 1–19.

LENNTORP, B. (1976): Paths in Space-Time Environments. A Time-Geographic Study of Movement Possibilities of Individuals. Lund Studies in Geography, Series B, 44. Lund, Royal University of Lund.

LÖSCH, A. (1940): Die räumliche Ordnung der Wirtschaft. Eine Untersuchung über Standort, Wirtschaftsgebiete und internationalem Handel. Jena, Gustav Fischer.

MACKA, M. [ed.] (1967): Economic Regionalisation. Praha, Academia.

MARTÍNEZ-BERNABEU, L., FLÓREZ-REVUELTA, F., CASADO-DÍAZ, J. M. (2012): Grouping genetic operators for the delineation of functional areas based on spatial interaction. Expert Systems with Applications, 39(8): 6754–6766.

MASSER, I., BROWN, P. J. B. (1975): Hierarchical aggregation procedures for interaction data. Environment and Planning A, 7(5): 509–523.

MASSER, I., SCHEURWATER J. (1978): The specification of multi-level systems for spatial analysis. In: Masser, I., Brown, P. J. B. [eds.]: Spatial representation and spatial interaction (pp. 151–172). Studies in Applied Regional Science, 10, Leiden–Boston, Martinus Nijhoff.

MASSER, I., SCHEURWATER J. (1980): Functional regionalisation of spatial interaction data: an evaluation of some suggested strategies. Environment and Planning A, 12(12): 1357–1382.

MITCHELL, W. F., WATTS, M. J. (2010): Identifying functional regions in Australia using hierarchical aggregate techniques. Geographical Research, 48(1): 24–41.

MORRILL, R. L. (1970): The Spatial Organization of Society. Belmont, Wadsworth Publishing Company.

MORRILL, R. L. (1974): The Spatial Organization of Society. 2nd edition. North Scituate, Duxbury Press.

MURPHY, A. B. (1991): Regions as Social Construct: The Gap between Theory and Practice. Progress in Human Geography, 15(1): 23–35.

NEL, J. H., KRYGSMAN, S. C., DE JONG, T. (2008): The identification of possible future provincial boundaries for

South Africa based on an intramax analysis of journey-to-work data. ORiON, 24(2): 131–156.

NEWELL, J. O., PERRY, M. (2005): Explaining continuity in New Zealand's local labour market areas 1991 to 2001. Australasian Journal of Regional Studies, 11(2): 155–174.

NYSTUEN, J. D., DACEY, M. F. (1961): A graph theory interpretation of nodal regions. Regional Science Association, Papers and Proceedings, 7(1): 29–42.

OPENSHAW, S. (1984): The Modifiable Areal Unit Problem. CATMOG 38. Norwich, GeoBooks.

PAASI, A. (1991): Deconstructing regions: notes on the scales of spatial life. Environment and Planning A, 23(2): 239–256.

PAPPS, K. L., NEWELL, J. O. (2002): Identifying functional labour market areas in New Zealand: a reconnaissance study using travel-to-work data. Discussion Paper 443. Bonn, Institute for the Study of Labor.

PERSYN, D., TORFS, W. (2011): Functional labour markets in Belgium: evolution over time and intersectoral comparison. In: Discussion Paper 17 (pp. 1–17). Vlaams Instituut voor Economie en Samenleving. Leuven, Katholieke Universiteit.

PHILBRICK, A. K. (1957): Principles of areal functional organization in regional human geography. Economic Geography, 33(4): 299–336.

PLATT, R. S. (1928): A detail of regional geography: Ellison Bay, community as an industrial organism. Annals of the Association of American Geographers, 18(2): 81–126.

PRED, A. (1977): The Choreography of Existence: Comments on Hägerstrand´s Time-geography and its Usefulness. Economic Geography, 53(2): 207–221.

ROBINSON, G. W. S. (1953). The geographical region: form and function. The Scottish Geographical Magazine, 69(2): 49–58.

SFORZI, F. [ed.] (1997): I sistemi locali del lavoro. 1991. Roma, Istat.

SLATER, P. B. (1976): A hierarchical regionalization of Japanese prefectures using 1972 interprefectural migration flows. Regional Studies, 10(1): 123–132.

SMART, M. W. (1974): Labour market areas: uses and definition. Progress in Planning, 2(4): 239–353.

SYMANSKI, R., NEWMAN, J. L. (1973): Formal, functional and nodal regions: three fallacies. Professional Geographer, 25(4): 350–352.

TAYLOR, P. J. (1991): A theory and practice of regions: the case of Europes. Environment and Planning D, 9(2): 183–195.

TERLOUW, K. (2001): Regions in geography and the regional geography of semiperipheral development. Tijdschrift voor economische en sociale geografie, 92(1): 76–87.

TIMMERMANS, H., ARENTZE, T., JOH, C.-H. (2002) Analysing space-time behaviour: new approaches to old problems. Progress in Human Geography, 26(2): 175–190.

TOBLER, W. (1970): A computer movie simulating urban growth in the Detroit region. Economic Geography, 46, Supplement: Proceedings. International Geographical Union, Commission on quantitative methods: 234–240.

ULLMAN, E. (1980): Geography as spatial interaction. Washington, University of Washington Press.

UNWIN, D. J. (1996): GIS, spatial analysis and spatial statistics. Progress in Human Geography, 20(4): 540–551.

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(2): 201–221.

WARD, J. H. (1963): Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association, 58(301): 236–244.

WATTS, M. (2004): Local labour markets in New South Wales: fact or fiction? Centre of Full Employment and Equity, Working Paper 04–12. Callaghan, the University of Newcastle, Australia.

WATTS, M. (2009): Rules versus hierarchy: an application of fuzzy set theory to the assessment of spatial grouping techniques. In: Kolehmainen, M., et al. [eds.]: Adaptive and Natural Computing Algorithms. Lecture Notes in Computer Science 5495 (pp. 517–526). Berlin-Heidelberg, Springer.

WATTS, M. (2013): Assessing different spatial grouping algorithms: an application to the design of Australia's new statistical geography. Spatial Economic Analysis, 8(1): 92–112.

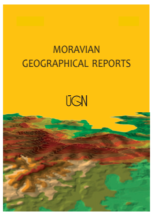
WILSON, A. (1969): Notes on some concepts in social physics. Regional Science Association, Papers, 22(1): 159–193.

YANG, M. S. (1993): A survey of fuzzy clustering. Mathematical and Computer modelling, 18(11): 1–16.

ZIPF, G. K. (1949): Human behavior and the principle of least effort. Cambridge, Mass., Addison-Wesley Press.

Please cite this article as:

KLAPKA, P., HALÁS, M. (2016): Conceptualising patterns of spatial flows: Five decades of advances in the definition and use of functional regions. Moravian Geographical Reports, 24(2): 2–11. Doi: 10.1515/mgr-2016-0006.



MORAVIAN GEOGRAPHICAL REPORTS

DE GRUYTER OPEN

Institute of Geonics, The Czech Academy of Sciences

journal homepage: **http://www.geonika.cz/mgr.html**

doi: **10.1515/mgr-2016-0007**

Intramax and other objective functions: The case of Slovenia

Samo DROBNE [a] *, Mitja LAKNER [a]

Abstract

The use of different objective functions in hierarchical aggregation procedures is examined in this paper. Specifically, we analyse the use of the original Intramax objective function, the sum-of-flows objective function, the sum-of-proportions-to-intra-regional-flows objective function, Smart's weighted interaction index, the first and second CURDS weighted interaction indices, and Tolbert and Killian's interaction index. The results of the functional regionalisation have been evaluated by self-containment statistics, and they show that the use of the original Intramax procedure tends to delineate operationally the most persuasive and balanced regions that, regarding the intra-regional flows, homogeneously cover the analysed territory. The other objective functions give statistically better but operationally less suitable results. Functional regions modelled using the original Intramax procedure were compared to the regions at NUTS 2 and NUTS 3 levels, as well as to administrative units in Slovenia. We conclude that there are some promising directions for further research on functional regionalisation using hierarchical aggregation procedures.

Keywords: *Intramax, objective function, contiguity constraint, hierarchical aggregation, functional region, functional regionalisation, Slovenia*

Article history: *Received 4 November 2016; Accepted 4 June 2016; Published 30 June 2016*

1. Introduction

The concept of the region has been traditionally one of the cornerstones of geographic thought. A region is a delimitated spatial system and an expression of organisational unity that differentiates it from another region (Abler et al., 1972; Gregory et al., 2009; Klapka et al., 2013). A functional region (FR) is a region organized by horizontal relations in a space in the form of spatial flows or interactions between parts of the region (Ullman, 1980), also called basic data units (BDUs)[1]. FR can be understood as a generalized pattern of spatial interactions. It can be defined by many different spatial interactions, including population flows (commuting to school or work, migration, shopping or recreation), traffic and goods flows (traffic and passenger flows by land/sea/air), commodity and financial flows, information flows (communications and newspaper circulation), gas/water/electricity flows (service connections), and so forth (Vanhove and Klaassen, 1987). In most of the literature, however, functional regions are defined by economic interactions. For example, Farmer and Fotheringham (2011) and Van der Laan and Schalke (2001) define a functional region as a spatially contiguous region in which aggregate supply and demand meet, and Karlsson and Olsson (2006) define a functional region as a territorial area characterised by a high frequency of intra-regional economic interactions (such as intra-regional trade in goods and services, labour commuting, and household shopping). Among different economic interactions, the daily interactions in the labour market are considered a good relative measure for the cohesion of a functional region (Ball, 1980; Cörvers et al., 2009; OECD, 2002). In this context, the basic characteristic of a functional region is the integrated labour market, in which intra-regional labour commuting, intra-regional job search, and search for labour demand are much more intensive than among the inter-regional counterparts (Karlsson and Olsson, 2006; Van der Laan and Schalke, 2001). Consequently, self-containment is the crucial characteristic of a functional region (Halás et al., 2015).

Several procedures for the delimitation of functional regions have been suggested (e.g. Coombes et al., 1986; Farmer, Fotheringham, 2011; Flórez-Revuelta et al., 2008;

[a] Department of Geodetic Engineering, Faculty of Civil and Geodetic Engineering, University of Ljubljana, Slovenia (*corresponding author: S. Drobne, e-mail: *Samo.Drobne@fgg.uni-lj.si*)

[1] Basic data units could be census units, statistical units, statistical local areas, settlements, communities, municipalities, postal zones, etc.

Kim et al., 2015; Masser and Brown, 1975; Slater, 1981). Farmer and Fotheringham (2011) identified three general classes of functional regionalisation procedures: hierarchical aggregation, multistage aggregation, and central place aggregation. Regardless of the approach, the aim of regionalisation procedures is to define as many functional regions as possible, subject to certain statistical constraints that ensure that the regions remain statistically and operationally valid (Casado-Diaz and Coombes, 2011).

In this paper, we analyse systems of hierarchical FRs modelled by the hierarchical aggregation procedure used in the original Intramax objective function (Masser and Brown, 1975, 1977) and six other objective functions. Seven sets of systems of hierarchical FRs[2] are analysed by well-known self-containment indicators (Goodman, 1970; Smart, 1974; Van der Laan and Schalke, 2001) and by the statistics suggested in this paper. The selected results of the hierarchical functional regionalisation are compared to regions at NUTS 2 and NUTS 3 levels and to administrative units (AUs) at LAU 1 level in Slovenia[3].

This paper is organised as follows. In section 2, we discuss the development and implementation of the Intramax method. In section 3, we introduce a methodology for modelling and evaluating seven sets of systems of hierarchical FRs in Slovenia, and for comparing the selected systems of FRs to official regions at NUTS 2 and NUTS 3 levels, as well as to AUs in Slovenia. The results are presented and discussed in section 4. Finally, section 5 concludes this topic of using different objective functions in the hierarchical aggregation procedure and suggests future research directions.

2. Development and implementation of the Intramax method

The Intramax method was developed by Masser and Brown (1975) for the purpose of analysing the structure of flows in a square interaction matrix (Brown and Pitfield, 1990). Masser and Brown (1977) emphasised two areas of the application of the Intramax procedure: the first of these was seen in dealing with the multi-level specification problem and with the association issue of data set reduction; and the second in functional regionalisation procedures. An example of the first application is the partitioning of a large interaction matrix into a number of spatially identifiable subsystems: *"within each of which there is observed to be a high level of flows but between which flows are small and links are weak"* (Brown and Pitfield, 1990: 60). Such principles are further discussed in Masser and Scheurwater (1978, 1980). The Intramax procedure was also suggested as a method of functional regionalisation (Masser and Brown, 1975). The results of such a regionalisation process are functional regions.

According to Masser and Brown (1975: 510): "… *[in each aggregation step, the Intramax procedure seeks] to maximise the proportion of the total interaction which takes place within the aggregations of basic data units that form the diagonal elements of the matrix, and thereby to minimise the proportion of cross-boundary movements in the system as a whole*". The authors reported, however, that the Intramax is a heuristic procedure which does not guarantee a global optimal solution to the partitioning problem (where maximum interaction flows would stay in the regions and less would cross the region's borders (Masser and Brown, 1977).

The incapability of the Intramax procedure to achieve a global optimal solution is mostly the consequence of the irreversibility of the aggregated BDUs/FRs (BDUs that are once aggregated in FR cannot be disaggregated any more). As reported many times (e.g. Alvanides et al., 2000; Casado-Diaz and Coombes, 2011; Coombes et al., 1986), the irreversibility of the aggregated BDUs/FRs is one of the most important shortcomings of the Intramax procedure, and a second problem is the indeterminacy of the number of FRs.

In its relative simplicity and its implementation in publicly available software[4], however, one sees the reason that the Intramax method has been used so many times to analyse FRs – of so many different kinds of interactions at very different levels of consideration, for example: labour market area delineation (Masser and Scheurwater, 1980; Feldman et al., 2005; Watts, 2009; Landré, 2012; Landré and Hĺkansson, 2013; Koo, 2012); housing market area delineation (Goetgeluk and de Jong, 2007; Brown and Hincks, 2008; Jaegal, 2013); commodity market delineation (Brown and Pitfield, 1990); world trade block delineation (Poon, 1997; Kohl and Brouver, 2014); functional economic region delineation (Mitchell et al., 2007, 2013; Mitchell and Stimson, 2010; Mitchell and Watts, 2010); telecommunication analysis (Fischer et al., 1993); to identify possible administrative or statistical regions (Nel et al., 2008; Drobne and Bogataj, 2012a, 2012b); transport regions (Krygsman et al., 2009); in the (allocation) analysis of services (Drobne and Bogataj, 2014, 2015), and so forth.

In the Intramax procedure, which is a modified version of Ward's (1963) hierarchical aggregation procedure, two spatial BDUs that show the most intensive relative interaction (in terms of both of the flows) are joined together and stay fused for the remainder of the aggregation process, which continues until all BDUs are fused. Here, we consider interaction flows as entries in the interaction matrix $T = [t_{ij}]$, where $t_{ij} \geq 0$ is an observed value of the cell entry in the i-th row and the j-th column, namely the interaction flow from BDU_i (origin i) to BDU_j (destination j). So, the Intramax analysis is a stepwise analysis. In each step of the aggregation (hereinafter "aggregation step" or

[2] In this paper, the term "system of hierarchical FRs" or, in short, "system of FRs", denotes delimitation of the analysed territory into FRs at each hierarchical level, but the term "set of systems of hierarchical FRs" or, in short, "set of systems of FRs" or "set of FRs", denotes all of the systems of hierarchical FRs modelled by an individual objective function.

[3] The NUTS (Nomenclature of Territorial Units for Statistics) classification is a hierarchical system for dividing up the economic territory of the EU for the purpose of: (a) the collection, development and harmonisation of EU regional statistics; (b) socio-economic analyses of the regions; and (c) framing of EU regional policies (EC, 2003; 2007). In Slovenia, there is only one region at NUTS 0 or NUTS 1 level (the whole state), there are two regions at NUTS 2 level and 12 regions at NUTS 3 level. Below the NUTS levels, there are two LAU (Local Administrative Units) levels in Slovenia: the LAU 1 level defines 58 administrative units; and for the LAU 2 level 212 municipalities are defined in 2016.

[4] The Intramax procedure is implemented in Flowmap, which is a software package dedicated to analysing and displaying interaction or flow data developed at the Faculty of Geosciences of the Utrecht University in the Netherlands (De Jong and Van der Vaart, 2013).

simply "step"), two BDUs/FRs[5], whose interaction gives the highest value of the objective function, are grouped together, and the interaction between them becomes the internal (or intrazonal) interaction for the resulting FR. This new FR now takes the place of the two parent BDUs/FRs in the next step of the analysis. Thus, with N basic data units, all BDUs are grouped together into one FR after N − 1 steps and all interactions become intrazonal (Brown and Pitfield, 1990; Masser and Brown, 1975; Masser and Scheurwater, 1980; Nel et al., 2008). The procedure, as well as the results of the hierarchical aggregation, can be presented in a tree structure of a dendrogram.

In their first paper on the Intramax procedure, Masser and Brown (1975) specified the objective function as the difference between the observed values, t_{ij}, and expected values, t^*_{ij}:

$$F_{ij} = (t_{ij} - t^*_{ij}) + (t_{ji} - t^*_{ji}) \qquad \max_{i \neq j} F_{ij} \tag{1}$$

where expected value t^*_{ij} was calculated similarly to the expected frequency in the i-th row and the j-th column in a contingency table for the Chi-square test using the sum of the i-th row, $r_i = \sum_j t_{ij}$, the sum of the j-th column, $c_j = \sum_i t_{ij}$, and the sum of all elements of matrix T, $t = \sum_{i,j} t_{ij}$:

$$t^*_{ij} = \frac{r_i c_j}{t} \tag{2}$$

Note that when the objective function is defined as (1), matrix T should be standardised by dividing T by t, so that $T' = [t'_{ij}]$, where $\sum_{i,j} t'_{ij} = 1$[6]. Masser and Brown (1975) applied the contiguity constraint, γ_{ij}, to restrict the search only for contiguous BDUs/FRs: $\gamma_{ij} = 1$ when BDU_i / FR_i and BDU_j / FR_j are contiguous, and $\gamma_{ij} = 0$ otherwise. They analysed the use of the objective function (1) using the commute-to-work data for Merseyside, England, and noted that, in practise, little or no difference may be expected in the results obtained with or without contiguity constraints and that the procedure would tend to favour small BDUs/FRs (in the first place, small BDUs/FRs are aggregated, then large ones).

Following Tyree's findings (1973), Hirst (1977) noted that Masser and Brown (1975) took insufficient account of the influence that the row totals, r_i, and column totals, c_j, in $T = [t_{ij}]$ had on the residual values, $t_{ij} - t^*_{ij}$, that appeared in the objective function (1). The difference between the observed and expected values will tend to increase for cells (interactions) in those rows and columns with large sums (large outflows from BDUs/FRs and large inflows to BDUs/FRs). Since the objective function is recalculated after each step in the grouping procedure, this bias will be cumulative. Using the methodological solutions for arbitrary origin-destination distribution as defined by Goodman (1963), Hirst (1977) suggested that the effects of unequal marginal distribution could be overcome by reformulating the objective function in the revised version by dividing the difference $t_{ij} - t^*_{ij}$ by an expected value t^*_{ij}, with t^*_{ij} corrected for $t_{ij} = 0$ in T. Consequently Masser and Brown revised the objective function (Masser and Brown, 1977) as follows:

$$F_{ij} = \frac{(t_{ij} - t^*_{ij})}{t^*_{ij}} + \frac{(t_{ji} - t^*_{ji})}{t^*_{ji}} \tag{3}$$

The entries in the objective function (3) are not necessarily standardised. Hirst (1997) tested function (3) using hypothetical data. He noted that using the objective function (3) in the Intramax procedure would still tend to favour groupings of smaller BDUs/FRs before larger ones (because of the differences between the values obtained for small opposed to large BDUs/FRs).

Masser and Scheurwater (1978, 1980) first applied the revised version of the objective function (3) to real data, specifically the migration data on four contiguous zones within a larger dataset in Greater London, the migration data on forty regions from the Netherlands, and the labour commuting data for Merseyside that were used in the earlier paper on Intramax analysis (Masser and Brown, 1975). They concluded:

a. that the Intramax procedure *"explicitly identifies regions that have more (direct) interaction with each other than with other areas at each stage of the grouping process"* (Masser and Scheurwater, 1980: 1361);
b. that *"the dendrogram obtained by the Intramax procedure has a well-developed tree structure in which basic data units combine to form broadly similar-sized clusters within the grouping process"* (Masser and Scheurwater, 1978: 161);
c. that *"stronger connections would appear between pairs of smaller zones containing a relatively low proportion of intrazonal interaction than between pairs of larger zones containing a relatively high proportion of intrazonal interaction, and that all other things being equal, the former would tend to fuse together before the latter"* (Masser and Scheurwater, 1980: 1380);
d. that *"the bias noted by Hirst (1977), far from being a disadvantage, is in fact a positive advantage in that it is a reflection of the inherent characteristics of the structure of spatial interaction in the matrix"* (Masser and Scheurwater, 1980: 1380); and
e. that the Intramax procedure might be readily applied to large data sets and might be adapted easily to deal with large, sparse matrices (Masser and Brown, 1977; Masser and Scheurwater, 1980).

Brown and Pitfield (1990) noted that the part that was subtracted in (3) was a constant and might thus be ignored. So, the objective function (3) can be re-expressed more simply as follows (Brown and Pitfield, 1990):

$$F_{ij} = \frac{t_{ij}}{t^*_{ij}} + \frac{t_{ji}}{t^*_{ji}} \tag{4}$$

The objective function (4) is referred to as an original objective function of the Intramax procedure. The entries in (4) are not necessarily standardised.

Some recent methodological studies on using Intramax for functional regionalisation are discussed in the following

[5] Notation on BDUs/FRs: In the aggregation procedure, two BDUs, which are fused first, form a first FR. Thus, in the second step of the aggregation, two BDUs or one BDU and one FR can be fused; in the third step, two BDUs or one BDU and one FR or two FRs can be fused; and so on, until all BDUs are aggregated into FRs. From this step, from which no singleton region exists, small FRs are aggregated into larger ones.

[6] The results of the objective function (1) are dependent on the standardisation, because a proportional change of elements in the interaction matrix T would not result in the same results as T'.

studies. Alvanides et al. (2000) analysed a set of systems of hierarchical FRs generated by the Intramax procedure and a set of systems of FRs modelled by ZDeSi (zone design system for interaction data; Openshaw and Rao, 1995; Openshaw and Alvanides, 1999) for 402 local authority districts in England and Wales using 1991 census data for labour commuting. The comparison of intra-regional flows showed the superiority of the ZDeSi model compared to the Intramax approach: the proportions of the intra-regional flows were higher for ZDeSi for all systems of FRs (more for systems of small FRs and less for systems of large FRs). It was reported also that it was evident from the results that *"the Intramax procedure gets trapped in local optima, producing low scores and fragmented regions"* (Alvanides et al., 2000: 127).

Watts (2009) evaluated hierarchical versus rule-based techniques of modelling FRs, namely the Intramax technique and the modified version of Coombes' updated algorithm (Coombes and Bond, 2008; Coombes, 2010), using the principles of fuzzy set theory (Feng, 2009), to explore the local properties of the two solutions. The application was carried out for 1,365 Australian statistical local areas and labour commuting data. He reported that both approaches to grouping had strong, but not robust, local optimisation properties (Watts, 2009). The robustness was analysed in relation to self-containment as defined by Smart (1974) and Van der Laan and Schalke (2001). Watts reported that *"the low minimum rate of closure [of the Intramax method] underlines the point, however, that the grouping criterion under Intramax is quite different than those criteria characterising the Coombes algorithm. The number of groups in the final solution is more readily controlled under the Intramax technique, since the stopping rule can easily be adjusted to achieve a given form of final solution, which could be based on a minimum closure requirement, rather than the aggregate intrazonal flow. There may be limits as to the desirable rates of closure across groups. Otherwise convergence may be achieved with a singleton group, unless there is genuine geographical separation of the labour markets"* (Watts, 2009: 525). He also noted the superiority of the Intramax method regarding the singleton regions.

Landre and Håkansson (2013) compared the performance of the Intramax procedure to the rule-based procedure used by Statistics Sweden's (2010) Local Labour Market Areas (LLMAs). The application was performed for 289 municipalities in Sweden. They reported that *"although the procedure used for the construction of LLMAs differs considerably from that of Intramax regions, the results obtained are quite similar for most of the country … Despite many similarities, the two methods differ fundamentally with regard to self-containment levels and the construction of regions in metropolitan areas. In the latter, Intramax analysis results in a fragmented pattern with unacceptable low levels of self-containment in a number of regions. However, LLMAs are clearly too large there"* (Landre and Håkansson, 2013: 15). The same differences for the urban areas, where Intramax gives more fragmented FRs, have been reported by Feldman et al. (2005) for Scotland, and by Mitchell et al. (2007) for Australia. Landre and HÍkansson concluded that *"both methods could benefit from additional controls in their procedures, especially when applied in situations where differences in land area are large… for Intramax analysis, it could be the application of self-containment constraints resulting in the amalgamation of regions if these constraints are not met"* (Landre and Håkansson, 2013: 15).

Koo (2012) suggested the use of a new objective function in the Intramax procedure that focused on the proportion of intra-regional flows,

$$F_{ij} = \frac{t_{ij}}{t_{ii}} + \frac{t_{ji}}{t_{jj}}, \qquad (5)$$

simultaneously with the use of a contiguity constraint and an area-balanced constraint. The algorithm was applied for the Seoul Metropolitan Area's data on labour commuting for a total of 1,180 BDUs. He reported that the *"algorithm has derived a set of improved functional regions that better serves the objective of the regionalisation which maximises the proportion of internal flows more compared to Intramax"* (Koo, 2012: 33). He noted that the constrained models gave better results than unconstrained ones with respect to the percentage of intra-regional flows. Koo also re-confirmed that the original Intramax model gave fragmented FRs in large urban areas.

Recently, Drobne and Lakner (2015, 2016) evaluated the simultaneous use of three different constraints in the Intramax procedure, namely: (a) the contiguity constraint; (b) the higher-proportion-of-intra-regional-flows constraint, which ensures that those regions are grouped together that gave a higher proportion of the intra-regional (inner) flows; and (c) the lower-coefficient-of-variation-of-intra-regional-flows constraint, which ensures that a grouping of BDUs/FRs gives FRs with a similar (proportion) of intra-regional-flows. They noted that, when using data on labour commuting, there is no need to include the contiguity constraint in the procedure. They reported as well that the use of the higher-proportion-of-intra-regional-flows constraint generates singleton regions, and that the lower-coefficient-of-variation-of-intra-regional-flows constraint forces the biggest BDU, as an isolated FR, up to a relatively high level of aggregation. They concluded that the Intramax procedure generates fragmented large urban areas, but the lower-coefficient-of-variation-of-intra-regional-flows constraint even more strictly delineates the metropolitan area into fragmented pieces.

3. Methodology

The Intramax procedure is a stepwise (hierarchical aggregation) procedure. In each step, the two BDUs/FRs which the interaction realises the highest value of the objective function, are grouped together. In this paper, we analysed the performance when using different objective functions in the aggregation procedure. The performance was measured by known self-containment indicators, as well as by the self-containment statistics suggested in this paper. Selected systems of hierarchical FRs were compared to official regions at NUTS 2 and NUTS 3 levels, as well as to the administrative units (AUs) in Slovenia.

The application was done for the inter-municipal labour commuting flows in 2011 in Slovenia. The initial dimension of the interaction matrix T was $N^2 = 210^2$. Of a total of 44,100 cells in the matrix, there were 31,557 (71.56%) empty cells. In 2011, there were 778,776 labour commuters in total, but only 388,376 (49.87%) of them commuted between municipalities. The rest (390,400; 50.13%) formed intra-municipal flows. The maximum inflow of 109,884 (28.29%) labour commuters terminated in the capital Ljubljana, which is the biggest employment centre of Slovenia, while the outflow from Ljubljana was 16,027 (4.13%) of labour commuters.

In the hierarchical aggregation procedure, we analysed the use of seven objective functions (see also Tab. 1): (F_1) the original Intramax objective function, as defined by Brown and Pitfield (1990); (F_2) the sum-of-flows objective function, which is the first logical alternative to the original objective function; it is also the only function analysed here that considers absolute values rather than proportions of flows; (F_3) the sum-of-proportions-to-intra-regional-flows objective function, which is the revised version of the objective function suggested by Koo (2012), where only proportions of flows in the origins have been suggested, but, here, we suggest using the sum of all proportions of flows, namely in origins and in destinations; (F_4) the first version of the CURDS weighted interaction index[7], used for analysing Travel-To-Work-Areas (TTWA) in Great Britain (Coombes et al., 1982), which is a sum-of-proportions-to-out-/inflows objective function; (F_5) Smart's weighted interaction index, which has been suggested on the ratio of gravity formula (Smart, 1974)[8]; (F_6) the second version of the CURDS weighted interaction index, as defined by Coombes et al. (1986) and ONS and Coombes (1998)[9]; and (F_7) Tolbert and Killian's interaction index (Tolbert and Killian, 1987; Killian and Tolbert, 1993), which aggregates two BDUs/FRs only if both flows of interaction are large[10]. Standardisation of the entries is not needed for any of the objective functions analysed here. For all functions (F_1–F_7), we assume that all $t_{ii} > 0$.

To analyse the performance of objective functions F_1 to F_7 and the use of the contiguity constraint in the Intramax procedure, we developed a programme code in Mathematica 10.3. To check the validity of contiguity,

Id	Objective function	Description	Eq.
F_1	$F_{1ij} = \frac{t_{ij}}{r_i c_j} + \frac{t_{ji}}{r_j c_i} = \frac{t_{ij}}{t^*_{ij}} + \frac{t_{ji}}{t^*_{ji}}$	Original Intramax objective function (Masser and Brown, 1977; Brown and Pitfield, 1990)	(6)
F_2	$F_{2ij} = t_{ij} + t_{ji}$	Sum-of-flows objective function	(7)
F_3	$F_{3ij} = \frac{t_{ij}}{t_{ii}} + \frac{t_{ij}}{t_{jj}} + \frac{t_{ji}}{t_{jj}} + \frac{t_{ji}}{t_{ii}}$	Sum-of-proportions-to-intra-regional-flows objective function (compare to Koo, 2012)	(8)
F_4	$F_{4ij} = \frac{t_{ij}}{r_i} + \frac{t_{ij}}{c_j} + \frac{t_{ji}}{r_j} + \frac{t_{ji}}{c_i}$	First version of CURDS weighted interaction index (Coombes et al., 1982)	(9)
F_5	$F_{5ij} = \frac{t_{ij}}{t_{ii}}\frac{t_{ij}}{t_{jj}} + \frac{t_{ji}}{t_{jj}}\frac{t_{ji}}{t_{ii}} = \frac{t^2_{ij} + t^2_{ji}}{t_{ii} t_{jj}}$	Smart's (1974) weighted interaction index	(10)
F_6	$F_{6ij} = \frac{t_{ij}}{r_i}\frac{t_{ij}}{c_j} + \frac{t_{ji}}{r_j}\frac{t_{ji}}{c_i} = \frac{t^2_{ij}}{r_i c_j} + \frac{t^2_{ji}}{r_j c_i}$	Second version of CURDS weighted interaction index (Coombes et al., 1986)	(11)
F_7	$F_{7ij} = \frac{t_{ij} + t_{ji}}{Min\{r_i, r_j\}}$	Tolbert & Killian's (1987) interaction index	(12)

Tab. 1: Objective functions analysed in the hierarchical aggregation procedure
Source: As noted in the table and authors' elaboration

---

[7] CURDS: Centre for Urban & Regional Development Studies in Newcastle University.

[8] Smart noted (1974: 270): "*Where an area has a large absolute movement of workers to another area, this is represented in the numerator of the fraction by its square, ensuring that strong central pulls are not 'overlooked'. They are, however, balanced by the effect of the size of different areas on the denominator of the fraction, which acts as 'antibody' in the system preventing the emergence of the inflated labour markets [i.e. FRs]*".

[9] Coombes and Bond (2008: 234) wrote that "*considerable experimentation has led to the choice of the formula to determine in which way a zone should be grouped to maximise the likelihood that the resulting TTWA definitions most closely meet their objectives. The key need in practice is to enable smaller places near major centres to consolidate as separable TTWAs (where commuting flows justify this) because otherwise the TTWAs that include major centres expand remorselessly to engulf all surrounding areas, with the result that the set of defined TTWAs is less numerous than the maximum possible which meet the set criteria*".

[10] As Tolbert and Killian (1987: 16) reported, "*the numerator reflects a concern for the total number of commuters between two counties [BDUs] (regardless of direction) and as such, provides a measure of the degree of interconnectedness between them. The denominator expresses the volume of shared commuters on a relative rather than absolute basis, thus ensuring that the analysis is not dominated by large counties. The resident labour force … is used as the base because it is not sensitive to the direction of commuting and because it is constant across all versions of frequency matrices*". They emphasised the importance of using an asymmetric relationship: "*The smaller county's resident labour force in the denominator ensures that even highly asymmetric commuting patterns are considered evidence of a strong labour market tie. Thus, while commuters from a small county may make up only a minor portion of a large county's labour force, those same commuters can represent a very substantial proportion of the smaller county's resident labour force. … The counties clearly depend on each other in ways that have important social, economic, and political implications. Use of the large county's resident labour force or the sum of the two counties' labour forces would average the asymmetric relationship and reduce the apparent relationship between the two counties*" (ibid.).

we translated spatial contiguity into a network tree-generation problem. In this way, regions and their adjacency relationships are expressed as nodes and edges in a graph, so that a region is verified as contiguous only if there is at least one path connecting all the spatial units within the region or if all the spatial units within the region are connected to the tree structure (Kim et al., 2015). In our programme code, contiguity is checked by the depth-first search (DFS) algorithm (Daras, 2005).

Seven sets of systems of 2–209 FRs, generated in the hierarchical aggregation procedure using objective functions F_1–F_7, were evaluated at each stage of the aggregation procedure by self-containment indicators as suggested by Goodman (1970) and Smart (1974) and used very often in the literature (e.g. Casado-Diaz, 2000; Casado-Diaz and Coombes, 2011; Landré and Håkansson, 2013; Van der Laan in Schalke, 2001; Watts, 2009); namely, the proportion of intra-regional flows, supply-side self-containment (SSSC), and demand-side self-containment (DSSC)[11]. The proportion of intra-regional (inner) flows was calculated as

$$t_{ii}^{\%} = \frac{1}{t}\sum_i t_{ii}\,; \tag{13}$$

and supply-side self-containment (SSSC) and demand-side self-containment (DSSC) were calculated as follows:

$$SSSC_i = \frac{t_{ii}}{r_i}, \tag{14}$$

$$DSSC_i = \frac{t_{ii}}{c_i}. \tag{15}$$

The basic property of the automatic hierarchical aggregation procedure should be the inclusion of all BDUs into FRs – without leaving any singleton regions. For this reason, we analysed the performance of objective functions also by the proportion of singleton regions, which was calculated as

$$n_{si}^{\%} = \frac{n_{si}}{n}, \tag{16}$$

where n_{si} is the number of singletons, and n is the total number of FRs (dimension of matrix T) at each step of aggregation.

Properties of different systems of FRs can be compared by absolute values like minimum, maximum, mean, and standard deviation of self-containment indicators. But, when comparing the homogeneity of FRs, relative statistics are the only valid approach to comparing the variation of self-containment of FRs[12]. In our study, we analysed the performance of using different objective functions in the hierarchical aggregation procedure to generate homogeneous FRs by relative self-containment indicators; namely by the coefficient of variation of SSSC and by the coefficient of variation of DSSC:

$$CV_{SSSC} = \frac{\mu_{SSSC}}{\sigma_{SSSC}}, \tag{17}$$

$$CV_{DSSC} = \frac{\mu_{DSSC}}{\sigma_{DSSC}}, \tag{18}$$

where $\mu_{SSSC} = \frac{1}{n}\sum_i SSSC_i,$, $\sigma_{SSSC} = \sqrt{\frac{1}{n}\sum_i (SSSC_i - \mu_{SSSC})^2}$,

and $\mu_{DSSC} = \frac{1}{n}\sum_i DSSC_i$, $\sigma_{DSSC} = \sqrt{\frac{1}{n}\sum_i (DSSC_i - \mu_{DSSC})^2}$.

We evaluated also the performance of objective functions F_1–F_7 to aggregate contiguous BDUs/FRs with the highest interactions when analysing labour commuting flows. As it has been shown several times (e.g. Brown and Pitfield, 1990; Drobne and Lakner, 2015, 2016; Feldman et al., 2005; Masser and Scheurwater, 1980), the inclusion of the contiguity constraint is, in general, not needed when analysing labour commuting flows in the original Intramax procedure. However, it should be considered when using a different objective function rather than the original one (Koo, 2012), in a combination with other constraints (Drobne and Lakner, 2015, 2016), or when analysing some other data (e.g. financial flows; Kohl and Brouver, 2014). When the contiguity constraint is used, two BDUs/FRs that give the maximal value of the analysed objective function are aggregated only if they are contiguous ($\gamma_{ij} = 1$). On the other hand, if the maximal value of the objective function is defined by two non-contiguous BDUs/FRs ($\gamma_{ij} = 0$), the contiguity constraint is used to seek the first contiguous BDUs/FRs. In this way, the constraint forces to aggregate two BDUs/FRs that do not provide the maximum value of the particular objective function. The sum of deviation from the maximum value of the objective function was measured by the sum of steps of seeking two contiguous BDUs/FRs (hereinafter "contiguity seeking steps", CSSs). The sum of deviation from the maximum value of the objective function, because of the contiguity constraint, measures the quality of the objective function to aggregate contiguous BDUs/FRs at a given interaction matrix. The inclusion of the contiguity constraint in the hierarchical aggregation procedure for a particular objective function is reported as $F_1^{(\gamma)}, F_2^{(\gamma)} \ldots F_7^{(\gamma)}$.

Following the basic objectives of the functional regionalisation by means of the hierarchical aggregation procedure, the most suitable systems of hierarchical FRs are defined by a relatively higher proportion of intra-regional flows ($t_{ii}^{\%}$), by a relatively lower proportion of singleton regions ($n_{si}^{\%}$), by a lower coefficient of variation of supply-side self-containment (CV_{SSSC}), by a lower coefficient of variation of demand-side self-containment (CV_{DSSC}), and by a geographically valid spatial extent[13]. We evaluated efficiency of using different objective functions in the hierarchical aggregation procedure by ranks of $t_{ii}^{\%}$, $n_{si}^{\%}$, CV_{SSSC}, CV_{DSSC}, and CSS at each stage of the aggregation procedure. The general efficiency for each objective function is calculated as a mean of ranks. The performance of the analysed objective functions was evaluated also by dendrograms

[11] In the analysis of labour commuting (journey-to-work) flows, supply-side self-containment (SSSC) is also called workplace-based self-containment (Goodman, 1970; Smart, 1974) or employment self-containment (Van der Laan and Schalke, 2001); similarly, demand-side self-containment (DSSC) is also called residence-based self-containment (Goodman, 1970; Smart, 1974) or housing self-containment (Van der Laan and Schalke, 2001).

[12] Systems of FRs modelled by different methods represent different populations. When comparing the variation of self-containment of FRs, only relative values should be used. However, there are many reports on standard deviation (absolute value) of self-containment in the literature (e.g. Casado-Diaz, 2000; Landre and Håkansson, 2013; Watts, 2009), but, unfortunately, there is no notion on relative statistics (e.g. coefficient of variation).

[13] Regarding geography, FRs of most of the interactions mentioned here (particularly of commuting and migration) should be geographically compact; however, there are some exceptions, like trade FRs.

and by animations of the hierarchical aggregation of BDUs into FRs, which were generated by our programme code in Mathematica 10.3.

The properties of FRs modelled in the original Intramax procedure were then compared to the current regions at NUTS 2 and NUTS 3 levels and administrative units in Slovenia. In Slovenia, there are 2 regions for the application of regional policies at NUTS 2 level also called "macro regions" or "cohesion regions", and there are 12 "statistical regions" at NUTS 3 level also called "development regions". Below the NUTS 3 level in Slovenia, there is the LAU 1 level where 58 "administrative units" are defined.

4. Results and discussion

First, we represent and discuss the results of the hierarchical aggregation procedure regarding different objective functions without the use of the contiguity constraint. Fig. 1 shows the results with respect to intra-regional flows, singleton regions, and the homogeneity of supply-side and demand-side self-containment of the hierarchical FRs. Comparing the performance of objective functions regarding the proportion of intra-regional flows (Fig. 1a), the superiority of F_2 and the inferiority of F_1 are obvious. From other objective functions, good results were obtained by the use of F_7 and F_4 for the regionalisation at the beginning of the aggregation procedure, but for the use of F_6 good results were found only at the end of the procedure.

Regarding the number of singleton regions, the results of using F_1 and F_2 are opposite of that for intra-regional flows (compare Fig. 1a and Fig. 1b). The most effective objective function at all aggregation levels that aggregates singletons is original Intramax function F_1. It starts to aggregate small BDUs first and leaves the most important destination BDU in Slovenia, i.e. the capital of Ljubljana, as a singleton region up to the last 10% of the aggregation steps (ASs). But, F_2 chains neighbouring BDUs to Ljubljana. So, the number (proportion) of remaining singletons is the highest for all stages of aggregation when using the sum-of-flows function F_2. Among other objective functions, at the beginning of the aggregation procedure, F_5 and F_6 reduce the number of singletons fast, but they miss some of them for the rest of the steps. F_3 and F_4 produce very similar results, aggregating singletons much earlier than F_7. Among the functions that solve the problem of aggregating small BDUs (and not leaving singletons), F_1 performs the best, as it aggregates the last singletons in the 190th AS; the next-best functions are F_3 and F_4, which aggregate the last singletons in the 191st and 192nd AS, respectively, whereas F_7 aggregates them in the 196th AS. However, F_1 performs quite differently from the other functions, as it leaves the bigger BDUs as separate FRs to compete with other FRs, whereas F_3, F_4, and F_7 produce singletons that cannot compete with other FRs at each stage of the aggregation procedure (in our case, singleton regions were different borderline municipalities whose inferiority was evident at each step before they were aggregated into FRs).

The hierarchical aggregation procedure should produce systems of homogeneously self-contained FRs with the smallest possible variation of supply-side and demand-side self-containment. Here, F_3 and F_4 give the best results for almost the whole procedure of the aggregation. F_1 is better only for the last steps of the aggregation, when modelling just 2–5 FRs. F_5 is also competitive, but only at the beginning and at the end of the aggregation procedure. The most dissimilar FRs regarding the self-containment of flows were obtained using F_2 and F_6. Comparison of homogeneity of self-containment of FRs generated by different objective functions shows also that, among the objective functions analysed here,

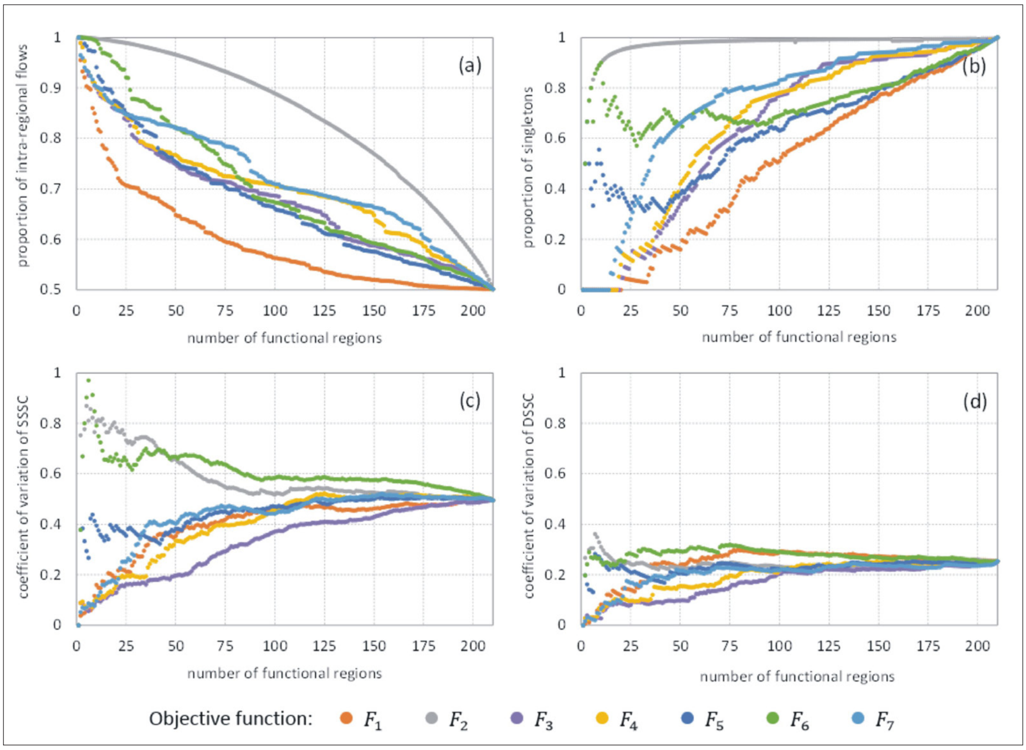


Fig. 1: (a) Proportion of intra-regional flows, (b) proportion of singleton regions, (c) coefficient of variation of supply-side self-containment, and (d) coefficient of variation of demand-side self-containment, hierarchical aggregation of municipalities regarding the labour commuting flows without the continuity constraint (Slovenia, 2011)
Source: authors´ calculations

the original Intramax objective function performs better, in general, regarding the supply-side self-containment than regarding the demand-side self-containment.

The performance of using the individual objective function in the hierarchical aggregation procedure can also be expressed by the capability to aggregate contiguous BDUs at a given interaction matrix. Such a capability of the analysed objective functions for data on labour commuting between Slovene municipalities in 2011 was measured by the sum of deviations from the maximum value of the objective function, namely, contiguity seeking steps, CSSs; see Fig. 2 where frequency distribution of CCSs is represented. Here, $F_2^{(\gamma)}$, which forced almost 490 CCSs in total, was the least effective objective function among all the analysed objective functions. It was followed by $F_3^{(\gamma)}$ with more than 150 CSSs, by $F_7^{(\gamma)}$ with more than 50 CSSs, and by $F_5^{(\gamma)}$ with 12 CSSs. Objective functions $F_1^{(\gamma)}$, $F_4^{(\gamma)}$, and $F_6^{(\gamma)}$ are very effective objective functions while they aggregated mostly contiguous BDUs/FRs: $F_4^{(\gamma)}$ forced only 3 CSSs, $F_6^{(\gamma)}$ only 2 CSSs, and $F_1^{(\gamma)}$ only 1 CSS.

From the aforementioned results of the comparative analysis of using vs. omitting the contiguity constraint and from Fig. 2, we summarise that the use of the contiguity constraint is not needed when modelling bigger labour commuting FRs (in our case 2 to 20 FRs) – this is valid for the use of all analysed objective functions. Especially when using the objective function that considers variations in all rows and columns in the interaction matrix – like the original Intramax function (F_1) or the first (F_4) and the second CURDS's weighted interaction function (F_6) – the results of modelling FRs are similar when using or omitting the contiguity constraint.

The evaluation of using different objective functions in the hierarchical aggregation procedure by ranks of intra-regional flows, proportion of singleton regions, coefficient of variation of supply-side self-containment, coefficient of variation of demand-side self-containment, and contiguity seeking steps that were calculated at each aggregation stage showed that, in general, F_3, F_4, F_5, and F_7 generate statistically better results than original Intramax objective function F_1; see Fig. 3. According to the general evaluation of statistical results, F_2 and F_6 are the only objective functions whose efficiency is statistically worse than that of using F_1. However, geographic and operational evaluation of systems of hierarchical FRs by dendrograms and by animations of the

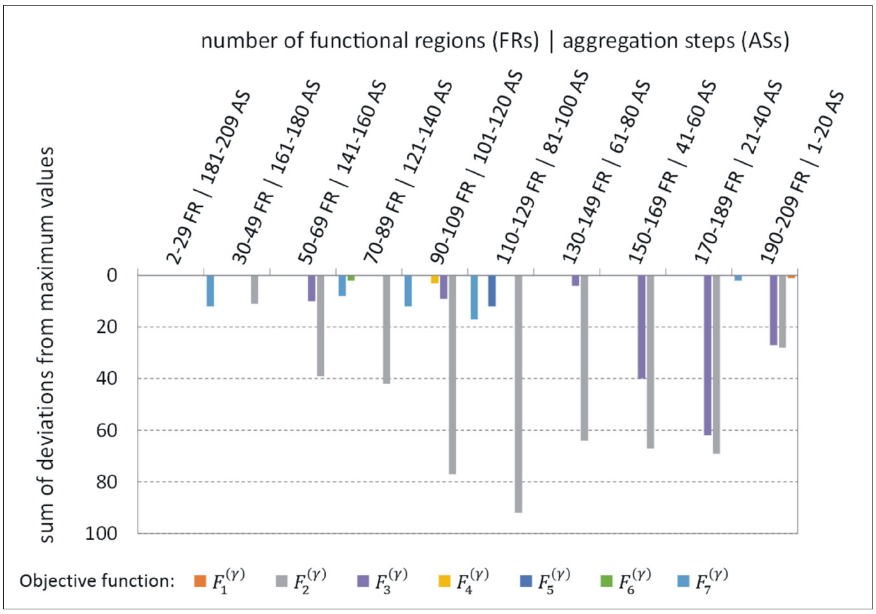


Fig. 2: Sum of deviations from maximum values of the objective function because of the contiguity constraint (Slovenia, 2011). Source: authors´ calculations

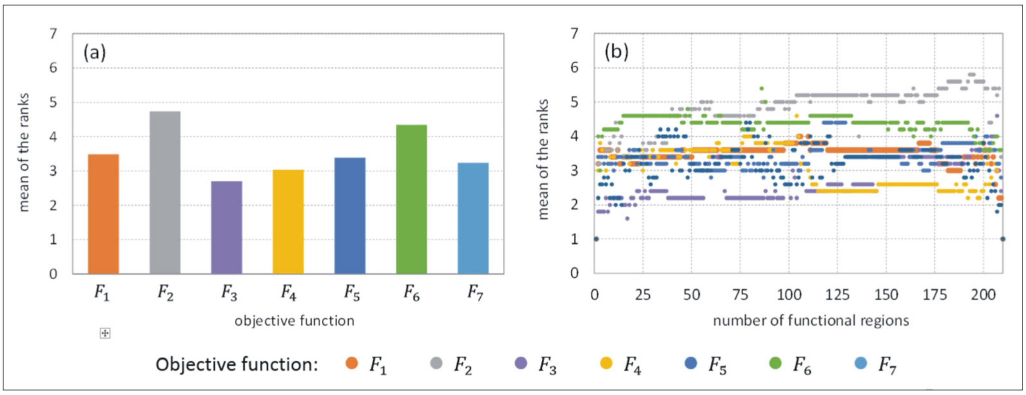


Fig. 3: General evaluation of sets of hierarchical functional regions modelled using objective functions F_1–F_7 by ranks of the analysed indicators: (a) mean of the ranks in the set of systems of functional regions, (b) mean of the ranks by systems of functional regions, hierarchical aggregation of municipalities regarding the labour commuting flows without the use of the contiguity constraint (Slovenia, 2011). Source: authors´ calculations
Note: A lower value indicates better ranking

aggregation of municipalities showed that the most suitable FRs were generated using the original Intramax procedure. Besides the original Intramax objective function, Smart's weighted interaction index models also persuasive labour commuting FRs (if singletons are corrected manually).

The efficiency of FRs modelled by the original Intramax procedure was analysed also by comparison to the delimitation of Slovenian territory into 2 "cohesion regions" at the NUTS 2 level (see Fig. 4), into 12 "statistical regions" at NUTS 3 level (see Fig. 5), and into 58 administrative units of Slovenia at LAU 1 level (see Fig. 6). While the cohesion regions have existed only since 2008, the first version of statistical regions dates back to the mid-1970s. At that time, statistical regions were established for the purpose of regional planning and cooperation in various sectors. The first regionalisation of statistical regions was supported by exhaustive gravity analysis of labour markets, education areas, and supply markets in twelve regional, and their sub-regional, centres. Up to 2011, statistical regions were fine-tuned several times – that is the reason why today's

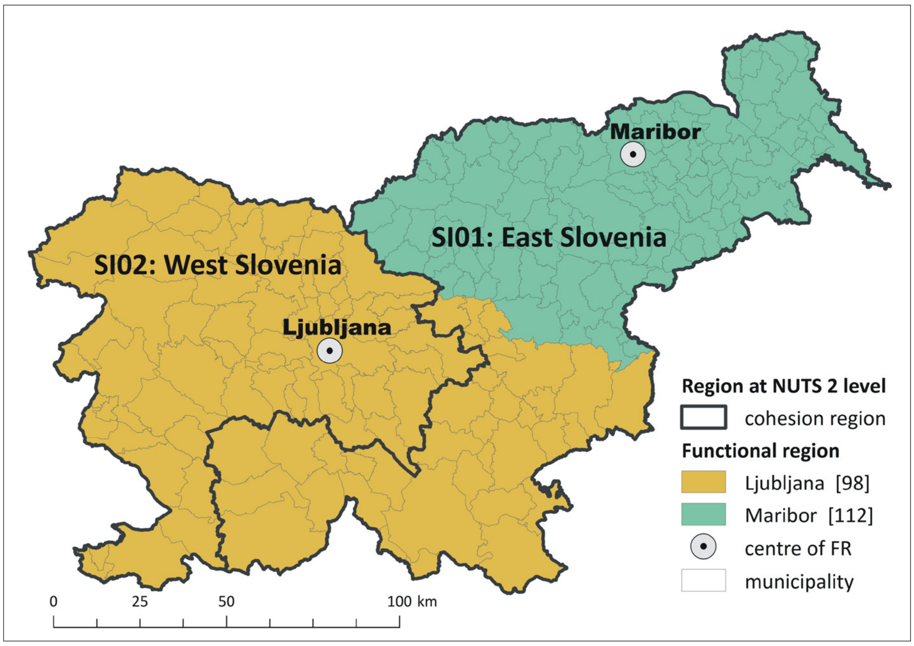


Fig. 4: Two regions at NUTS 2 level and two functional regions modelled by original Intramax procedure (Slovenia, 2011). Source: authors´ calculations
Notes: The number of the municipalities in the FR is given in square bracket. In 2011, there were 210 municipalities in Slovenia.

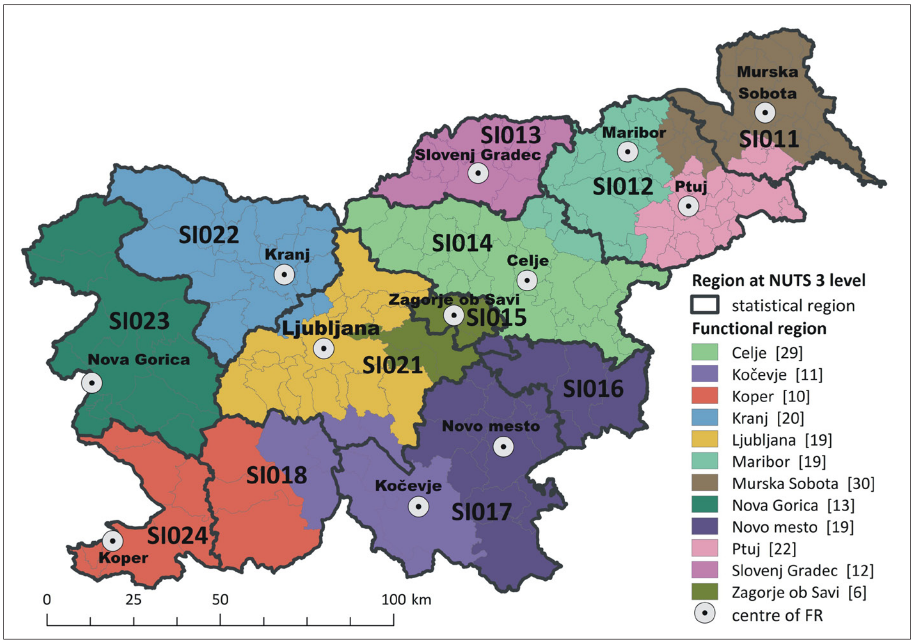


Fig. 5: 12 statistical regions at NUTS 3 level and 12 functional regions modelled by original Intramax procedure (Slovenia, 2011). Source: authors´ calculations
Notes: Regions at NUTS 3 level are: Pomurska (SI011), Podravska (SI012), Koroška (SI013), Savinjska (SI014), Zasavska (SI015), Posavska (SI016), South-East Slovenia (SI017), Primorsko-notranjska (SI018), Central Slovenia (SI021), Gorenjska (SI022), Goriška (SI023), and Obalno-kraška (SI024). The number of the municipalities in the FR is given in square brackets. In 2011, there were 210 municipalities in Slovenia

Slovenian regions at NUTS 3 level are very stable (Drobne and Bogataj, 2012a; SORS, 2016). Some 58 administrative units were set up in 1991 to optimise administrative tasks between citizens and state. They were delimitated on the base of 62 old municipalities that were transformed into much smaller ones (147) in 1994. From that time, municipalities in Slovenia were changed several times to the 212 municipalities existing in 2016.

The comparison of FRs to official regions and administrative units showed that 12 statistical regions and 58 administrative units, which were delimitated on the long-term bases of functional interactions in Slovenian territory, demonstrated higher self-containment than the corresponding number of FRs. From Tab. 2, it is evident that the proportion of intra-regional flows at the state level is higher, as well as the homogeneity regarding intra-regional flows, supply-side and demand-side self-containment for statistical regions and administrative units, than for FRs modelled using the original Intramax procedure.

But, both cohesion regions at NUTS 2 level, which were established for the application of regional policies, demonstrate lower self-containment than FRs. Among others, the reason for that is rooted in Slovenian tradition. In older European Union member countries, administrative regions are the units in which regional economic policies are designed and executed and the members of regional

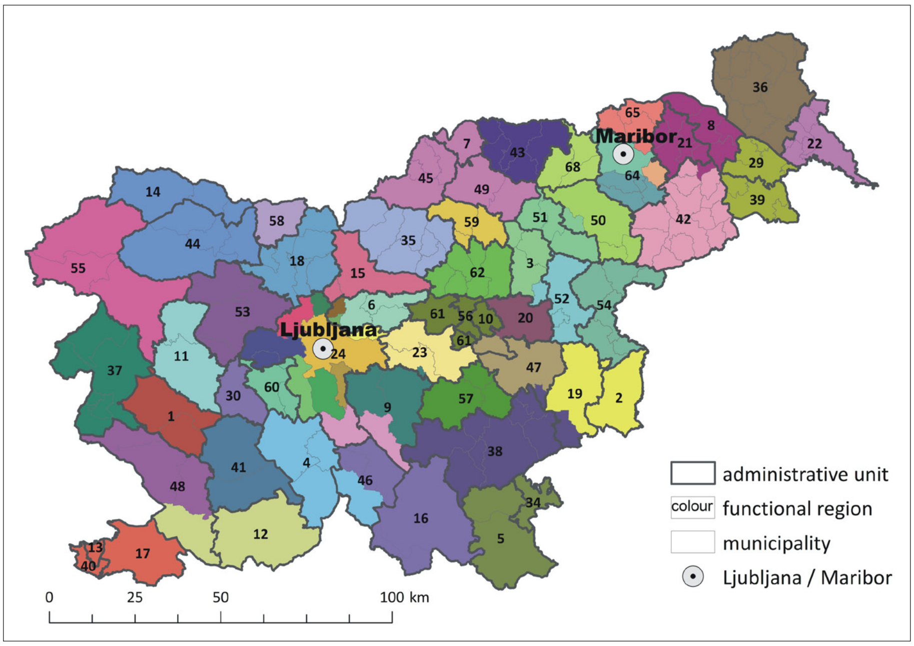


Fig. 6: 58 administrative units at LAU 1 level and 58 functional regions modelled by original Intramax procedure (Slovenia, 2011). Source: authors´ calculations
Notes: The maps indicates the codes for administrative units. The list of 58 administrative units is available at http://www.upravneenote.gov.si/. In 2011, there were 210 municipalities in Slovenia

	2 NUTS 2 (cohesion) regions	2 FRs	12 NUTS 3 (statistical) regions	12 FRs	58 AUs	58 FRs
T_{ii} or *SSSC* or *DSSC* at state level	92.0%	95.5%	83.7%	81.3%	64.5%	63.4%
$CV_{T_{ii}}$	0.053	0.038	0.124	0.147	0.240	0.388
$Min_{T_{ii}}$	87.2%	90.9%	59.5%	51.5%	29.9%	17.0%
$Max_{T_{ii}}$	97.0%	98.1%	92.7%	91.9%	86.9%	85.4%
CV_{SSSC}	0.053	0.038	0.124	0.147	0.240	0.388
Min_{SSSC}	87.2%	90.9%	59.5%	51.5%	29.9%	17.0%
Max_{SSSC}	97.0%	98.1%	92.7%	91.9%	86.9%	85.4%
CV_{DSSC}	0.047	0.006	0.057	0.075	0.140	0.254
Min_{DSSC}	88.0%	95.1%	73.9%	68.6%	45.1%	6.7%
Max_{DSSC}	96.7%	96.3%	94.0%	94.0%	89.7%	89.7%

Tab. 2: Comparison of functional regions in 2011 to the delimitation of the Slovenian territory at NUTS 2 and NUTS 3 levels and to administrative units (Slovenia, 2011). Source: authors' elaboration
Notes: FR denotes functional region, T_{ii} denotes intra-regional flow, SSSC is supply-side self-containment, DSSC is demand-side self-containment, CV is coefficient of variation, Min denotes minimum, and Max denotes maximum

governments and regional parliaments find their constituencies. Regional economic policy thus primarily reflects the administrative regions' interests; therefore they are the proper starting points for most purposes of regional economic policy analysis. However, there was no "middle layer" of government between the central state and the municipal level in most new member states of the European Union – so, there was no such tradition. This was also the case in Slovenia (Drobne et al., 2009).

The shortcoming of the Intramax procedure to model too many FRs in a metropolitan area can be observed in Fig. 6, where the metropolitan areas of the two most important employment centres in Slovenia, Ljubljana and Maribor, are fragmented into a number of FRs. On the other hand, most of the rest of Slovenian territory is delimitated into FRs with a similar area. The administrative unit of Ljubljana consists of 10 municipalities that belong to 9 different FRs, and Maribor covers the territory of 6 municipalities that belong to 4 different FRs. The fragmentation of the metropolitan areas of Ljubljana and Maribor is visible even at the level of 12 FRs; see Fig. 5.

As Landre and Håkansson (2013) have already reported for Sweden, this was also the case for Slovenia, i.e. that FRs generated by the Intramax procedure resulted in a fragmented pattern with (unacceptable) low levels of self-containment in the metropolitan area. From Fig. 7 it is evident that the two most important employment centres in Slovenia (Ljubljana and Maribor) are surrounded by municipalities/FRs with a very low proportion of intra-regional flows and low SSSC and DSSC. DSSC of near

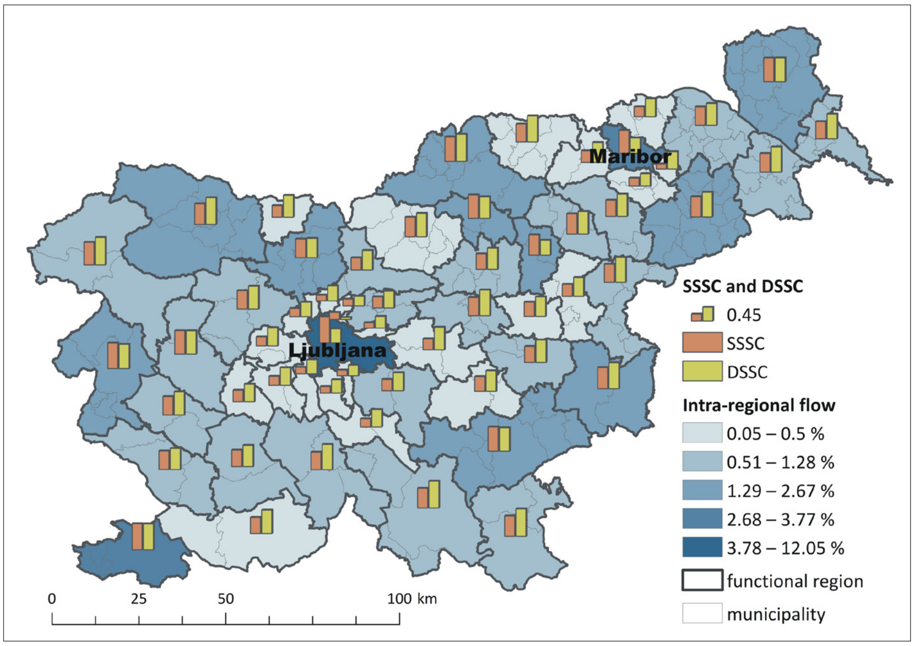


Fig. 7: 58 functional regions modelled by original Intramax procedure (Slovenia, 2011)
Source: authors´ calculations

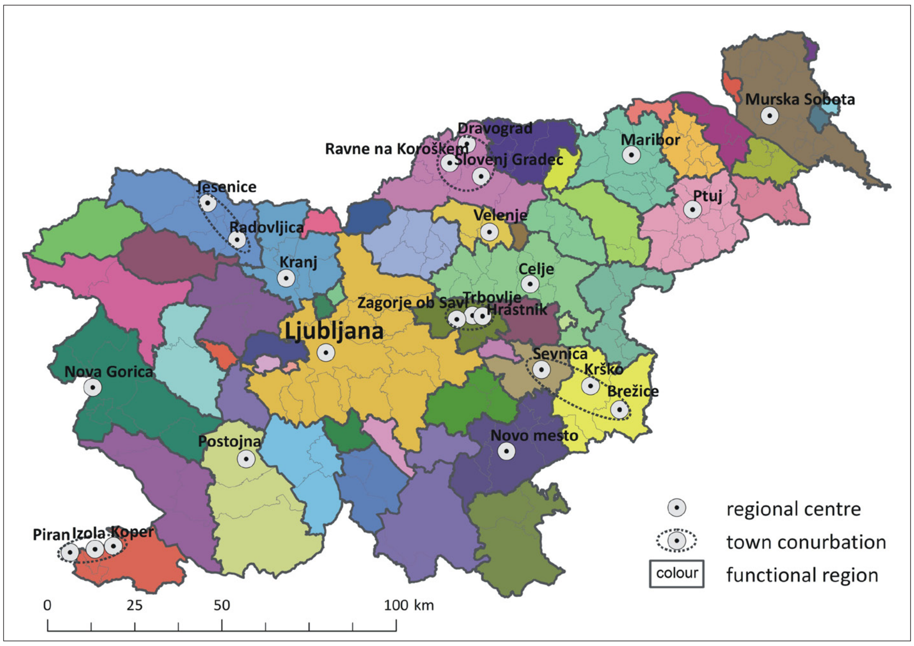


Fig. 8: 58 functional regions modelled using Smart's weighted objective function and 15 regional centres of Slovenia as defined in the Spatial Development Strategy of Slovenia (SDSS, 2004) (Slovenia, 2011)
Source: authors´ calculations and SDSS (2004)

municipalities/FRs is much higher than their SSSC. So, the final result modelled by Intramax procedure could be controlled by additional self-containment criteria to amalgamate BDUs/FRs that do not meet them. In this way, most neighbour municipalities/FRs around Ljubljana and Maribor would be aggregated into two bigger FRs that reflected the functional urban area much more realistically than the delimitation shown in Fig. 7. On the other hand, the problem of fragmented (metropolitan) urban areas can be (partly) solved by using Smart's weighted interaction index F_5 instead of original Intramax objective function F_1. Figure 8 shows the result of using objective function F_5 in the hierarchical aggregation procedure. Here, 58 FRs generated by Smart's weighted interaction index are compared to 15 regional centres in Slovenia that are defined in the Spatial Development Strategy of Slovenia (SDSS, 2004).

5. Conclusions

There are many different approaches and methodologies to delineate functional regions (some of which are mentioned in this paper). Intramax is a hierarchical aggregation procedure that tends to delimitate, in its original form (when using the original Intramax objective function), homogeneous FRs regarding intra-regional flows. Variation of intra-regional flows, delimitated by other objective functions analysed herein, is always higher. The tendency to generate FRs with similar intra-regional flows is the reason why the original Intramax function divides metropolitan urban areas into smaller sub-regions (Drobne and Lakner, 2016).

Use of the original Intramax objective function in the hierarchical aggregation procedure operationally delineates the most persuasive regions, but self-containment statistics, and especially the proportion of inner flows, are less acceptable. In general, other objective functions give statistically more persuasive results but operationally less suitable regions. More precisely, if the problem of singleton regions and small isolated FRs were manually corrected, Smart's weighted interaction index (Smart, 1974: see also equation (10) in Tab. 1) would aggregate regions that would be operationally acceptable for the case of Slovenia. Smart's weighted interaction index also generates non-fragmented functional urban areas.

In this case study, we compared functional regions delimitated by the original Intramax procedure to three official delimitations of the Slovenian territory. Two delimitations that are based on long-term analyses of functional interactions, fine-tuned and optimised several times in the past, demonstrate higher self-containment than functional regions. On the other hand, the recent delimitation of Slovenia into two cohesion regions at NUTS 2 level shows lower self-containment than the two functional regions calculated in the original Intramax procedure, using labour commuting inter-municipal interactions.

As already noted, hierarchical aggregation procedures, like Intramax, do not guarantee a global optimal solution to the regionalisation problem. But, the most important advantage of the hierarchical aggregation procedure is its capability to reveal the structure of the grouping process. Masser and Brown (1978: 17) concluded that "the development of special types of hierarchical aggregation procedure is a useful starting point for dealing with the multi-level specification problem. Procedures of this kind have the advantage that they give insights into the structure of the grouping process which can be used to select the desired level of spatial representation and they also give some indication as to the possible configuration of basic data units that occurs at different levels in the grouping process".

From this point of view, the step-by-step use of two or more objective functions in the same aggregation procedure could be an interesting topic for future research. Here, a combination of the original Intramax objective function and Smart's weighted interaction index to avoid the fragmentation of metropolitan areas should be looked into first. On the other hand, and as already noted by Landre and Håkansson (2013), the fragmented pattern of functional regions in metropolitan areas could be improved by the inclusion of the additional self-containment criteria in the whole aggregation procedure or in some of the last aggregation steps. Searching for a new theoretically defined objective function could be another promising research direction. Variations of the objective functions compared in this paper could be analysed, starting with non-symmetric functions ($F_{ij} \neq F_{ji}$), in which different weights could be assigned to origins or destinations.

References:

ABLER, R., ADAMS, J. S., GOULD, P. (1972): Spatial Organization. London, Prentice-Hall.

ALVANIDES, S., OPENSHAW, S., DUKE-WILLIAMS, O. (2000): Designing zoning systems for flow data. In: Atkinson, P., Martin, D. [eds]: Innovations in GIS 7, GIS and geocomputation, New York, Taylor and Francis, 115–134.

BALL, R. M. (1980): The use and definition of travel-to-work areas in Great Britain: some problems. Regional Studies, 14(2): 125–139.

BERRY, B. J. L., GARRISON, W. L. (1958): The functional bases of the central place hierarchy. Economic Geography, 34(2): 145–154.

BROWN, P. J. B., HINCKS, S. (2008): A framework for housing market area delineation: Principles and application. Urban Studies, 45(11): 2225–2247.

BROWN, P. J. B., PITFIELD, D. E. (1990): The Intramax derivation of commodity market structures from freight flow data. Transportation Planning and Technology, 15(1): 59–81.

CASADO-DIAZ, J. M. (2000): Local labour market areas in Spain: A case study. Regional Studies, 34(9): 843–856.

CASADO-DIAZ, J. M., COOMBES, M. G. (2011): The delineation of 21st century local labour market areas: a critical review and a research agenda. Boletín de la Asociación de Geógrafos Españoles, 57: 7–32.

COOMBES, M. G. (2010): Defining labour market areas by analysing commuting data: innovative methods in the 2007 review of Travel-to-Work Areas. In: Stillwell, J., Duke-Williams, O., Dennett, A. [eds.]: Technologies for migration and commuting analysis: Spatial interaction data applications. Hershey, IGI Global, 227–241.

COOMBES, M. G., BOND, S. (2008): Travel-to-Work Areas: the 2007 review. Office for National Statistics, London.

COOMBES, M. G., DIXON, J. S., GODDARD, J. B., OPENSHAW, S., TAYLOR, P. J. (1982): Functional regions for the population census of Great Britain. In: Herbert, D. T., Johnston, R. J. [eds.]: Geography and the Urban Environment. Progress in Research and Applications 5, Chichester, John Wiley and Sons Ltd., 63–112.

COOMBES, M. G., GREEN, A. E., 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(10): 943–953.

CÖRVERS, F., HENSEN, M., BONGAERTS, D. (2009): Delimitation and coherence of functional and administrative regions. Regional Studies, 43(1): 19–31.

DARAS, K. (2005): An information statistics approach to zone design in the geography of health outcomes and provision. PhD dissertation, University of Newcastle. Newcastle, United Kingdom.

DE JONG, T., VAN DER VAART, N. (2013): Manual Flowmap 7.4.2, Faculty of Geographical Sciences, Utrecht University, The Netherlands. [online]. [cit. 14.08.2015]. Available at: < http://flowmap.geo.uu.nl/downloads/FM742_Manual.pdf>

DROBNE, S., BOGATAJ, M. (2012a): A method to define the number of functional regions: An application to NUTS 2 and NUTS 3 levels in Slovenia = Metoda opredelitve števila funkcionalnih regij: Aplikacija na ravneh NUTS 2 in NUTS 3 v Sloveniji. Geodetski vestnik, 56(1): 105–150.

DROBNE, S., BOGATAJ, M. (2012b): Evaluating functional regions. In: Babić, Z. et al. [eds]: 14th International conference on operational research, Trogir, Croatia, September 26–28, 2012, Croatian operational research review, 3: 14–26. [online]. [cit. 10.08.2015]. Available at: <hrcak.srce.hr/file/142254>

DROBNE, S., BOGATAJ, M. (2014): Regions for servicing old people: Case study of Slovenia. Business systems research journal, 5(3): 19–36.

DROBNE, S., BOGATAJ, M. (2015): Optimal allocation of public service centres in the central places of functional regions. In: 15th IFAC/IEEE/IFIP/IFORS Symposium on Information Control Problems in Manufacturing, INCOM 2015. Ottawa, IFAC, 2449–2454.

DROBNE, S., LAKNER, M. (2015): Intramax and constraints. In: Zadnik-Stirn, L., Žerovnik, J., Kljajić Borštner, M., Drobne, S. [eds.]: The 13th International Symposium on Operations Research in Slovenia, SOR '15. Bled, SDI-SOR, 433–438.

DROBNE, S., LAKNER, M. (2016): Use of constraints in the hierarchical aggregation procedure Intramax. Business Systems Research Journal, 7 (in print).

DROBNE, S., LISEC, A., KONJAR, M., ZAVODNIK LAMOVŠEK, A., POGAČNIK, A. (2009): Functional vs. Administrative regions: Case of Slovenia. In: Vujošević, M., Perić, J. [eds.]: International Scientific Conference Regional Development, Spatial Planning and Strategic Governance: Thematic Conference Proceedings, Belgrade, 395–416.

FARMER, C. J. Q., FOTHERINGHAM, A. S. (2011): Network-based functional regions. Environment and Planning A, 43(11): 2723–2741.

FELDMAN, O., SIMMONDS, D. C., TROLL, N., TSANG, F. (2005): Creation of a system of functional areas for England and Wales and for Scotland. European Transport Conference, Strasbourg (on CD), PTRC, London. [online]. [cit. 06.08.2015]. Available at: <http://www.davidsimmonds.com/files/resourcesmodule/@random4767836fd9d18/1200086162_ETC_2005_Feldman_et_al.pdf >

FENG, Z. (2009): Fuzziness of Travel to Work Areas. Regional Studies, 43(5): 707–720.

FISCHER, M. M., ESSLETZBICHLER, J., GASSLER, H., TRICHTL, G. (1993): Telephone communication patterns in Austria – A comparison of the IPFP-based graph-theoretic and the Intramax approaches. Geographical Analysis, 25(3): 224–233.

GOETGELUK, R., DE JONG, T. (2007): What about the spatial dimension of subsidiarity in housing policy? ENHR 2007 International conference on Sustainable Urban Areas. ENHR, Rotterdam. [online]. [cit. 06.08.2015]. Available at: < https://www.yumpu.com/en/document/view/26672955/what-about-the-spatial-dimension-of-subsidiarity-in-housing-policy>

GREGORY, D., JOHNSTON, R. J., PRATT, G., WATTS, M., WHATMORE, S. [eds.] (2009): The Dictionary of Human Geography, 5th edition. Oxford, Wiley-Blackwell.

HIRST, M. A. (1977): Hierarchical aggregation procedures for interaction data: A comment. Environment and Planning A, 9(1): 99–103.

JAEGAL, Y. (2013): Delineating housing market areas in the Seoul metropolitan area using a geo-computational approach. Journal of the Association of Korean Geographers, 2(1): 7–20.

KARLSSON, C., OLSSON, M. (2006): The identification of functional regions: theory, methods, and applications. The Annals of Regional Science, 40(1): 1–18.

KIM, H., CHUN, Y., KIM, K. (2015): Delimitation of functional regions using a p-regions problem approach. International Regional Science Review, 38(3): 235-263.

KILLIAN, M. S., TOLBERT, C. M. (1993): Mapping social and economic space: the delineation of local labour markets in the United States. In: Singelmann, J., Desaran, F. A. [eds.]: Inequalities in Labour Market Areas. Boulder, Westview, 69–79.

KLAPKA, P., HALÁS, M., TONEV, P. (2013): Functional regions: Concept and types. In: 16th International Colloquium on Regional Sciences, Conference Proceedings. Brno, Masaryk University, 94–101.

KOHL, T., BROUVER, A. E. (2014): The development of trade blocs in an era of globalisation. Environment and Planning A, 46(7): 1535–1553.

KOO, H. (2012): Improved hierarchical aggregation methods for functional regionalization in the Seoul metropolitan area. Journal of the Korean Cartographic Association, 12(2): 25–35.

KRYGSMAN, S., DE JONG, T., NEL, J. (2009): Functional transport regions in South Africa: An examination of national commuter data. Proceedings of the 28th South African transport conference. Pretoria, South Africa, 144–154. [online]. [cit. 06.08.2015]. Available at: < http://repository.up.ac.za/bitstream/handle/2263/11952/Krygsman_Functional%282009%29.pdf>

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): 1–19.

MASSER, I., BROWN, P. J. B. (1975): Hierarchical aggregation procedures for interaction data. Environment and Planning A, 7(5), 509–523.

MASSER, I., BROWN, P. J. B. (1977): Spatial representation and spatial interaction. Papers of the Regional Science Association, 38(1): 71–92.

MASSER, I., BROWN, P. J. B. (1978): Spatial representation and spatial interaction: an overview. In: Masser, I. and Brown, P. J. B. [eds]: Spatial representation and spatial interaction. Studies in applied regional science, Leiden and Boston, Martinus Nijhoff Social Sciences Division, 1–23.

MASSER, I., SCHEURWATER, J. (1978): The specification of multi-level systems for spatial analysis. In: Masser, I. and Brown, P. J. B. [eds]: Spatial representation and spatial interaction. Studies in applied regional science, Leiden and Boston, Martinus Nijhoff Social Sciences Division, 151–172.

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

MITCHELL, W., BAUM, S., FLANAGAN, M., HANNAN, M. (2013): CofFEE functional economic regions. AURIN project. Centre of Full Employment and Equity. Darwin, Australia. [online]. [cit. 06.08.2015]. Available at: < http://e1.newcastle.edu.au/coffee/functional_regions/>

MITCHELL, W., BILL, A., WATTS, M. (2007): Identifying functional regions in Australia using hierarchical aggregation techniques. Centre of Full Employment and Equity. Working Paper No. 07-06. Newcastle, Australia. [online]. [cit. 06.08.2015]. Available at: < http://e1.newcastle.edu.au/coffee/pubs/wp/2007/07-06.pdf>

MITCHELL, W., STIMSON, R. (2010): Creating a new geography of functional economic regions to analyse aspects of labour market performance in Australia. In: Dalziel, P. [ed.]: Innovation and regions: Theory, practice and policy. Lincoln, New Zealand, AERU Research Unit, 178–220.

MITCHELL, W., WATTS, M. (2010): Identifying functional regions in Australia using hierarchical aggregation techniques. Geographical Research, 48(1): 24–41.

NEL, J. H., KRYGSMANY, S. C., DE JONG, T. (2008): The identification of possible future provincial boundaries for South Africa based on an Intramax analysis of journey-to-work data. Orion, 24(2): 131–156.

OECD (2002): Redefining territories: the functional regions. Paris, OECD.

ONS, COOMBES, M. G. (1998): 1991-based Travel-to-Work Areas. Office for National Statistics, London.

OPENSHAW, S., ALVANIDES, S. (1999): Applying geocomputation to the analysis of spatial distributions. In: Longley, P. A., Goodchild, M. F., Maguire, D. J., Rhind, D. W. [eds.]: Geographical Information Systems: Principles, Techniques, Applications and Management, Chichester, Wiley, 1: 267–282.

OPENSHAW, S., RAO, L. (1995): Algorithms for re-engineering 1991 census geography. Environment and Planning A, 27(3): 425–446.

OPENSHAW, S., WYMER, C. (1995): Classifying and regionalizing census data. In: Openshaw, S. [ed.]: Census Users Handbook. Cambridge, United Kingdom, GeoInformation International, 239–270.

POON, J. P. (1997): The cosmopolitanization of trade regions: Global trends and implications, 1965–1990. Economic Geography, 73(4): 390–404.

SLATER, P. B. (1981): Comparisons of aggregation procedures for interaction data: An illustration using a college student international flow table. Socio-Economic Planning Sciences, 15(1): 1–8.

SMART, M. W. (1974): Labour market areas: uses and definition. Progress in Planning, 2(4): 239–353.

SORS (2016): Explanations of territorial changes of statistical region, Statistical Office of the Republic of Slovenia, Ljubljana [online]. [cit. 06.08.2015]. Available at: http://www.stat.si/dokument/8486/Explanations-territorial-changes-statistical-regions.pdf.

SDSS (2004): Spatial development strategy of Slovenia, Ministry of the Environment, Spatial Planning and Energy, Ljubljana [online]. [cit. 06.08.2015]. Available at: http://www.mop.gov.si/fileadmin/mop.gov.si/pageuploads/publikacije/drugo/en/sprs_eng.pdf; retrieved 25 February 2016.

STATISTICS SWEDEN (2010): Construction and use of labour market areas in Sweden, Enterprise and Register-based Employment Statistics Unit, Örebro [online]. [cit. 06.08.2015]. Available at: http://www.scb.se/statistik/_publikationer/AM0207_2009A01_BR_AM95BR1001.pdf; retrieved 23 January 2012.

TYREE, A. (1973): Mobility ratios and association in mobility tables. Population Studies, 27(3): 577–588.

ULLMAN, E. L. (1980): Geography as spatial interaction. Seattle and London: University of Washington Press.

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(2): 201–221.

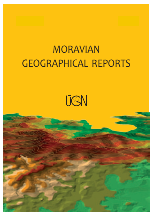
VANHOVE, N., KLAASSEN, L. H. (1987): Regional policy: a European approach, 2nd edition. Avebury, Gower Publishing Company Limited, Aldershot.

WARD, J. H. (1963): Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association, 58(301): 236–244.

WATTS, M. (2009): Rules versus hierarchy: An application of fuzzy set theory to the assessment of spatial grouping techniques. In: Kolehmainen, M. et al. [eds.]: Adaptive and naturals computing algorithms. Berlin Heidelberg, Springer-Verlag, 517–526.

Please cite this article as:

DROBNE, S., LAKNER, M. (2016): Intramax and other objective functions: The case of Slovenia. Moravian Geographical Reports, 24(2): 12–25. Doi: 10.1515/mgr-2016-0007.



MORAVIAN GEOGRAPHICAL REPORTS

DE GRUYTER OPEN

Institute of Geonics, The Czech Academy of Sciences

journal homepage: **http://www.geonika.cz/mgr.html**

doi: **10.1515/mgr-2016-0008**

Delineating zones to increase geographical detail in individual response data files: An application to the Spanish 2011 Census of population

Lucas Martínez-Bernabeu [a *], José Manuel Casado-Díaz [b]

Abstract

Due to confidentiality considerations, the microdata available from the 2011 Spanish Census have been codified at a provincial (NUTS 3) level except when the municipal (LAU 2) population exceeds 20,000 inhabitants (a requirement that is met by less than 5% of all municipalities). For the remainder of the municipalities within a given province, information is only provided for their classification in wide population intervals. These limitations, hampering territorially-focused socio-economic analyses, and more specifically, those related to the labour market, are observed in many other countries. This article proposes and demonstrates an automatic procedure aimed at delineating a set of areas that meet such population requirements and that may be used to re-codify the geographic reference in these cases, thereby increasing the territorial detail at which individual information is available. The method aggregates municipalities into clusters based on the optimisation of a relevant objective function subject to a number of statistical constraints, and is implemented using evolutionary computation techniques. Clusters are defined to fit outer boundaries at the level of labour market areas.

Key words: *labour market areas, census, microdata, regionalisation, clustering, evolutionary computation, Spain*

Article history: *Received 30 September 2015; Accepted 10 December 2015; Published 30 June 2016*

1. Introduction

As in many other countries, microdata derived from a Census of Population are a very rich source of information for the analysis of socio-economic phenomena in Spain. One of the potential applications of this dataset is conducting labour market analyses at very detailed territorial levels, something that is not feasible when using other, more frequently updated, sources of information such as a Labour Force Survey, due to their sampling limitations. When attempting to conduct analyses based on Census microdata, however, researchers and other potential users are faced with the fact that (e.g., in the Spanish case) the geographic reference provided for the majority of indicators appears at a provincial level (NUTS 4), with information on the reference municipality (LAU 2) available only when the population of said municipality exceeds some threshold (e.g., 20,000 inhabitants in Spain). In the remainder of the cases, the information is aggregated into four groups of municipalities[1] for each province, mainly due to confidentiality and sampling constraints. In these cases, apart from the province code, there is only information provided on the population category to which the municipality of residence belongs (e.g. the municipality of residence belongs to province x and is in the range 2,001–5,000 inhabitants). Since about 95% of the 8,116 Spanish municipalities have less than 20,000 inhabitants, this characteristic of the microdata set results in the loss of a large amount of potentially useful information[2]. These restrictions apply to seven territorial variables: place of residence, place of birth, previous place of residence, place of residence one year ago, place of residence ten years ago, place of second residence and place of work.

The motivation for this article is therefore practical. It seeks to produce a geography that allows the re-codification

[a] International Economics Institute, University of Alicante, Spain (*corresponding author: L. Martínez-Bernabeu, e-mail: *lucas.martinez@ua.es*)

[b] International Economics Institute and Department of Applied Economic Analysis, University of Alicante, Spain

[1] Group 1: municipalities with less than 2,001 inhabitants; group 2: between 2,001 and 5,000; group 3: between 5,001 and 10,000; and group 4: between 10,001 and 20,000.

[2] This fact is discussed in detail in section 5.

of the territorial variables in the Census microdata file in order to regain as much spatial information as possible, so that the currently vague reference to population intervals for less-populated municipalities may be substituted by a reference to specific clusters of municipalities. Such clusters are designed to meet the statistical constraints imposed by the National Institute of Statistics, which in this case refer only to a minimum population threshold, and to nest in the upper-level geography of labour market areas (LMAs, a set of functional areas which was defined in other research: Martínez-Bernabeu et al., 2016)[3]. Some of the LMAs have a reasonable level of spatial resolution to be used as the geographic reference in order to re-codify the microdata, but many of them could be further sub-divided if the only requirement to be fulfilled is that of having a minimum population size of 20,000. Therefore, to further increase territorial detail, this article aims at sub-dividing these LMAs into so-called "municipality clusters", with populations over 20,000 inhabitants for which a minimum self-containment level is not required. As in the case of LMAs, this regionalisation is characterised by an exhaustive coverage of the entire territory under consideration, not allowing overlapping between the resulting areas and enforcing contiguity between the municipalities making up each cluster.

Fulfilling the objective of this article, the subdivision of the Spanish LMAs into their constituting clusters of municipalities to increase territorial detail in the Census 2011 individual data sets, involved the definition of a new procedure based on evolutionary computation. Such a procedure has been tailored to fit the specific characteristics of the problem, since despite being guided by the commuting links between the municipalities and the interaction between clusters in these same terms, the process of delineation of clusters is quite different from the identification of LMAs. Thus, while the aim of the definition of the Spanish LMAs was to maximise the internal interaction between the constituting municipalities within each LMA subject to the fulfilment of both a minimum population condition and a trade-off rule between self-containment and area (Martínez-Bernabeu et al., 2016), the aim of this work is to maximize the number of clusters identified so that each LMA is sub-divided into as many clusters as possible, each of which must exceed the minimum population threshold.

The remainder of this article is organized as follows: Section 2 provides the background for the analysis and in Section 3 the different elements making up the problem are described in detail. The latter include the problem formulation as an optimization procedure subject to certain constraints, guided by a fitness function based on an interaction (in terms of travel-to-work) index. In Section 4, the evolutionary algorithm (structure, chromosome representation, operators and configuration/parameters) used is described in detail. The resulting set of 931 municipality clusters is presented and discussed in Section 5. Finally, Section 6 offers some conclusions.

2. Background

The problem addressed in this article -- grouping a set of elements with an associated size (or cost) into as many disjoint groups subject to reach a minimum size as possible – is a specific case of the more general Set Partitioning (SP) problem (Balas and Padberg, 1976). In the SP problem, the input is a finite set of elements, U, called a universe, and a set of possible subsets of the universe, S, each with an associated cost. The task is to find the partition P (i.e. a subset of S so that all sets in P are pairwise disjoint and the union of P is equal to the universe) with minimum total cost, calculated as the sum of the costs of each subset in the partition. This is a complex problem (NP-complete) having numerous real-life applications, e.g. airline crew scheduling (Barnhart et al., 2003) and vehicle routing (Toth and Vigo, 2001). Most of the applications of the SP problem solve it through integer programming for small instances and approximation algorithms for instances that become computationally intractable through exact methods (Laporte, 1992). Other forms of optimisation methods, particularly genetic algorithms, have also been successfully used (e.g. Levine, 1996), and are particularly useful when facing large instances in which linear relaxations and approximations for the integer programming approach do not suffice to make them computationally tractable.

This article focuses on a specific instance of this problem. Such an instance has some peculiarities compared with the general SP problem: its objective is to maximise the value of the partition instead of minimising its cost; and, more importantly, the number of possible subsets of the universe, S, is not an input to the problem (i.e. it is unknown a priori). Instead of generating a huge set of possible subsets in a first step and then solving the associated SP problem, these approaches solve both problems simultaneously by applying a stochastic optimisation method that performs a randomised search over possible partitions.

One example of such a family of applications is the delineation of Census "output areas" (OAs). This consists of the grouping of a given set of spatial building blocks into subsets which are argued to be appropriate for the publication and the integration of different datasets derived from a Census of Population. In the case of the UK, OAs of the 2011 Census were defined[4] for England and Wales using the "automated zoning procedure" (AZP) originally designed by Openshaw (1977a and b) and further refined by Openshaw and Rao (1995). This procedure departs from a possible regionalisation of OAs (the definition of such areas specifically produced for the previous Census), and iteratively re-allocates building blocks, chosen at random, between OAs, accepting one specific re-allocation if it improves the design criteria and otherwise rejecting it, until no more positive re-allocations are found after a certain number of iterations. In the case of the OAs (Martin et al., 2001), such criteria included a constraint in terms of minimum population and three objectives to be optimised (with each given the same weight): a target population criterion (minimising the sum

[3] LMAs were defined using a variation of the so-called GEA method (Martínez-Bernabeu, et al., 2012). The output of such a process was the partition of the 8,116 Spanish municipalities into a total of 260 non-overlapping LMAs made up of one or more contiguous municipalities, with each LMA having a population exceeding 20,000 inhabitants and a self-containment of over 70% (i.e., at least 70% of local jobs are taken by residents of the area, and at least 70% of the residents work locally).

[4] On the occasion of the new 2011 Census of Population, maintenance of the existing set of OAs was preferred to the complete re-design of this set of zones. This involved splitting, merging or re-designing a small sub-set of existing OAs, a process that was based on AZP but that required more manual intervention as compared to the original delineation process conducted in 2001 (Cockings et al., 2011).

of the squared differences between OA populations and the specified target population within each administrative area), the (within zone) social homogeneity (measured as the intra-area correlation in terms of dwelling type and tenure); and morphological compactness (which implied minimising the squared perimeter divided by area). The stochastic nature of this procedure allows for an automatic search over the possible regionalisations without the need to implement complex heuristics, but it does have some handicaps. First, it only considers single building blocks re-allocations and only accepts them if they improve the design criteria. Therefore, it does not allow for an exhaustive search of the solutions' space and will get trapped in local maxima if the problem is not trivial. Second, it does not allow changes to be made to the initial number of OAs, which remains fixed as the number of regions of the initial solution or by user input if no initial regionalisation was provided, and that is a problem when there is no a priori knowledge regarding the appropriate number of regions.

A different group of SP problems that has connections with the one on which this article focuses, is that of tackling with the definition of LMAs: areas aimed at capturing the local dimension of labour markets understood as the spaces where local supply and demand for labour meet. Ideally, each LMA should be characterised by being externally self-contained in terms of commuting to work (i.e., there are few commuters travelling between different LMAs), and by being internally integrated in those same terms (i.e., the ideal LMA should consist of basic building blocks among which daily commuting flows are abundant). Although international experience is quite extensive (see, for example, Casado-Díaz and Coombes, 2011), only a limited number of authors have dealt with the problem of delineation of LMAs as a SP problem[5]. Authors who have addressed this issue include Flórez-Revuelta et al. (2008), Farmer and Fotheringham (2011), Fusco and Caglioni (2011), Martínez-Bernabeu et al. (2012), Chakraborty et al. (2013) and Alonso et al. (2015).

Flórez-Revuelta et al. (2008) proposed a grouping evolutionary algorithm (a general-purpose optimisation technique used in Artificial Intelligence, with genetic operators specifically designed to fit grouping problems) in order to optimise a fitness function that measures the interaction within LMAs, subject to reach certain minimum self-containment and population thresholds. Their fitness function is based on the interaction index (originally proposed by Smart, 1974) that is used to define the official Travel-to-Work Areas (TTWAs) in the UK (Coombes et al., 1986; Coombes and Bond, 2008; ONS, 2015) and the Sistemi Locali del Lavoro in Italy (ISTAT, 1997; 2005; 2014), their local version of LMAs. Martínez-Bernabeu et al. (2012) further improved upon the work by Flórez-Revuelta et al. (2008) by designing renovated search operators that allow for higher quality results and a reduction in computational costs. Alonso et al. (2015) propose and exemplify a delineation scheme based on these grouping evolutionary algorithms.

The work by Farmer and Fotheringham (2011) and Fusco and Caglioni (2011) use a different objective function, the modularity quality index. This function, borrowed from Newman and Girvan (2004), was originally developed for the detection of (social) communities in networks. It accumulates the difference between the interaction links within each community and their expected value in a network having the same nodes but with uniformly distributed flows (the null model). The use of the modularity function has been criticised in the context of community detection (Fortunato and Barthelemy, 2007; Lancichinetti and Fortunato, 2011), since it is unable to identify communities (that are obvious to the human eye) when the number of nodes vary sufficiently between different communities (or, in the LMA context, when large variations between the actual LMAs are observed in population terms). Moreover, the expected interaction value in the null model increases with the size of the territory under analysis, while the actual LMAs for a given region should not depend on whether or not some other unrelated regions are included in the analysis. These drawbacks of the modularity function lead to our preference for the interaction function of Flórez-Revuelta et al. (2008), as well as their general methodology, which has been found to produce better results than the widely- applied TTWAs method, in terms of the number of identified LMAs and cohesion values for the same levels of minimum self-containment, while the works based on modularity have not been compared with alternative approaches.

Since this article focuses on the problem of identifying subsets of municipalities within each LMA, it was considered important to retain the assessment of the commuting links at a cluster level as part of the delineation process (and this is a type of variable that is not considered in the OAs definition process, which is based on the attributes of the building blocks and not on the functional relationships observable between them). Moreover, AZP suffer some technical inconveniences that have been outlined above. This led us to favour the adaptation of a different grouping algorithm (GEA: see below) in order to tackle this specific problem instead of adopting any of the other obvious alternatives. This process has involved defining a fitness function, constraints and a set of operators adapted to this specific grouping problem.

3. Problem statement

As stated in the previous sections, the problem consists of the within-LMA grouping of basic spatial units (BSU), in this case municipalities, into as many geographically continuous clusters of municipalities with a minimum size of 20,000 inhabitants as possible. Thus, the number of identified clusters is the main objective to be maximised.

We also introduce the maximisation of the interaction between municipalities within each area as a secondary objective. That is, we shall always prefer producing (continuous) groupings consisting of $(n + 1)$ clusters over groupings of n clusters, but when facing two alternative groupings with the same number of clusters, we shall prefer the one with the higher inner interaction. Thus, the defined clusters shall be as connected as possible, avoiding the identification of clusters composed of BSUs that are not linked by commuting flows whenever possible.

3.1 Problem formulation

Let $U = \{1, 2, \dots N\}$ be a set of $N = |U|$ BSUs (the LMA to be divided into clusters of municipalities); T, the matrix of commuting flows, so that T_{ij} is the number of commuters

[5] The methods applied can be more often characterised as greedy: they use one or more heuristics that quickly produce a reasonable but sub-optimal regionalisation, through methods that are not based on a fitness function to be maximised and therefore cannot be characterised as optimisation procedures.

from BSU i to BSU j; and P, the vector of populations, so that P_i is the population of BSU i. The objective is to obtain the set of clusters $C = \{C1, C2, ... CK\}$ that maximises the fitness function $f(T,P,C)$, described in section 3.2, subject to

- C being a partition of U (i.e., $C_i \neq \emptyset \ \forall \ C_i \in C$; $\cup_{i=1...K} C_i = U$, and $C_i \cap Cj = \emptyset \ \forall \ C_i, C_j \in C, i \neq j$),
- $\sum_{x \in C_i} P_x \geq 20{,}000 \ \forall \ Ci \in C$, and
- each cluster C_i being geographically continuous.

3.2 Commuting interaction index

To assess the degree of commuting interaction between a pair of clusters or BSUs, we use the interaction index proposed by Flórez-Revuelta et al. (2008), a generalisation of the index used in the TTWA method (Coombes et al., 1986). This indicator takes into account the commuting flows in both directions as well as the relative size of both areas to weight the flows between them. Thus, the flows between small interdependent areas are not eclipsed by the flows between larger areas. Let the interaction index between two clusters II(Ci,Cj) be defined as:

$$\mathrm{II}(C_i, C_j) = \mathrm{II}(C_j, C_i) = \frac{T(C_i, C_j)^2}{R_i \cdot J_j} + \frac{T(C_j, C_i)^2}{R_j \cdot J_i} \qquad (1)$$

in which $T(C_i,C_j)$ is the number of commuters from any of the BSUs in C_i to any of the BSUs in C_j; $R_k = T(C_k, U)$ is the total number of workers residing in C_k; and $J_k = T(U,C_k)$ is the total number of jobs in C_k.

3.3 Fitness function

The main objective of maximising the number of identified clusters may be directly represented by the number of identified clusters. The interaction within clusters may be measured with the same fitness function used in Flórez-Revuelta et al. (2008):

$$g(C) = \sum_{i \in U} \mathrm{II}(\{i\}, C^{(i)}) \qquad (2)$$

in which $C^{(i)}$ represents the cluster to which BSU i belongs minus the own BSU i, and $\{i\}$ represents the cluster formed by i alone. This function accumulates the interaction value between (a) each BSU i and (b) the aggregation of the rest of BSUs in the cluster which that specific BSU i is a part of (excluding T_{ii}).

In order to include the interaction value in the fitness function as a secondary objective, to the number of identified areas we add the average global interaction per BSU, with values in the range [0, 1] (that in practice are always close to 0). Thus, the secondary objective can never force the choice of a grouping of n areas over one of $(n + 1)$ areas, but different groupings of n areas will have different evaluations, depending on the associated interaction levels, and it will allow us to choose the one having more within-clusters interaction:

$$f(C) = \mathrm{card}(C) + \frac{g(C)}{N} \qquad (3)$$

4. Optimisation algorithm

We base our proposal on the grouping evolutionary algorithm (GEA) by Martínez-Bernabeu et al. (2012). This type of algorithm, within the family of genetic algorithms (Goldberg, 1989), is based on the principles of natural evolution and the selection of the fittest. Generally speaking, genetic algorithms are stochastic optimisation techniques, and the specific class of grouping genetic algorithms (Falkenauer, 1998) use tailored genetic operators working over an encoding that can represent groupings of elements, in this case clusters of municipalities within LMAs.

Departing from an initial population of solutions (called *individuals*), which are codified as numeric *chromosomes*, new solutions are created by combining the current individuals (as in sexual reproduction) and applying random changes (as in genetic mutations) to the chromosomes. Then, the new individuals are evaluated using a *fitness function* and some of them are chosen (using a *selection scheme* that favours solutions with better evaluations) to remain in the population for the next iteration (called *generation*) of the algorithm, until a certain *stop condition* is met. This is described in detail in the following subsections, where three forms of stochastic selection are used: random (i.e. uniform probability), probability proportional to the attraction (self-explanatory), and 3-way tournament[6] over a certain characteristic (attraction, size, etc.).

4.1 Structure of the optimisation algorithm

The structure of the GEA algorithm follows these steps:

1. Initialise population: Generate N_p valid solutions by taking the whole set of BSUs in U as mono-BSU clusters and apply the greedy heuristic SHA (described in section 4.3) over them;
2. Evaluate fitness and rank population;
3. Repeat until no improvement of the best solution is found for N_g generations:

3.1 Apply genetic operators until N_o new valid individuals are produced, as follows:

3.1.1 Select a parent from the current population with a probability proportional to the fitness ranking;

3.1.2 Randomly select an operator with uniform probability;

3.1.3 If the operator is the crossover, select a second, different parent with a probability proportional to the fitness ranking;

3.1.4 Create a new individual as a copy of the (first) parent;

3.1.5 Apply the selected operator to the new individual;

3.1.6 If the operator terminates successfully, the resulting individual is evaluated; otherwise its fitness will be set to 0 (invalid);

3.2 Rank individuals in the population by their fitness; and

3.3 From the current pool of previous and new individuals, select the N_p individuals that will stay in the population for the next generation, using selection by ranking with elitism for the best.

The N_p parameter defines the population size, the N_o parameter controls how many new individuals are generated in each generation, and the N_g parameter controls how many generations without further improvement will be performed before stopping the search. In our application we set $N_p = 25$ and $N_o = 10$ and $N_g = 5{,}000$.

---

[6] This is performed by randomly selecting three elements and then selecting the one with highest (or lowest) score in the given characteristic.

In contrast with regular genetic algorithms, the crossover and mutation operators are treated equally in a single stage, so that No mutations and no crossovers (or vice versa) could be applied in a given generation.

4.2 Chromosome representation

We use exactly the same representation as in Martínez-Bernabeu et al. (2012), referred to as group-number encoding: the chromosome of an individual is a vector of *N* integers (one for each BSU in *U*), so that the BSUs with the same integer value are allocated to the same cluster. This representation ensures that the solution is a partition and that the corresponding constraints are automatically met (that is, each BSU is assigned to one and only one cluster). The integer values on each chromosome are forced to follow an ascending order to avoid the possibility of having different representations for the same partition, so that for a partition of *x* clusters, the first BSU is always assigned to group 0, the following BSU allocated to a different group will be assigned group 1 (and so on), and the maximum group number will be $x - 1$.

4.3 Stochastic Hierarchical Agglomeration

For the creation of the initial population (step 1 in the GEA algorithm, see section 4.1), as well as for the reparation of the invalid clusters that may result from the crossover operator (described in section 4.4), we adapt the greedy heuristic presented in Martínez-Bernabeu et al. (2012), the Stochastic Hierarchical Agglomeration (SHA). This algorithm starts from a given partition of a set: one cluster per BSU in the case of step 1 of GEA, or the partition resulting from the crossover operator (that will normally include clusters with several BSUs). Then, it iteratively chooses a cluster with population lower than 20,000 and another adjacent area with low population, and merges them, repeating these steps until all of the clusters have at least 20,000 inhabitants. The exact procedure followed in this work is as follows:

1. Terminate successfully if all the clusters have at least 20,000 inhabitants;
2. Select a cluster *G* by 3-way tournament over the inverse of population;
3. Select a cluster *H* adjacent to *G*, with a probability proportional to the inverse of its population; and
4. Merge clusters *G* and *H* and go to step 2.

4.4 Grouping genetic operators

Martínez-Bernabeu et al. (2012) describe ten group-based genetic operators: one crossover and nine mutations designed to cover all general operations over what in mathematical terms are known as disjoint sets. In this study, we have used the crossover and only five of those mutation operators (*M*, *I*, *E*, *D* and *N*), adapted to the particular objectives of our specific grouping problem.

This has affected the attraction criteria between pairs of clusters: while the original operators use the commuting interaction index (eq. 1), to help maximise the main objective of LMA definition (interaction within LMAs), our variants use the inverse of the summation of both cluster's population, to contribute to the maximisation of the number of clusters identified, the main objective of the process:

$$a(C_x, C_y) = \frac{1}{\sum_{i \in C_x} P_i + \sum_{i \in C_y} P_i} \quad (4)$$

The following subsections describe the precise algorithms of each operator used in this work.

4.4.1 Crossover

This operator is based on the standard grouping crossover as described by Falkenauer (1998). A random selection of clusters from one parent is copied over the other, changing the codification so that none of the copied clusters share their code number with any of those already present in the recipient parent. The integrity of the copied clusters is maintained while the clusters of the other parent sharing BSUs with them can became invalid in terms of size or contiguity. Any non-continuous cluster is fragmented into (smaller) contiguous clusters. Then, the SHA procedure (see section 4.3) is applied to all individuals, so that invalid fragments of clusters are merged with adjacent clusters until they are all valid. This procedure can also modify the initially preserved clusters (those that absorb other invalid clusters):

1. Copy all of the information from the first parent into the child;
2. Randomly select a number *r* with uniform distribution between 1 and 66% of the amount of clusters in the second parent;
3. Randomly select *r* distinct clusters from the second parent and copy them into the child, changing the codification so that none of that clusters share their code with any cluster in the offspring;
4. Check each cluster and divide those clusters that are not continuous into their continuous parts;
5. Apply SHA over all of the clusters of the child (reparation of broken clusters from the first parent); and
6. Terminate successfully.

4.4.2 Mutation M: random re-allocations

This operator randomly selects *border*[7] BSUs with low interaction in their clusters, and attempts to re-allocate them to other adjacent clusters. This is the operator closer to the concept of standard mutation in general genetic algorithms:

1. Randomly select a number *r* between 1 and 2% of the total number of BSUs;
2. Repeat *r* times:

2.1 Choose a border BSU *i* with low attraction with its micro-area C_i by 3-way tournament;

2.2 Select a cluster C_j adjacent to *i*, with probability proportional to the attraction to *I*;

2.3 Re-allocation of i from C_i to C_j if both clusters continue being valid; and

3. If at least one effective re-allocation occurred, terminate successfully; otherwise terminate unsuccessfully.

4.4.3 Mutation I: inclusion into a cluster of adjacent BSUs

This operator attempts to increase the size of a cluster with a low population by absorbing some of the adjacent,

[7] A border BSU is one that is adjacent to at least one cluster other than the one that it is currently part of.

unnecessary[8] BSUs into the surrounding clusters with which it shares higher interaction (the opposite of the mutation *E*):

1. Select a cluster C_i with low population by 3-way tournament;
2. Randomly select a number *r* between 1 and 10% of the average number of BSUs per cluster;
3. Repeat r times:

   3.1 Select a BSU *i* adjacent to C_i and belonging to a cluster $C_j \neq C_i$ with a probability proportional to the attraction to C_i;

   3.2 Re-allocate *i* from C_j to C_i if both clusters continue being valid; and
4. If at least one effective re-allocation occurred, terminate successfully; otherwise terminate unsuccessfully.

4.4.4 Mutation E: exclusion of border BSUs with high external attraction from a cluster

This operator attempts to reduce the size of a large cluster by choosing some border BSUs with lower interaction with the rest of the cluster to which it is currently assigned and reassigning them to other related, adjacent clusters. Its process is inverse to that of mutation *I*:

1. Select a cluster C_i of high population by 3-way tournament;
2. Randomly select a number *r* between 1 and 20% of the amount of BSUs in C_i;
3. Repeat *r* times:

   3.1 Select a border BSU *i* from C_i having a low population by 3-way tournament;

   3.2 Select a cluster C_j adjacent to i with a probability proportional to the attraction to I;

   3.3 Re-allocate *i* from C_j to C_i if both clusters continue being valid; and
4. If at least one effective re-allocation occurred, terminate successfully; otherwise terminate unsuccessfully.

4.4.5 Mutation D: dismembering of a cluster and assignation of its constituent BSUs to the adjacent clusters

This operator uses the same mechanism as in mutation E, but finishes only when the cluster disappears. This operator will always reduce the number of clusters by one, worsening the fitness value, but this may allow that subsequent operations find a better solution and help the search process to escape from a local maximum:

1. Select a cluster C_i with low population by3-way tournament;
2. Repeat until there are no remaining BSUs in C_i:

   2.1 Select a border BSU *i* from C_i with low population by 3-way tournament;

   2.2 Select a cluster C_j adjacent to *i* with a probability proportional to the attraction to I;

   2.3 Re-allocate *i* from C_i to C_j if C_j continues being valid after the re-allocation, otherwise terminate unsuccessfully; and
3. Terminate successfully.

4.4.6 Mutation N: creation of a new cluster using a border BSU as seed

This operator chooses an unnecessary, border BSU in a cluster of low population, creates a new cluster from that BSU, and then tries to absorb other unnecessary, adjacent BSUs from surrounding clusters, until the new cluster reaches the minimum population or there are no more available BSUs to absorb:

1. Select a cluster C_i with a high population by 3-way tournament;
2. Select an unnecessary, border BSU *i* from C_i with a low population by 3-way tournament. If it cannot be found, terminate unsuccessfully;
3. Create a new cluster C_j conformed by *I*;
4. Repeat while population of C_j is smaller than 20,000:

   4.1 Select BSU *k* from the BSUs adjacent to C_j, with a probability proportional to the attraction to C_j;

   4.2 Re-allocate *k* from its cluster C_k to C_j if C_k continues being valid after the re-allocation, otherwise terminate unsuccessfully; and
5. Terminate successfully.

5. Results

Of the 260 LMAs which, according to the objective of this article, should be divided into clusters, 86 already had a population of less than 40,000 inhabitants, and therefore a subdivision was not possible. Thus, the grouping technique described in this paper was applied to the remaining 174 LMAs whose populations exceeded 40,000 inhabitants. Of these, 21 LMAs could not be divided because one of the BSUs concentrated most of the population and any grouping of the remaining BSUs could not reach the minimum of 20,000 inhabitants (10 cases), or because they were formed by only one BSU (2 cases: the cities of Ceuta and Melilla, in the north of Africa), or because the contiguity restriction did not allow for a proper division (9 cases). The remaining 153 LMAs were divided into 824 clusters (totalling 931 clusters with the undivided LMAs[9]). As expected, the LMAs that were sub-divided into a larger number of clusters are those centred in the largest metropolitan areas: Madrid (with 60 clusters), Barcelona (34), Valencia (32), Terrassa (32), Sevilla (22) and Bilbao (22).

To assess the extent to which the results increase and improve the territorial detail of the original reference geography of municipalities grouped in ranges of population, we have compared both regionalisations. Figure 1 depicts the geography that currently serves as a territorial reference in the conventionally distributed Census microdata, as described in Section 1. Such geography consists of the 402 municipalities whose population exceeds 20,000 inhabitants (coloured in dark blue), plus the within-province aggregation of the remaining municipalities into groups according to their population range (these groups have been coloured in blue shades according to the specific population group to which their municipalities belong[10]). The combination of both territorial references (large municipalities plus the within-

[8] An unnecessary BSU is one that can be re-allocated to another cluster without breaking the constraints of minimum population and contiguity.

[9] These clusters include one for which the population minimum is not reached: El Hierro (in the Canary Islands). This is a very specific case whose separate consideration is justified since it is the only populated island not reaching the minimum population threshold, despite being one of the territories with higher self-containment levels.

[10] It is noticeable that 28 of such population-range sub-provincial clusters have in fact less than 20,000 inhabitants.

province population groups) results in the 587 clusters that are presented in Figure 1. It is noticeable that in many provinces, one specific cluster (that consisting of the municipalities under 2001 inhabitants) covers most of the area, and that, in general, clusters based on population ranges are formed by fragmented parts among which distances may be very large.

On the other hand, Figure 2 shows the 931 clusters of municipalities obtained with our methodology, which as previously noted, has been applied within each of the LMAs defined in a previous article (the colour scale reflects the clusters' population levels). In this regionalisation, only 171 clusters are formed by a single municipality. Figure 3 focuses on a specific example: the province of

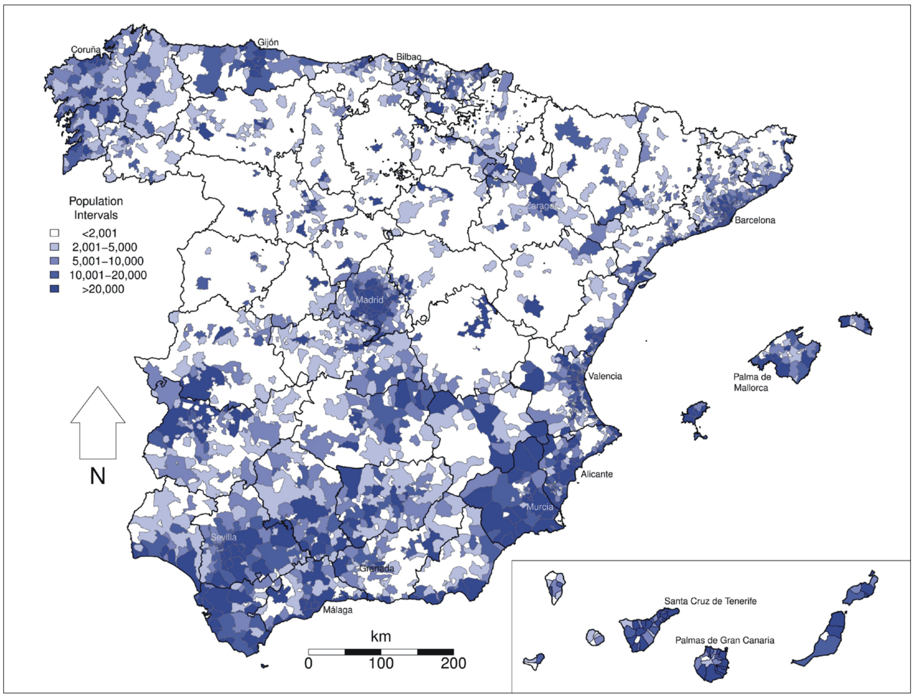


Fig. 1: Geographic reference currently in use in the Spanish Census microdata file 2011. Source: Authors' results based on data from the Spanish Census of Population 2011 (Instituto Nacional de Estadística, INE). Notes: Black lines mark provincial boundaries. Microdata are currently referenced to 587 regions (402 individual municipalities whose population exceeds 20,000 inhabitants – marked in the Figure with the darkest shade – plus within-province groupings of the remaining municipalities according to the population ranges depicted in the Figure's legend – within each province municipalities marked with the same colour belong to one cluster)

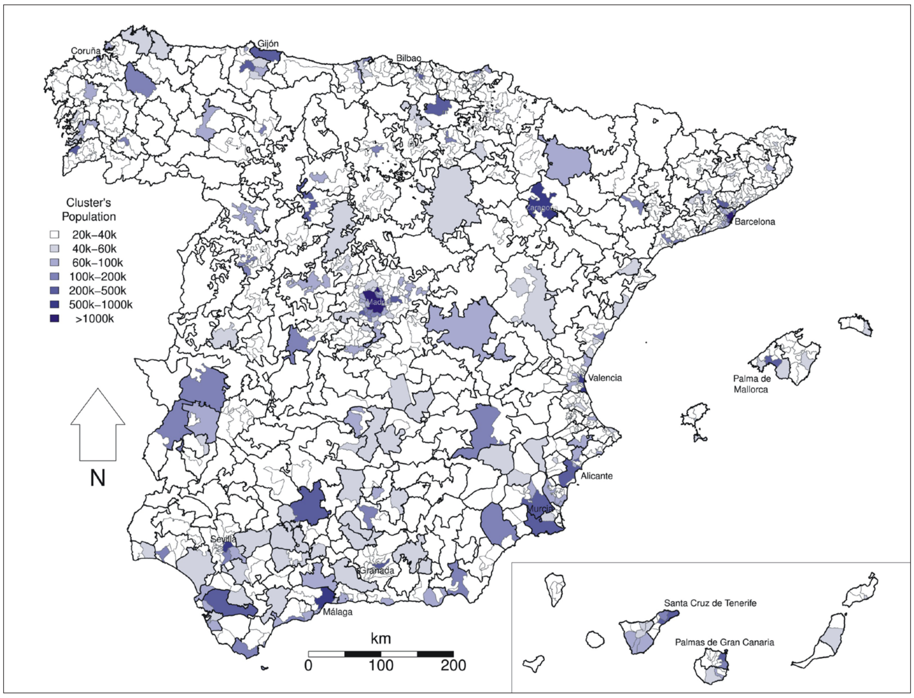


Fig. 2: The proposed geography. Source: Authors' results based on data from the Spanish Census of Population 2011 (Instituto Nacional de Estadística, INE). Notes: Black lines mark LMAs' boundaries. Grey lines mark clusters' boundaries (all clusters are formed by continuous municipalities). The colour scale characterises each cluster according to its population

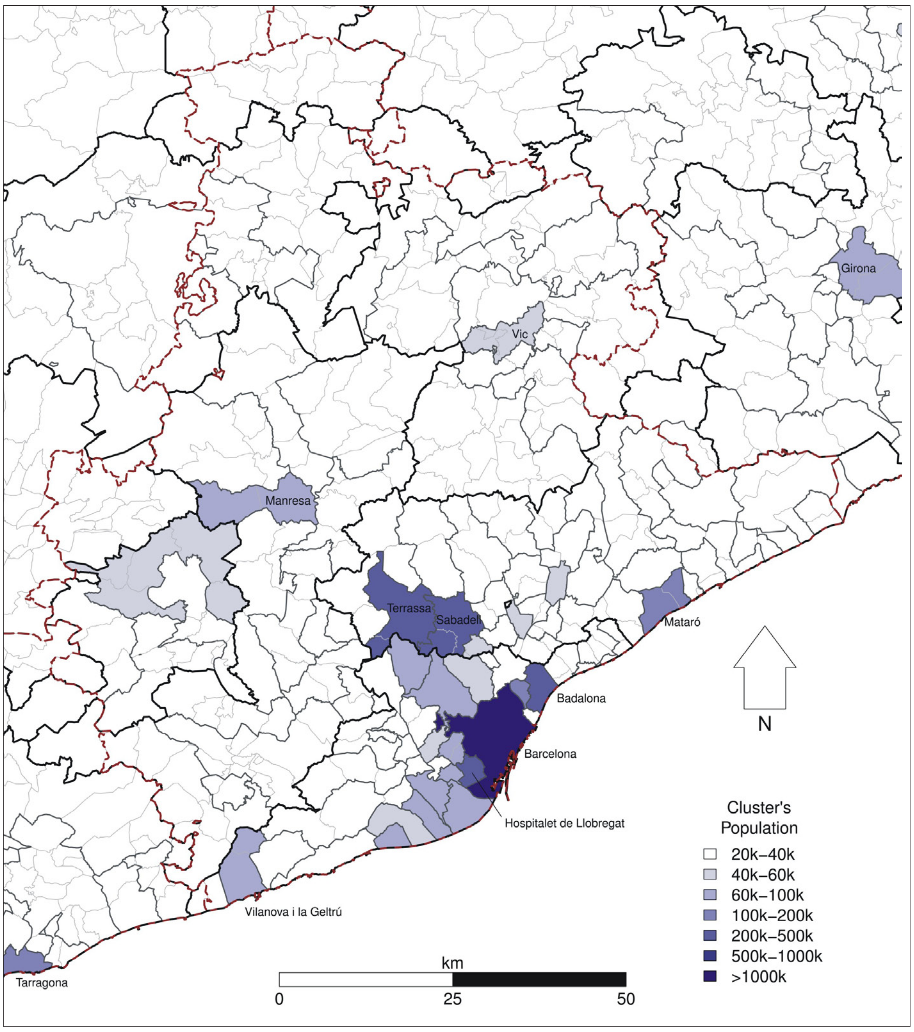


Fig. 3: The proposed geography. Detail of the province of Barcelona. Source: Authors' results based on data from the Spanish Census of Population 2011 (Instituto Nacional de Estadística, INE). Notes: Red line marks provincial boundaries. Black lines mark LMAs' boundaries. Darker grey lines mark clusters' boundaries (all clusters are formed by continuous municipalities). Municipality boundaries in light grey lines. The colour scale characterises each cluster according to its population

Barcelona. This figure illustrates how clusters are structured at a municipality level. Thus that figure shows the actual groupings of municipalities within each cluster of the approximately six LMAs that cover the province.

Morphologically, we see four relevant differences between the original regionalisation (Figure 1) and our results (Figures 2 and 3): (a) our proposal involves a great increase in the microdata territorial detail since it consists of almost 60% more regions, that are more comparable in terms of area; (b) specifically, our proposal divides the large clusters of municipalities that almost completely cover some of the Spanish inner provinces in the current geography into several clusters; (c) many of the municipalities that form a singleton cluster in the currently-used geography are grouped with other smaller municipalities in our proposal, although none of them becomes considerably large; and (d) all of the clusters in our proposal are contiguous. And, obviously given the design of the methodology applied, the clusters of our proposal honours the boundaries of the LMAs of this study.

To complete the description of the proposed geography and its comparison with that which is currently in use, Tables 1 and 2, respectively, depict the number of clusters by area and population intervals. Each table includes, for each regionalisation (the one currently used and the geography proposed here), one column with information for all clusters and a second column in which only the clusters that group at least two (2) municipalities are considered (this column has been labelled "> 1 municipalities").

As shown in both tables, the regionalisation resulting from the method applied in this article offers a much higher level of territorial detail. Thus, in Table 1, it is noticeable that while in the currently-used geography, 19 clusters have an area over 6,000 km^2, this threshold is not exceeded in any of the clusters included in our proposal, and only 6 clusters are over 4,000 km^2, so that microdata records can be referenced to smaller, more specific geographical places. In terms of population (Table 2), the proposed geography of clusters also involves a great increase in the level of detail. In this case, most of the gain (compared with the current geography) occurs in the range between 60,000 and 300,000 inhabitants.

Thus the current territorial division has 335 clusters in the range of 20,000 to 60,000 inhabitants (i.e. 57% of clusters), whereas in our proposal 793 clusters (85.2%) fall within that interval. On the other hand, the geography currently in use includes 209 clusters in the range of 60,001 to 300,000 (35.6% of clusters), while our proposal reduces that number to 123 (13.2%). In comparison, the number of

Land area intervals	Proposed geography		Currently used geography	
	All	> 1 municipalities	All	> 1 municipalities
< 1,000	773	602	478	84
1,000–2,000	96	96	47	39
2,000–3,000	36	36	15	15
3,000–4,000	20	20	16	16
4,000–5,000	3	3	8	8
5,000–6,000	3	3	4	4
6,000–7,000	0	0	5	5
7,000–8,000	0	0	1	1
8,000–9,000	0	0	1	1
9,000–10,000	0	0	2	2
10,000–11,000	0	0	2	2
11,000–12,000	0	0	2	2
12,000–13,000	0	0	4	4
13,000–14,000	0	0	1	1
> 14,000	0	0	1	1
Total	**931**	**760**	**587**	**185**

Tab. 1: Number of clusters by area intervals (km^2). Currently used and proposed geographies. Source: authors' results based on data from the Spanish Census of Population 2011 (Instituto Nacional de Estadística, INE).

Population intervals	Proposed geography		Currently used geography	
	All	> 1 municipalities	All	> 1 municipalities
< 20001	1	1	28	19
20,001–40,000	706	623	246	29
40,001–60,000	87	64	89	38
60,001–80,000	41	20	64	27
80,001–100,000	31	17	54	27
100,001–120,000	10	6	22	15
120,001–140,000	10	4	16	7
140,001–160,000	2	2	10	5
160,001–180,000	7	4	7	1
180,001–200,000	7	3	11	5
200,001–220,000	5	3	13	6
220,001–240,000	4	3	6	2
240,001–260,000	5	3	2	0
260,001–280,000	0	0	2	1
280,001–300,000	1	0	2	1
300,001–350,000	5	4	4	0
350,001–400,000	1	0	2	1
400,001–500,000	2	1	2	0
500,001–750,000	3	2	4	1
750,001–1,000,000	1	0	1	0
1,000,001–2,000,000	1	0	1	0
> 2,000,000	1	0	1	0
Total	**931**	**760**	**587**	**185**

Tab. 2: Number of clusters by population intervals. Currently used and proposed geographies. Source: authors' results based on data from the Spanish Census of Population 2011 (Instituto Nacional de Estadística, INE). Note: see footnotes 9 and 10 for group < 20,001.

clusters with the largest populations, which in most cases correspond to the main cities (already classified as single-municipality clusters), reveals few differences between both regionalisations.

6. Conclusions

This paper deals with a problem that is frequently seen when microdata from different statistical operations are made available for academic research and other uses: the lack of territorial detail derived from sampling or confidentiality restrictions. More specifically, microdata frequently refer to large units such as regions or provinces (NUTS 2 or NUTS 3 in the EU terminology) geographical levels that hamper detailed territorial analyses. When lower-level administrative units are included in the diffusion programmes, they are typically subject to a minimum population restriction. The microdata file associated with the Census of Population 2011 in the Spanish case exemplifies this situation: while the geography currently in use includes the specification of a local territorial reference in the case of municipalities over 20,000 inhabitants, the remaining municipalities are grouped within each province into a maximum of four clusters depending on the population interval to which they belong.

In this article we propose an approach in which a new geography is produced. This partition of the territory is designed to maximise the number of identified clusters of municipalities (so that the detail of the territorial reference used in the microdata file is increased), each of which exceeds a certain minimum population level, with the maximisation of the commuting links between the municipalities that constitute each cluster acting as a secondary objective. Such clusters are identified as sub-divisions of a pre-existing set of LMAs (Martínez-Bernabeu et al., 2016). To achieve that goal, we have designed a new method based on a novel optimisation approach recently applied in the field of functional regionalisation, an evolutionary optimisation technique (GEA: Martínez-Bernabeu et al., 2012) previously used to define the LMAs. In this study, we have adapted the fitness function and the search operators of this technique to adapt to the objectives and restrictions of this specific problem.

The results, an application of the approach to the Spanish case, are designed to increase the territorial detail in the 2011 Census of Population microdata file to permit a more accurate analysis of the labour market at local levels. The resulting geography consists of 931 clusters of municipalities. Some of them (approximately 400) are roughly similar to the ones currently used (they basically correspond to municipalities exceeding the 20,000 inhabitants threshold). The rest (more than 500) are subdivisions of the 185 clusters that in the currently-used territorial division are formed by the aggregation of the municipalities with less than 20,000 inhabitants into four groups within each province. The new clusters are logically characterised by lower figures of both population and area, and allow for an increase in territorial resolution in the microdata file, while respecting the statistical constraints established by the National Institute of Statistics for the diffusion of individual data.

Since the new clusters have been conceived as sub-divisions of the Spanish LMAs, this new regionalisation also permits an analysis at the level of LMAs (Martínez-Bernabeu et al., 2016), in contrast with the reference geography currently included in the microdata file, in which many of the clusters of municipalities are not contiguous and the diverse parts of the clusters are frequently separated by large distances. Moreover, pre-existing clusters have excessively (and unnecessarily) large areas and/or populations, and most of them consist of municipalities from different LMAs, to the detriment of an analysis of the interactions between and within LMAs. None of these drawbacks are present in the alternative regionalisation presented in this article. Moreover, if subsequent analyses find it useful, this subdivision of the territory into smaller clusters would allow for minor adjustments of the LMAs' boundaries. A forthcoming step in this research programme will involve the inclusion of the clusters' territorial codes in the Census 2011 microdata dataset for the seven variables listed previously, and its use in the analysis of commuting and migration behaviour at an individual level, as well as the analysis of the influence of the characteristics of the LMA/cluster of residence on the labour market outcomes of such individuals.

Finally, one incidental contribution of this article is the illustration of how the GEA (Martínez-Bernabeu et al., 2012) algorithm, originally designed for the delineation of LMAs, may be easily adapted to other related contexts through the modification of its fitness function and restrictions, according to the nature of the specific instance of regionalisation to which it is applied.

Acknowledgement

This work was supported by the Spanish Ministry of Economy and Competitiveness (grant number CSO2014-55780-C3-2-P, National R&D&i Plan 2013-2016). Census data were provided by the Instituto Nacional de Estadística (INE), an institution that is not responsible for the use of such data in this research.

References:

ALONSO, M. P., BEAMONTE, A., GARGALLO, P., SALVADOR, M. (2015): Local labour markets delineation: an approach based on evolutionary algorithms and classification methods. Journal of Applied Statistics, 42(5): 1043–1063.

BALAS, E., PADBERG, M. W. (1976): Set partitioning: A survey. SIAM review, 18(4): 710–760.

BARNHART, C., COHN, A. M., JOHNSON, E. L., KLABJAN, D., NEMHAUSER, G. L., VANCE, P. H. (2003): Airline crew scheduling. In Handbook of transportation science. New York, Springer US, 517–560.

CHAKRABORTY, A., BEAMONTE, M. A., GELFAND, A. E., ALONSO, M. P., GARGALLO, P., SALVADOR, M. (2013): Spatial interaction models with individual-level data for explaining labor flows and developing local labor markets. Computational Statistics and Data Analysis, 58: 292–307.

COCKINGS, S., HARBOOT, A., MARTIN, D., HORNBY, D. (2011): Maintaining existing zoning systems using automated zone-designed techniques: methods for creating the 2011 Census output geographies for England and Wales. Environment and Planning A, 43: 2399–2418.

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.

FALKENAUER, E. (1998): Genetic Algorithms and Grouping Problems. New York: John Wiley & Sons.

FLÓREZ-REVUELTA, F., CASADO-DÍAZ, J. M., 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.

GOLDBERG, D. E. (1989): Genetic Algorithms in Search, Optimization, and Machine Learning. Addison, Wesley.

ISTAT (1997): I Sistemi Locali Del Lavoro 1991. Rome, Istituto Nazionale di Statistica.

ISTAT (2005): I Sistemi Locali Del Lavoro 2001. Rome, Istituto Nazionale di Statistica.

ISTAT (2014): Sistemi Locali Del Lavoro. Nota metodologica. Rome, Instituto Nazionale di Statistica [online]. Available at: http://www.istat.it/it/files/2014/12/nota-metodologica_SLL2011_rev20150205.pdf

LANCICHINETTI, A., FORTUNATO, S. (2011): Limits of modularity maximization in community detection. Physical Review E, 84: 066122.

LEVINE, D. (1996): A parallel genetic algorithm for the set partitioning problem. New York, Springer US, 23–35.

LAPORTE, G. (1992): The vehicle routing problem: An overview of exact and approximate algorithms. European Journal of Operational Research, 59(3): 345–358.

MARTIN, D. (2000): Towards the geographies of the 2001 UK Census of Population. Transactions of the Institute of British Geographers, New Series, 25: 321–332.

MARTÍNEZ-BERNABEU, L., CASADO-DÍAZ, J. M., FLÓREZ-REVUELTA, F. (2016): Improving the objective and constraint functions in optimisation-based functional regionalisation methods. Mimeo.

MARTÍNEZ-BERNABEU, L., FLÓREZ-REVUELTA, F., CASADO-DÍAZ, J. M. (2012): Grouping genetic operators for the delineation of functional areas based on spatial interaction. Expert Systems with Applications, 39(8): 6754–6766.

NEWMAN, M. E. J., GIRVAN, M. (2004): Finding and Evaluating Community Structure in Networks. Physical Review E, 69: 026113. doi: http://dx.doi.org/10.1103/PhysRevE.69.026113

ONS (2015): Methodology note on 2011 Travel to Work Areas. London: Office for National Statistics [online]. Available at http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/other/travel-to-work-areas/index.html

OPENSHAW, S. (1977a): A geographical solution to scale and aggregation problems in region-building, partitioning and spatial modelling, Transactions of the Institute of British Geographers, New Series, 2: 459–472.

OPENSHAW, S. (1977b): Algorithm 3: a procedure to generate pseudo-random aggregations of N zones into M zones, where M is less than N, Environment and Planning A, 9: 1423–1428

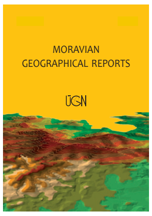
OPENSHAW, S., RAO, L. (1995): Algorithms for re-engineering 1991 Census geography, Environment and Planning A, 27: 425–446.

SMART, M. W. (1974): Labour market areas: uses and definition. Progress in Planning, 69: 238–353.

TOTH, P., VIGO, D. (2001): The vehicle routing problem. Philadelphia, US, Society for Industrial and Applied Mathematics (SIAM).

Please cite this article as:

MARTÍNEZ-BERNABEU, L., CASADO-DÍAZ, J. M. (2016): Delineating zones to increase geographical detail in individual response data files: An application to the Spanish 2011 Census of population. Moravian Geographical Reports, 24(2): 26–36. Doi: 10.1515/mgr-2016-0008.



MORAVIAN GEOGRAPHICAL REPORTS

Institute of Geonics, The Czech Academy of Sciences

journal homepage: http://www.geonika.cz/mgr.html

doi: 10.1515/mgr-2016-0009

DE GRUYTER OPEN

A functional interaction approach to the definition of meso regions: The case of the Czech Republic

Martin ERLEBACH [a]*, Martin TOMÁŠ [a], Petr TONEV [b]

Abstract

The definition of functional meso regions for the territory of the Czech Republic is articulated in this article. Functional regions reflect horizontal interactions in space and are presented as a useful tool for various types of geographical analyses, and also for spatial planning, economic policy designs, etc. This paper attempts to add to the discussion on the need to delineate areal units at different hierarchical levels, and to understand the functional flows and spatial behaviours of the population in a given space. Three agglomerative methods are applied in the paper (the CURDS regionalisation algorithm, Intramax, and cluster analysis), and they have not been used previously in Czech geography for the delineation of functional meso regions. Existing functional regions at the micro-level, based on daily travel-to-work flows from the 2001 census, have served as the building blocks. The analyses have produced five regional systems at the meso level, based on daily labour commuting movements of the population. Basic statistics and a characterisation of these systems are provided in this paper.

Key words: *functional region, local labour market area, travel-to-work flows, meso regional level, functional regionalisation, Intramax, Czech Republic*

Article history: *Received 4 November 2015; Accepted 10 May 2016; Published 30 June 2016*

1. Introduction

The definition of functional regions has played an important role in geographical and spatial research for several decades. The organisation of a space is crucial for an understanding of and an explanation for socio-economic phenomena, and is necessary for the needs of regional development, territorial planning and the implementation of regional policies. Functional regions based on daily travel-to-work flows appear to be a very useful tool in finding a solution to a number of socio-economic issues, unlike administrative or political areal units, which often ignore real geographic functional linkages. Suitably-defined functional regions for a particular research task can avoid the spatial bias caused by the incorrectly defined borders of administrative units (Wong, 2009). In many OECD countries, functional regions are used as statistical units for various analyses (OECD, 2002).

Geographic research often works on various hierarchical levels. The definition of functional regions is no exception (Drobne and Bogataj, 2012; Halás et al., 2014a). Functional regions defined at various hierarchical levels can be used to solve different socio-economic problems. Functional regions can be understood as the spatial delineation of the spheres of influence of settlement system centres. These centres are hierarchically organised according to their significance (size, function), and, obviously, so are the functional regions. In this respect, meso regions identify the spheres of influence of meso regional centres in a settlement system, in this case the Czech Republic. Functional regions at the micro level, such as local labour market areas (LLMA) (Ball, 1980), can be used for labour market analyses, the implementation of local employment policies and the identification of regional disparities, while functional regions at the meso level can serve as a tool for regional planning (regional development strategies).

The definition of functional meso regions and the identification of the main intraregional interactions can also benefit the optimisation of transport systems of a regional and inter-regional importance. Variants of functional meso regions discussed in this paper conform to the NUTS 2 definitions, possibly to the NUTS 3 level in terms of their size. The definition of NUTS regions, however, is based on administrative concepts, unlike the presented variants of functional meso regions. Presently, meso regions (NUTS 3

[a] Department of Geography, Faculty of Science, Palacký University Olomouc, Czech Republic (*corresponding author: Martin Erlebach, e-mail: *Martin.Erlebach@seznam.cz*)

[b] Department of Regional Economics and Administration, Faculty of Economics and Administration, Masaryk University, Brno, Czech Republic

and 2 levels) are used for the implementation of particular EU policies. Each hierarchical level of a functional region can be suitable for the solution of various tasks, therefore it is important to differentiate between them.

The daily rhythm of population movements, particularly daily travel-to-work flows, is a determinative process in the formation of regional systems at the micro level. Based on their flow patterns and further criteria, various types of functional regions can be defined (Klapka et al., 2013a; Erlebach et al., 2014). Similarly, as in the case of the definition of functional micro regions, daily travel-to-work flows are also determinative for the identification of functional meso regions (Halás et al., 2014b). The daily labour commuting rhythm (daily travel-to-work flows) is an aggregation of the regular movements of individuals, describing spatial behaviours for a considerable part of the population. These movements are relatively stable over time (unlike travel for retail services, for example). The frequency of retailing trips and commuting for services, however, is decreasing and they can no longer be considered as regular daily movements (Halás and Zuskáčová, 2013; Halás et al., 2014b).

The main objective of this paper is to define meso regions for the territory of the Czech Republic, using several acknowledged functional interaction approaches. Micro regions (functional regions based on daily travel-to-work flows from the 2001 census, variant FRD 2B) defined by Klapka et al. (2014), serve as the building blocks for the analyses presented. A subsequent objective of the paper is to apply three methods for the definition of functional meso regions, which either were not applied in the Czech Republic at all, or were used only to define functional micro regions. The first method applies the adjusted variant of the CURDS regionalisation algorithm to the delimitation of LLMAs (Coombes, et al., 1986; Coombes, 2010). The second method applies the Intramax regionalisation procedure (Masser and Brown, 1975). The third method uses the approach of a general cluster analysis (e.g. Ward, 1963). The paper also provides the relevant characteristics for, and an assessment of the resulting meso regional systems (see for example, Klapka et al., 2014; Konjar et al., 2010).

2. Functional regions and methods for their definition

The term functional region was introduced into geography in the second half of the 20th century (Philbrick, 1957; Nystuen and Dacey, 1961; Haggett, 1965). The term was inspired by the theories and work of spatial economists (von Thünen, 1826; Christaller, 1933; Lösch, 1940; Isard, 1956). Since then the concept of a functional region has gone through many adjustments (Sýkora and Mulíček, 2009). The primary characteristics of a functional region are its size and its self-containment in relation to its surroundings (Coombes et al., 1986; Casado-Díaz, 2000; Klapka et al., 2013a). Size can be expressed as the population, the number of jobs or some other relevant criterion. Self-containment requires the maximising of the flows within a particular region: at least 50% of the flows incident to a region should occur within its boundaries (Smart, 1974; Farmer and Fotheringham, 2011).

The research literature also provides slightly different views on the definition of a functional region. The OECD (2002) sees it as an areal unit based on socio-economic linkages, regardless of historical and physical geographical conditions. Karlsson and Olsson (2006) see a functional region as an area with a high frequency of socio-economic interactions which occur within the region. Similarly, Sýkora and Mulíček (2009) see a functional region as an area which contains the integrated socio-economic activities of the population. In the geographical literature, several specific types of functional regions can be observed, based on specific characteristics, particularly regarding the type and pattern of region-organising interaction.

The first such instance is a functional urban region (FUR), which is determined by the specific character of its core (Bezák, 2000; Karlsson and Olsson, 2006). A core should have an urban character in this case and interactions should be directed at this core. A similar instance, often confused with the former, is a daily urban system (DUS). Daily urban systems (Berry, 1973) are based on a more complex concept of the limits of a spatial economy (coherence and self-containment), where differences between a core and its adjacent hinterland are less distinct (Hall and Hay, 1980). Unlike the preceding type, DUS is based on daily interactions, and the daily rhythm of flows is significant. Local labour market areas (LLMA) and travel-to-work areas (TTWA) are almost near-synonyms. Local labour market areas (Smart, 1974) are considered to have a slightly wider meaning than travel-to-work areas (Ball, 1980). They need not necessarily be based on labour commuting, but in practice they are, almost without exception. Travel-to-work areas directly refer to the region-organising process, while local labour market areas first refer to their applied use, such as labour market analysis and the implementation of labour market policy (Tonev, 2013).

Basically, three approaches to the definition of functional regions can be identified: a) a graph-theoretic approach; b) a numerical taxonomy approach; and c) a rule-based approach. Graph-theoretic methods are based on the analysis of significant flows and have a long tradition in geographical research (Nystuen and Dacey, 1961; Holmes and Haggett, 1977). The graph-theoretic approach has even been used in some recent regionalising tasks, albeit in its more sophisticated form (Karlsson and Olsson, 2006; and recently via three-step procedure by Kropp and Schwengler, 2014). Methods of numerical taxonomy (e.g. Brown and Holmes, 1971; Masser and Brown, 1975) are based on the analysis of functional (taxonomic) distances. The rule-based / multistage methods were proposed, for example, by Smart (1974) and Coombes, et al. (1982, 1986). The latter proposal represents a relatively complex regionalisation algorithm which has probably had the greatest response from geographical researchers (see Casado-Díaz, 2000; Papps and Newell, 2002; Drobne et al., 2009; Konjar et al., 2010; Bezák, 2000; Halás et al., 2014a, b; Klapka et al., 2013b, 2014; Tonev, 2013). Application of the latter two approaches is a part of this contribution.

All approaches have their pros and cons (see for example, Coombes, 2010; Mitchell and Watts, 2010; Landré, 2012; Landré and Håkansson, 2013). Clustering methods (numerical taxonomy methods) are based on a selected criterion, which is gradually decreased until all the resulting regions meet it (Watts, 2009). The Intramax procedure is seen by Masser and Scheurwater (1980) as a suitable method for the analyses of large data sets and the amalgamation of regions on the basis of the strongest links, without their dissolution during later stages of analysis. On the other hand, these methods are not able to make amendments during the running of the procedure. Multistage methods are able to set several criteria for each step in a procedure and to control the required size and self-containment of the resulting regions (Coombes et al., 1986; Coombes, 2010; Casado-Díaz, 2000; Halás et al., 2015).

3. Procedures

The first step in the analysis was to identify the basic building blocks (basic spatial units) to be amalgamated into meso regions. All four variants of functional micro regions as defined by Klapka et al. (2014) were tested using three different regionalisation methods. An extensive set of the systems of functional meso regions resulted from the analysis. For the purposes of this paper, five variants of functional meso regions were selected on the basis of maximising the total self-containment of respective regional systems. All of these five variants were based on the functional micro regional system FRD 2B (Klapka et. al., 2014); its characteristics are presented in Table 1, where the β_1 and β_2 values determine the lower and upper limits of the self-containment of functional micro regions, and the β_3 and β_4 values determine the lower and upper limits of the size of functional micro regions.

For further analysis, these functional micro regions were arranged into a square interaction matrix (98 × 98 cells), which stored intra-regional and inter-regional flows. This matrix was the basis for the application of the three regionalisation approaches.

3.1 The local labour market area approach

The most well-known and acknowledged multistage procedure for defining functional regions (local labour market areas in this case) was developed by the Centre for Urban and Regional Development Studies (CURDS), see Coombes et al., 1982, 1986), and Coombes (2010). In this paper its adjusted third variant (Coombes and Bond, 2008; Coombes, 2010) has been applied. The adjustments are discussed in Halás et al. (2015). The third variant has a more general hierarchical clustering character – from the beginning of the procedure it considers all basic spatial units as proto-regions. A successful solution is identified through the use of the constraint function, which controls the trade-off between the self-containment and the size of resulting regions, where the upper and lower limits of self-containment and size are set by the researcher. The values for these four parameters (β_1, β_2, β_3 and β_4) have resulted from extensive testing and are presented in Table 2. The size parameter is based on the number of employed persons. The amalgamation of basic spatial units (proto regions) is based on Smart's measure (Smart, 1974), which is expressed by the formula:

$$\frac{T_{ij}^2}{(\sum_k T_{ik} * \sum_k T_{kj})} + \frac{T_{ji}^2}{(\sum_k T_{jk} * \sum_k T_{ki})} \quad (1)$$

where T_{ij} is the flow from spatial unit i into spatial unit j, T_{ji} is the flow from spatial unit j to spatial unit i, $\sum_k T_{ik}$ denotes all out-going flows from i, $\sum_k T_{kj}$ denotes all in-going flows to j, $\sum_k T_{jk}$ denotes all out-going flows from j, and $\sum_k T_{ki}$ denotes all in-going flows to i. After each amalgamation the interaction matrix is updated. An indisputable benefit of this method is that it is possible to set and adjust input parameters for the need to optimise the resulting regional systems.

3.2 The Intramax approach

The second approach to defining functional meso regions in this paper is based on maximising intra-regional flows – Intramax (Masser and Brown, 1975). The Intramax algorithm identifies functional regions through the hierarchical aggregation of basic spatial units, with the objective of forming homogeneous clusters (Landré, 2012). The interaction measure is used as an objective function in the hierarchical clustering algorithm and it can be expressed by the formula:

$$\frac{T_{ij}}{(\sum_k T_{ik} * \sum_k T_{kj})} + \frac{T_{ji}}{(\sum_k T_{jk} * \sum_k T_{ki})} \rightarrow \max \quad (2)$$

The Intramax is a bottom-up procedure, which starts with the same number of clusters as the number of basic spatial units. These clusters are gradually amalgamated until one cluster is formed which consists of all basic spatial units (de Jong and van der Vaart, 2010). Thus, the Intramax is a stepwise analysis: there are n basic spatial units, and $(n - 1)$ steps. The objective of the method is to maximise the share of the inner interactions of a region in the sum of column and row values of the interaction matrix (Masser and Brown, 1975). The adjusted standardised matrix is an input in the aggregation procedure. The Intramax method uses, for the standardisation of the interaction matrix, the following formula:

$$\sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij} = 1 \quad (3)$$

where a_{ij} is a value for an input matrix cell. Each pair of regions is examined separately in each step and is amalgamated on the basis of the maximum value for the objective function. The objective function can be calculated only if all $\sum_k T_{ik} > 0$ and all $\sum_k T_{kj} > 0$ (Landré, 2012).

β_1 value	β_2 value	β_3 value	β_4 value	Self-containment of reg. system	Number of regions	Self-containment (mean)	Self-containment (median)	Self-containment (var. coef.)
0.63	0.75	6,000	70,000	0.915	98	0.828	0.835	0.064
Economically active population (mean)	Economically active population (median)	Economically active population (var. coef*)	Population (mean)	Population (median)	Population (var. coef.*)	Area km^2 (mean)	Area km^2 (median)	Area km^2 (var. coef.*)
53,606.12	37,351.00	1.450	104,388.40	75,130.50	1.391	805.24	736.86	0.468

Tab. 1: Characteristics of regional system FRD 2B. Source: Klapka et al. (2014)
*Note: *var. coef. = Coefficient of variation*

3.3 Cluster analysis approach

The third approach used in this paper to define functional meso regions comes from general cluster analysis methods. Each basic spatial unit is initially considered as a cluster and the procedure runs until all basic spatial units form one single cluster. The interaction matrix in this case is considered to be a matrix of the taxonomic distances (taxonomic dissimilarities) between individual basic spatial units and it was relativised by the CURDS interaction measure (Coombes, et al., 1982):

$$\frac{T_{ij}}{\sum_k T_{ik}} + \frac{T_{ij}}{\sum_k T_{kj}} + \frac{T_{ji}}{\sum_k T_{jk}} + \frac{T_{ji}}{\sum_k T_{ki}} \quad (4)$$

If absolute values for interaction enter the clustering process, there is a disadvantage when the analysis of taxonomic distances (dissimilarities) clusters typologically similar basic spatial units, in this case units with a similar structure of interactions and not units with strong mutual linkages. Therefore the input matrix values have to be "relativized" and correlated. More concretely, the values in the interaction matrix have been normalised by the above-mentioned interaction measure (CURDS). The values in the normalised matrix have been correlated by the use of Pearson's coefficient. The analysis produces a dendrogram, where, on the x axis, there are basic spatial units and on the y axis there are distances (similarities) between clusters. The Silhouette method (Rousseeuw, 1987) has been used for validation of the resulting dendrogram. This method calculates the value of the width of silhouette for each object, the mean value of the width of a silhouette for each cluster, and the mean value of the width of silhouette for the whole data set. This approach compares mean silhouette widths for a given cluster. Silhouette represents the proportion of similarity and dissimilarity to other clusters.

The calculation of the Silhoutte coefficient is done in three consequent steps. Firstly for each object (i) its average distance (a_i) to all remaining objects in a cluster is calculated. Secondly for each object (i) and any cluster not containing the object, the object's average distance from all the objects in the given cluster is calculated. The minimum value (b_i) is found with respect to all clusters. Finally for the object (i) the Silhouette coefficient is calculated according to:

$$s_i = (b_i - a_i)/\max(a_i, b_i) \quad (5)$$

The measurement of similarity of basic spatial units and their clustering into meso regions is based on the Euclidean distance:

$$d_r(x_i, x_j) = \sqrt{\sum_{m=1}^{p} \left(x_{im} - x_{jm}\right)^2} \quad (6)$$

where $d_r(x_i, x_j)$ is the taxonomic distance between object i and j, x_{im} is the value of criterion m for object i, and x_{jm} the value of criterion m for object j. There are several approaches to the clustering of objects on the basis of their taxonomic distance or similarity (Gustafson, 1973). In this paper, Ward's method, based on the loss of information during clustering, has been applied (see Ward, 1963: 239–243). The clustering criterion is the total sum of square errors of each object from the group centroid, to which it belongs.

4. Results and discussion

Table 2 presents the basic statistics for five functional meso regional systems, where systems M-FRD 1A and 1B were produced by the multistage local labour market area approach, systems M-FRD 2A and 2B were produced by the cluster analysis approach, and system M-FRD 3 was produced by the Intramax approach. The Job Ratio function has been calculated according to:

$$\frac{\sum_j T_{ji}}{\sum_j T_{ij}}. \quad (7)$$

If its value exceeds 1, the region offers more jobs than the number of employed persons in the region. Table 2 comprises three types of self-containment. The first type was proposed by Halás, et al., (2015) and is calculated by:

$$\frac{T_{jj}}{(\sum_k T_{jk} + \sum_k T_{kj}) - T_{jj}}. \quad (8)$$

The remaining two types of self-containment (Casado-Díaz, 2003) express the so called supply-side self-containment:

$$\frac{T_{ii}}{\sum_j T_{ij}} \quad (9)$$

and demand-side self-containment:

$$\frac{T_{ii}}{\sum_j T_{ji}} \quad (10)$$

The former expresses the share of employed persons working locally out of the total employed persons in the region, while the latter expresses the share of employed persons working locally out of the total number of jobs in the region.

Five variants of functional meso regional systems of the Czech Republic are presented in figures 1–5, and laid over basic physical geographical conditions. The largest cities in each meso region (in terms of the number of employed persons) are labelled on the maps. The inner structure of meso regions and inter-regional interactions are also presented on the maps. The relationships between basic spatial units (functional micro regions – see Fig. 6 for their names) are used for this purpose, and three ways of representing the interactions are used in order to provide the highest diversity of views on the Czech regional system. Thus, the inter-regional relationships are based on the Smart's measure (Figs. 1 and 2), the CURDS measure (Figs. 3 and 4) and the absolute numbers for daily travel-to-work flows (Fig. 5). It should be noted here that all three were applied to all five meso regional systems; the selection presented in this paper is a result of limited space.

The variant M-FRD 1A (Fig. 1) consists of eight meso regions. The lowest level of self-containment (0.959) is in the meso region in eastern Bohemia (Hradec Králové), with significant cross-border interactions between the pairs of micro regions Jičín–Mladá Boleslav, Vrchlabí–Semily, and Vysoké Mýto–Svitavy. In contrast, the highest level of self-containment (0.981) is in the region of northern Moravia and Silesia (Ostrava), with strong interactions within this region. The overall pattern of this meso regional system very markedly resembles the administrative regional division of Bohemia as it has existed in the Czech Republic (Czechoslovakia) since 1960. In Moravia and Silesia,

Attributes for regional systems	M-FRD 1A	M-FRD 1B	M-FRD 2A	M-FRD 2B	M-FRD 3
β_1 (self-containment, lower limit)	0.8	0.75	x	x	X
β_2 (self-containment, upper limit)	0.9	0.85	x	x	X
β_3 (size of region, lower limit)	290,000	250,000	x	x	X
β_4 (size of region, upper limit)	300,000	251,000	x	x	X
Number of regions	8	9	12	8	9
Size (mean)	554,332	492,740	369,555	554,332	492,740
Size (median)	522,015	484,429	288,810	521,781	380,693
Job ratio function (mean)	1.008	0.994	0.942	0.959	0.919
Job ratio function (median)	0.668	0.741	0.756	0.682	0.858
Self-containment (mean)	0.969	0.965	0.957	0.969	0.965
Self-containment (median)	0.968	0.965	0.958	0.970	0.968
Population (mean)	1,278,758	1,136,673	852,505	1,278,758	1,136,673
Population (median)	1,261,083	1,111,630	643,953	1,261,554	852,794
Supply-side self-containment (mean)	0.983	0.981	0.976	0.983	0.981
Supply-side self-containment (median)	0.984	0.982	0.975	0.984	0.982
Demand-side self-containment (mean)	0.986	0.984	0.980	0.986	0.984
Demand-side self-containment (median)	0.987	0.984	0.983	0.986	0.985

Tab. 2. Characteristics for variants of meso regional systems of the Czech Republic. Sources: authors' computations

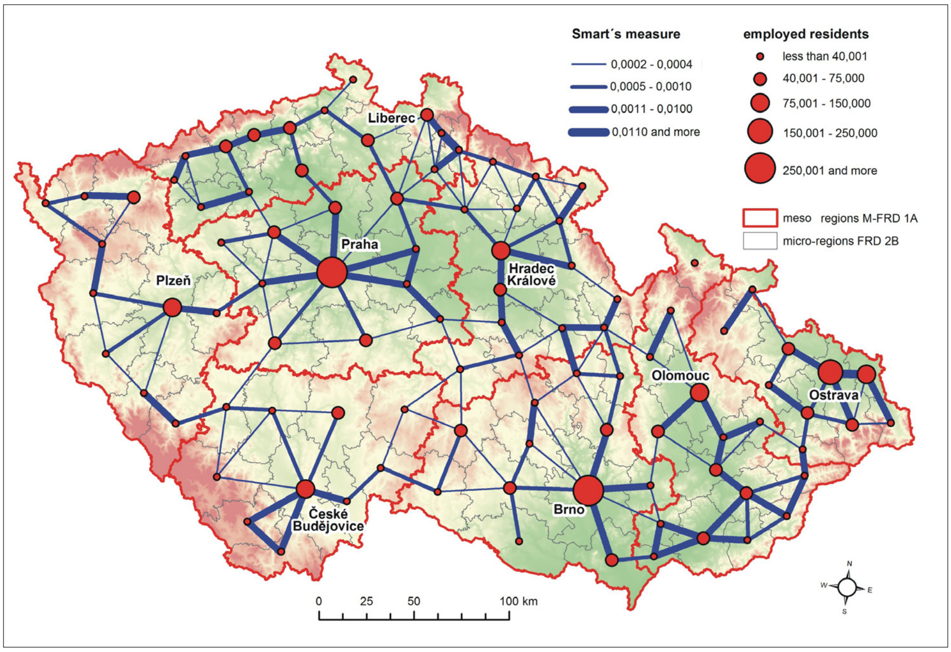


Fig. 1: Regional system M-FRD 1A. Sources: authors' elaboration, Klapka et al., (2014)

however, there is no such significant resemblance. Minor changes in the parameters of the constraint function (slightly lower values for size and self-containment) have produced the variant M-FRD 1B (Fig. 2). In this case the meso region in central Moravia has been split into two (Olomouc and Zlín), with the latter having the lowest level of self-containment (0.952), because of strong cross-border interactions between the pairs of micro regions Kroměříž–Přerov and Vsetín–Valašské Meziříčí. These two variants can be considered as the most suitable for further geographical and socio-economic analyses. In our opinion, they best reflect the geographic space of the Czech Republic and respect a large portion of the country's natural geographic boundaries.

The two variants of functional meso regions (M-FRD 2A, see Fig. 3; M-FRD 2B, see Fig. 4) result from a single run of the regionalisation procedure based on the cluster analysis approach. Both variants of meso regional systems have been derived from a dendrogram, where the analysis of the procedure based on the Silhouette method offered two possibilities corresponding to the meso level. The highest average values for the Silhouette coefficient correspond for 8 and 12 clusters (i.e. meso regions); in other words, for 8 and 12 clusters there were the most distinct differences between two consequent taxonomic distances in both directions. Inter-regional interactions between constituent functional micro regions are expressed through the values of the CURDS interaction measure (as was noted earlier),

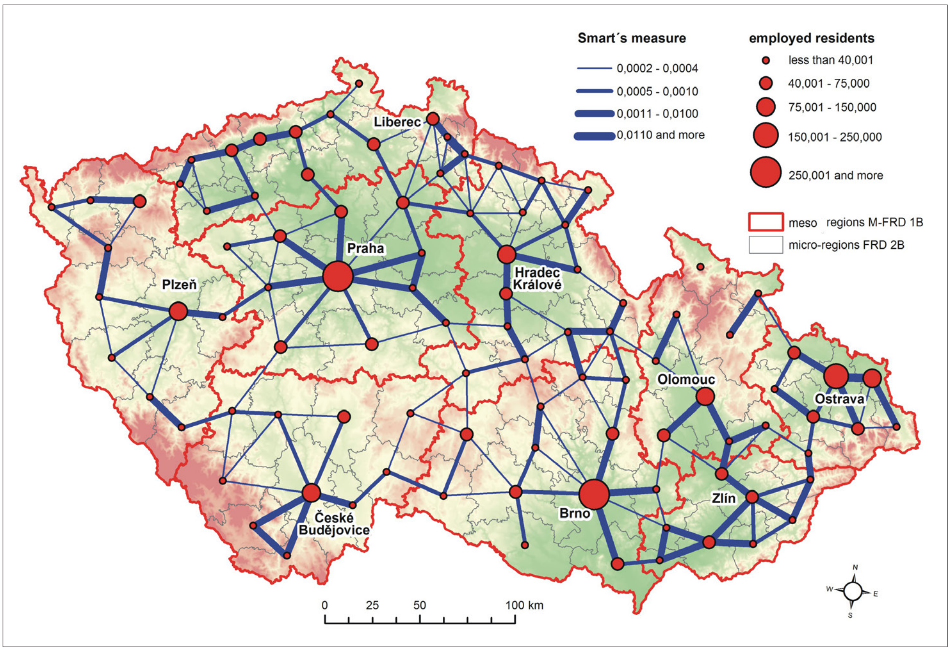


Fig. 2: Regional system M-FRD 1B. Sources: authors' elaboration, Klapka et al. (2014)

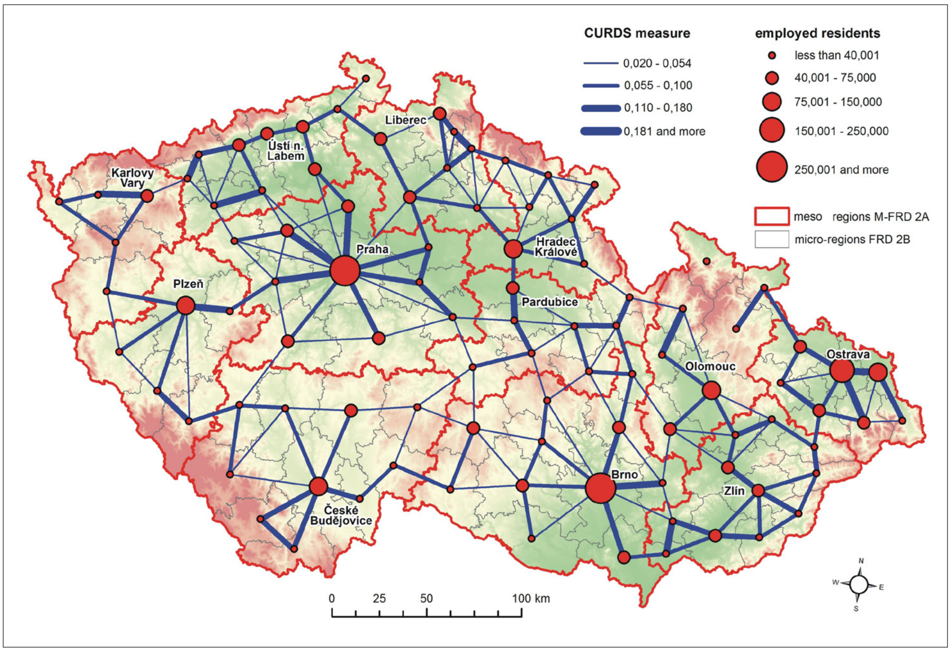


Fig. 3: Regional system M-FRD 2A. Sources: authors' elaboration, Klapka et al. (2014)

because the CURDS measure has been used for adjustments to the input interaction matrix. The variant M-FRD 2A consists of twelve meso regions, which is also the highest number of meso regions reached by the approaches used in this paper. The highest self-containment (0.982) is again in the region of northern Moravia and Silesia (Ostrava). The lowest self-containment is recorded in eastern Bohemia with two meso regions (Hradec Králové: 0.939, Pardubice: 0.926), which have significant mutual cross-border interactions due to the spatial proximity of their largest centres.

The variant M-FRD 2B has amalgamated some meso regions from the variant M-FRD 2A; according to the rank of amalgamation of regions in the resulting dendrogram, the Zlin and Olomouc regions were amalgamated in Moravia, the Hradec Králové and Pardubice regions in the east of Bohemia, the Plzeň and Karlovy Vary regions in the west of Bohemia, and Liberec and Ústí nad Labem regions in the north of Bohemia. This variant consists of eight meso regions, similar to the variant M-FRD 1A, both being very much alike in their spatial patterns. The difference between them lies in the meso regional affinity of some oscillating micro regions, such as Mladá Boleslav, which was removed from central Bohemia (Prague) and assigned to northern Bohemia (Liberec).

The last variant (M-FRD 3, see Fig. 5) has been produced using the Intramax approach. This meso regional system significantly differs from the preceding two pairs of systems. It consists of nine meso regions and arguably the most visible and unusual feature of this system is the spatial extent of the Brno region, which has spread along the historical border between Bohemia and Moravia, including the historically Bohemian micro-regions of Jindřichův Hradec,

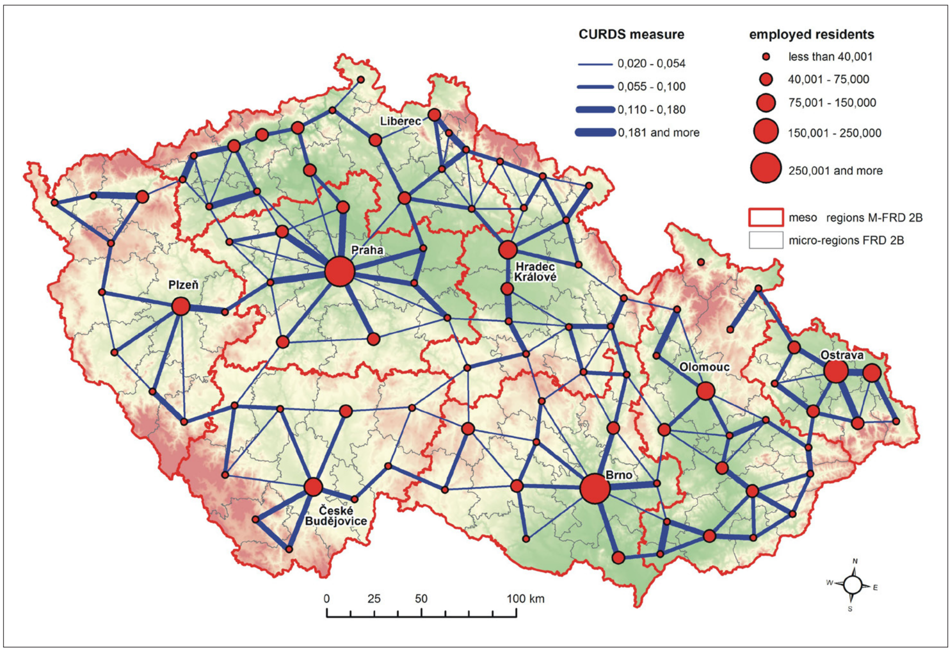


Fig. 4: Regional system M-FRD 2B. Sources: authors' elaboration, Klapka et al. (2014)

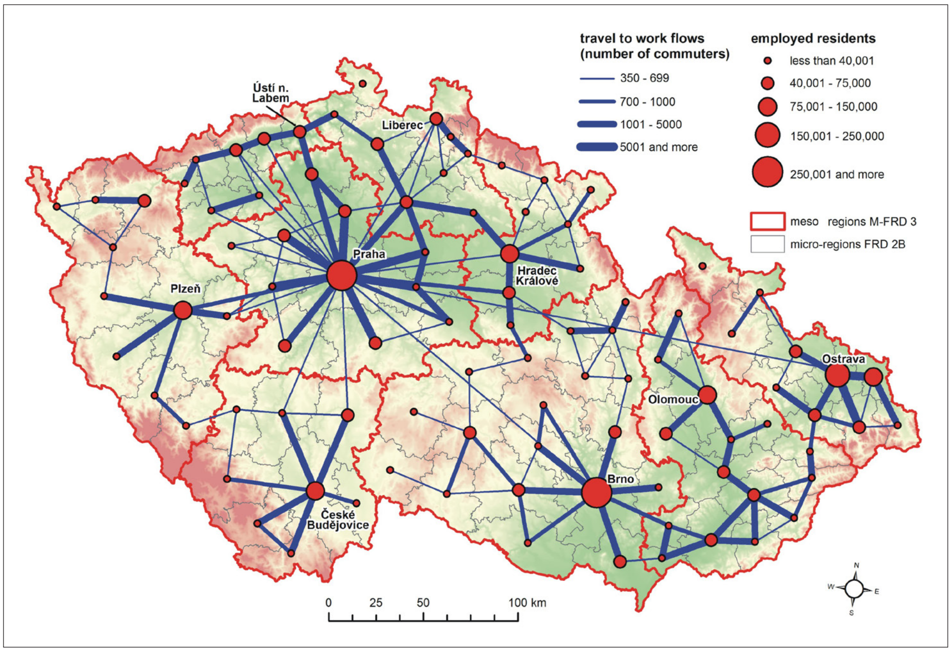


Fig. 5: Regional system M-FRD 3. Sources: authors' elaboration, Klapka et al. (2014)

Pelhřimov, Havlíčkův Brod, Hlinsko, Vysoké Mýto, Česká Třebova, Letohrad, and part of the Bohemian micro region of Svitavy. The spatial extent of northern Bohemia (Liberec) is similarly arguable. This meso region gained the micro regions of Jičín (typically assigned to eastern Bohemia) and Děčín (typically assigned to north-western Bohemia). Unlike the variant M-FRD 2B, central Bohemia (Prague) took over the micro region of Litoměřice from north-western Bohemia (Ústí nad Labem). This shift is responsible for the very low self-containment (0.945) of north-western Bohemia (Ústí nad Labem), however, which is caused by the strong cross-border interactions between the pair of micro regions Ústí nad Labem and Litoměřice.

Figure 6 presents all the functional micro regions that can be understood as oscillating. This means that during the analyses presented in this paper, they belonged to a different meso region in at least one case, when comparing all five functional meso regional systems. Naturally, the cases caused by the basic amalgamation of two meso regions which reflect different totals of meso regions in each system, are excluded from the results presented in Figure 6. It has to be admitted that a crucial role has been played by the variant M-FRD 3 in this respect, and it represents the largest spatial difference in the definitions of meso regions.

The identification of oscillating basic spatial units can be useful in processes which optimise the boundaries of regions and which are based on fuzzy set theory. The fuzziness of TTWA is examined and analysed for instance by Feng (2009), who proposed the optimisation of regional systems through the analysis and identification of the maximum values of the membership function for oscillating

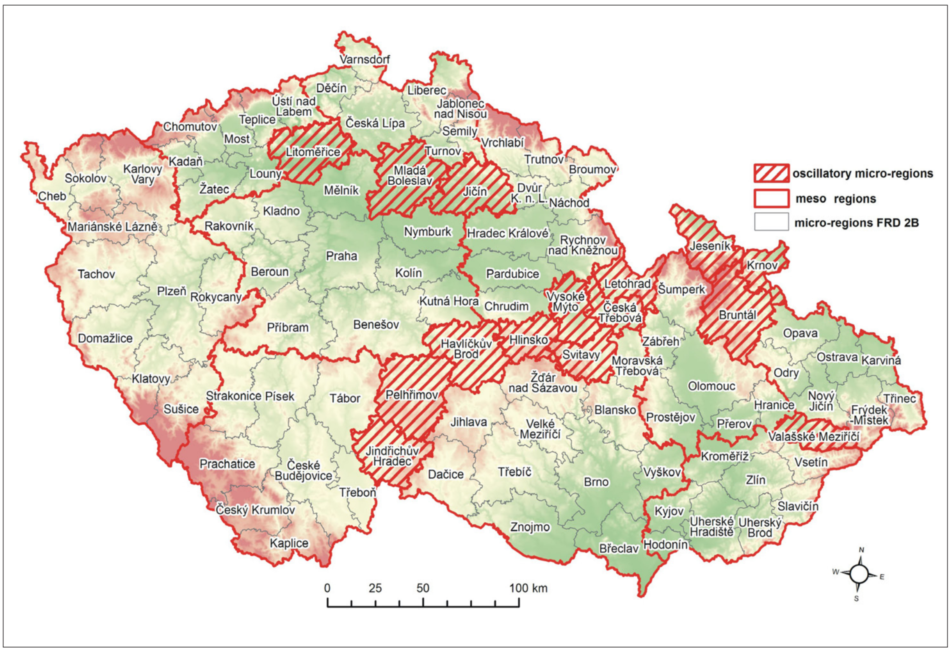


Fig. 6: Oscillating functional micro regions. Sources: authors' elaboration, Klapka et al. (2014)

basic spatial units. A similar principle was used by Kropp and Schwengler (2014) for the optimisation of LLMA.

Four spatial clusters of oscillating micro regions can be identified. The first can be found along the historical Bohemian-Moravian border (see above). The ambiguity in the affinity of these regions is conditioned by frequent and significant historical changes in the administrative division of the Czech Lands, which affected the area along the historical border between Bohemia and Moravia. The change of importance of regional centres is related to the change in commuting patterns. Until the mid of 20th century administrative divisions had respected the historical border. This was violated by the formation of the region around Jihlava. Other regions of this type (Letohrad, Česká Třebová, Svitavy) also changed their regional affinity between the regional centres of eastern Bohemia and Moravia. A certain degree of ambiguity has been preserved until present. The second cluster is formed along the north and north-east border of central Bohemia, where the influence of the capital city Prague and the regional centres of Hradec Králové and Liberec oscillates (the micro regions of Litoměřice, Mladá Boleslav and Jičín). The third spatial cluster can be found along the western part of the border between the historical regions of Moravia and Silesia (micro regions of Jeseník, Krnov and Bruntál), whose meso regional affinity is not quite clear. The last spatial "cluster" is formed only by a single micro region (Valašské Meziříčí), which oscillates between the spheres of influence of the meso regional centres of Ostrava, Zlín or Olomouc, depending on the number of meso regions.

5. Conclusions

Extensive analyses of inter-micro-regional daily travel-to-work flows in the Czech Republic have provided a number of meso regional patterns, out of which five have been selected for the purposes of this paper, on the basis of the total self-containment of a particular meso regional system. Three methods for the definition of meso regions have been applied, based on the labour market area approach, the Intramax approach, and the cluster analysis approach, and previously none of these methods have been used for the Czech Republic. In order to provide sufficient insight into intra-meso regional and inter-meso regional interactions, the flows between constituent micro regions have also been analysed, using three different expressions of these flows.

The LLMA approach used an adjusted third variant of the CURDS regionalisation algorithm (multistage agglomerative procedure), which has produced two variants of the Czech meso regional system (M-FRD 1A and M-FRD 1B). These two variants are based on different parameters regarding the levels of size and self-containment for the resulting functional meso regions. Both variants generally manifest a high degree of similarity; however, the application of multistage methods and the adjustments of their parameters do not produce nested regional systems.

The Intramax approach uses a hierarchical clustering procedure, which gives a single variant of the Czech meso regional system (M-FRD 3). The cluster analysis approach has provided two variants of the Czech meso regional systems (M-FRD 2A and M-FRD 2B). Even though both methods come from the principles of cluster analysis, they differ in linkage measures and in the adjustment of the input interaction (data matrix), and their results are considerably different. Unlike the multistage methods, they are able to provide nested variants for regional systems of different and similar hierarchical levels (see the variants M-FRD 2A and M-FRD 2B). This is secured by the analysis of outputs in the form of a dendrogram.

All five variants for functional meso regional systems of the Czech Republic conform to the normalised categories of either NUTS 2 regions or NUTS 3 regions (the case of variant M-FRD 2A). Generally, it can be concluded that all three methods are suitable for a definition of functional meso regions, even though there are some differences in respective meso regional patterns. In this respect, four spatial clusters of so-called oscillating functional micro regions have been identified. Finally, the functional micro regions have shown that they can act as suitable building blocks for their amalgamation into higher-level hierarchical regions.

Acknowledgement

The paper has been supported by the project of the Czech Science Foundation: "Spatial uncertainty and fuzzy regional systems: identification, analysis and implications of pulsating functional regions" (16-13502S). The authors are grateful for this support.

References:

BALL, R. M. (1980): The use and definition of travel-to-work areas in Great Britain: some problems. Regional Studies, 14(2): 125–139.

BERRY, B. J. L. (1973): Growth centers in the American urban system: Community development and regional growth in the sixties and seventies. Cambridge: Ballinger Pub. Co.

BEZÁK, A. (2000): Funkčné mestské regióny na Slovensku. Geographia Slovaca, 15, Bratislava: Geografický ústav SAV.

BROWN, L. A., HOLMES, J. (1971): The delimitation of functional regions, nodal regions, and hierarchies by functional distance approaches. Journal of Regional Science, 11(1): 57–72.

CASADO-DÍAZ, J. M. (2000): Local labour market areas in Spain: a case study. Regional Studies, 34(9): 843–856.

CASADO-DÍAZ, J. M. (2003): The use of commuting data to define local labour market areas and urban areas in Spain. Paper presented at 7th NECTAR Conference, Umeĺ University, Sweden.

CHRISTALLER, W. (1933): Die zentralen Orte in Süddeutschland. Jena: Fischer.

COOMBES, M. G. (2010): Defining labour market areas by analysing commuting data: innovative methods in the 2007 review of Travel-to-Work Areas. In: Stillwell, J., Duke-Williams, O., Dennett, A. [eds.]: Technologies for migration and commuting analysis: spatial interaction data applications (pp. 227–241). Hershey: IGI Global.

COOMBES, M. G., BOND, S. (2008): Travel-to-Work Areas: the 2007 review. Office for National Statistics, London.

COOMBES, M. G., DIXON, J. S., GODDARD, J. B., OPENSHAW, S., TAYLOR, P. J. (1982): Functional regions for the population census of Great Britain. In: Herbert, D. T., Johnston, R. J. [eds.]: Geography and the Urban Environment. Progress in Research and Applications 5, Chichester: John Wiley, p. 63–112.

COOMBES, M. G., GREEN, A. E., OPENSHAW, S. (1986): An Efficient Algorithm to Generate Official Statistical Reporting Areas: The Case of the 1984 Travel-to-Work Areas Revision in Britain. The Journal of the Operational Research Society, 37(10): 943–953.

De JONG, T., Van der VAART, N. (2010): Manual Flowmap, Faculty of Geographical Sciences, Utrecht University, Utrecht [online]. [cit. 13.10.2015], Available at: http://flowmap.geo.uu.nl/downloads/FM740_Manual.pdf

DROBNE, S., BOGATAJ, M. (2012): Evaluating Functional Regions. Croatian Operational Research Review, 3: 14–26.

DROBNE, S., KONJAR, M., LISEC, A. (2009): Delimitation of Functional Regions Using Labour Market Approach. In: Zadnik Stirn, L., Žerovnik, J., Drobne, S., Lisec, A. (ed.): Proceedings of SOR'09, 10th International Symposium on Operational Research in Slovenia, Slovenian Society Informatika, Section for Operational Research (pp.417– 425). Ljubljana, Slovenia.

ERLEBACH, M., KLAPKA, P., HALÁS, M., TONEV, P. (2014): Inner structure of functional region: theoretical aspects. In: 17th International Colloqium on Regional Science. Conference Proceedings. (Hustopeče 18.–20.6. 2014) (pp. 722–727). Masarykova univerzita, Brno.

FARMER, C. J. Q., FOTHERINGHAM, A. S. (2011): Network-based functional regions. Environment and Planning A, 43(11): 2723–2741.

FENG, Z. (2009): Fuzziness of Travel-to-Work Areas. Regional Studies, 43(5): 707–720.

GUSTAFSON, G. C. (1973): Numerical Classification Procedures. LARS Symposia. Paper 16.

HAGGETT, P. (1965): Locational Analysis in Human Geography. London: Arnold.

HALÁS, M., KLAPKA, P. (2010): Regional division of Czechia on the basis of spatial interaction modelling. Geografie, 115(2): 144–160.

HALÁS, M., ZUSKÁČOVÁ, V. (2013). Travelling to services in the daily urban system of Olomouc. Bulletin of Geography, Socio-economic Series, 20: 23–41.

HALÁS, M., KLAPKA, P., TONEV, P. (2014a): A contribution to human geographical regionalisation of the Czech Republic at the mezzo level. In: 17th International Colloqium on Regional Science. Conference Proceedings. (Hustopeče 18.–20.6. 2014) (pp. 715–721). Masarykova univerzita, Brno.

HALÁS, M., KLAPKA, P., BLEHA, B., BEDNÁŘ, M. (2014b): Funkčné regióny na Slovensku podľa denných tokov do zamestnania. Geografický časopis, 66(2): 89–114.

HALÁS, M., KLAPKA, P., BEDNÁŘ, M., TONEV, P. (2015): An alternative definition and use for the constraint function for rule-based methods of functional regionalisation. Environment and Planning A, 47(5): 1175–1191.

HALL, P. G., HAY, D. (1980): Growth centres in the European urban system. London: Heinemann.

HOLMES, J. H., HAGGETT, P. (1977): Graph theory interpretation of flow matrices: a note on maximization procedures for identifying significant links. Geographical Analysis, 9(4): 388–399.

ISARD, W. (1956): Location and space economy: A general theory relating industrial location, market areas, land use, trade and urban structure. New York: John Wiley.

KARLSSON, C., OLSSON, M. (2006): The identification of functional regions: theory, methods, and applications. The Annals of Regional Science, 40(1): 1–18.

KLAPKA, P., HALÁS, M., TONEV, P. (2013a): Functional regions: concept and types. In: 16th International Colloquium on Regional Sciences, Conference Proceedings (pp. 94–101). Brno: Masaryk University.

KLAPKA, P., HALÁS, M., TONEV, P., BEDNÁŘ, M. (2013b): Functional regions of the Czech Republic: comparison of simple and advanced methods of regional taxonomy. Acta Universitatis Palackianae Olomucensis, Facultas Rerum Naturalium, Geographica, 44(1): 45–57.

KLAPKA, P., HALÁS, M., ERLEBACH, M., TONEV, P., BEDNÁŘ, M. (2014): A multistage agglomerative approach for defining functional regions of the Czech Republic: The use of 2001 commuting data. Moravian Geographical Reports, 22(4): 2–13.

KONJAR, M., BOYANDIN, I., LALANNE, D., LISEC, A., DROBNE, S. (2010): Using flow maps to explore functional regions in Slovenia. The Proceedings of the 2nd International Conference on Information Society and Information Technologies. Dolenjske Teplice.

KROPP, P., SCHWENGLER, B. (2014): Three-Step Method for Delineating Functional Labour Market Regions. Regional Studies, 50(3): 429–445.

LANDRÉ, M. (2012): Geoprocessing Journey-to-Work Data: Delineating Commuting Regions in Dalarna, Sweden. ISPRS Int. J. Geo-Inf, 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): 1–19.

LÖSCH, A. (1940): The Economics of Localization. Jena, Fischer.

MASSER, I., BROWN, P. J. B. (1975): Hierarchical aggregation procedures for interaction data. Environment and Planning A, 7(5): 509–523.

MASSER, I., SCHEURWATER J. (1980): Functional regionalisation of spatial interaction data: an evaluation of some suggested strategies. Environment and Planning A, 12(12): 1357–1382.

MITCHELL, W. F., WATTS, M. J. (2010). Identifying functional regions in Australia using hierarchical aggregate techniques. Geographical Research, 48(1): 24–41.

NYSTUEN, J. D., DACEY, M. F. (1961): A graph theory interpretation of nodal regions. Regional Science Association, Papers and Proceedings, 7(1): 29–42.

OECD (2002): Redefining Territories: The Functional Regions. Paris: OECD.

PAPPS, K. L., NEWELL, J. O. (2002): Identifying functional labour market areas in New Zealand: a reconnaissance study using travel-to-work data. Discussion Paper 443. Bonn: Institute for the Study of Labor.

PHILBRICK, A. K. (1957): Principles of areal functional organization in regional human geography. Economic Geography, 33(4): 299–336.

ROUSSEEUW, P. J. (1987): Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics, 20: 53–65.

SMART, M. W. (1974): Labour market areas: uses and definition. Progress in Planning, 2(4): 239–353.

SÝKORA, L., MULÍČEK, O. (2009): The micro-regional nature of functional urban areas (FUAs): lessons from the analysis of Czech urban and regional system. Urban Research and Practice, 2(3): 287–307.

THÜNEN, von J. H. (1826), Der Isolierte Staat in Beziehung auf Landschaft und Nationalökonomie. Trans. By C. M. Wartenberg (1966): Von Thünen's Isolated State. Oxford: Pergamon Press.

TONEV, P. (2013): Změny v dojížďce za prací v období transformace: komparace lokálních trhů práce. Diploma thesis. Brno: Masaryk University.

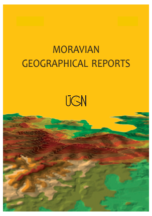
WARD, J. H. (1963): Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association, 58: 236–244.

WATTS, M. (2009): Rules versus hierarchy: an application of fuzzy set theory to the assessment of spatial grouping techniques. In: Kolehmainen, M., Toivanen, P. J., Beliczynski, B. [eds.]: ICANNGA 2009, LNCS, 5495: 517–526.

WONG, D. (2009): The modifiable areal unit problem (MAUP). In: Fotheringham, A., Rogerson, P. [eds.]: The SAGE handbook of spatial analysis (pp. 105–125). London: SAGE Publications Ltd.

Please cite this article as:

ERLEBACH, M., TOMÁŠ, M., TONEV, P. (2016): A functional interaction approach to the definition of meso regions: The case of the Czech Republic. Moravian Geographical Reports, 24(2): 37–46. Doi: 10.1515/mgr-2016-0009.



MORAVIAN GEOGRAPHICAL REPORTS

DE GRUYTER OPEN

Institute of Geonics, The Czech Academy of Sciences

journal homepage: http://www.geonika.cz/mgr.html

doi: 10.1515/mgr-2016-0010

The efficiency of areal units in spatial analysis: Assessing the performance of functional and administrative regions

Pavel KLAPKA [a *], Marián HALÁS [a], Pavlína NETRDOVÁ [b], Vojtěch NOSEK [b]

Abstract

An attempt to provide a procedure for the assessment of the efficiency of various regional systems for the purposes of spatial analysis is presented in this paper. Functional regions as well as approximated functional regions and the existing administrative regions in the Czech Republic are evaluated, as examples of regional systems to be compared and assessed. Functional regions and approximated functional regions are defined according to the adjusted third variant of the CURDS regionalisation algorithm, using the latest knowledge on the operation of the constraint function. The comparisons of individual regional systems are based on LISA maps and particularly on the assessment of regional variability, including the measures of internal homogeneity and external variability in the regional systems.

Keywords: *spatial analysis, functional regions, administrative regions, regional variability, LISA, Czech Republic*

Article history: *Received 4 November 2015; Accepted 10 May 2016; Published 30 June 2016*

1. Introduction

This paper attempts to provide a procedure which uses the concept of functional region to assess the efficiency of agglomerated areal units for the purposes of spatial analysis, particularly for the assessment of regional variability. Each grouping of arbitrary basic spatial units into larger regions is part of the so-called modifiable areal unit problem (MAUP – Openshaw, 1984; Fortheringham and Wong, 1991; Unwin, 1996; Grasland et al., 2006). As there is an extensive number of ways to organise basic spatial units into regions, the question of identification of an optimal or near-optimal solution is raised. If an inappropriate solution is chosen, important characteristics of the spatial distribution of geographical phenomena may remain concealed, and the application of such regions for spatial analysis of phenomena such as regional inequalities would be compromised in such a case.

Spatial analysis is a set of techniques and models that explicitly use various scales of reference and data related to phenomena and objects/cases in a spatially arrayed manner. The 'correct' grouping of these objects into more "manageable" cases can be seen as a necessary precondition for sensitive spatial analyses that explain or predict the spatial distribution of geographical phenomena. One of the objectives of spatial analysis can be the identification and characterisation of areal units (as regions, for example) that manifest either a higher degree of internal homogeneity and external separation (formal regions), or a large degree of internal cohesion and external self-containment in relation to other areal units or regions (functional regions).

This paper examines two types of regions: functional regions and administrative regions. Correctly-defined functional regions (i.e. those based on informed choices) can serve better as a geographical tool for administrative use than unsuitably- and arbitrarily-delineated administrative regions, which has been acknowledged long ago by Haggett (1965) and Dziewoński (1967). It can be assumed that suitable administrative regions, particularly at a micro-regional level, should be based on functional spatial relations. In addition, it can be generally assumed that functional regions better capture the geographical variability of spatial information for spatial analyses. If similar measurements of geographical variability are obtained for administrative regions, then it would indicate that such administrative regions are defined according to spatial functionality and suitable for spatial analyses. The hypothesis of this paper is that a regional system of functional regions (or approximated functional regions) should manifest at least the same value

[a] Department of Geography, Faculty of Science, Palacký University Olomouc, Czech Republic (*corresponding author: P. Klapka, e-mail: *pavel.klapka@upol.cz*)

[b] Department of Social Geography and Regional Development, Faculty of Science, Charles University in Prague, Czech Republic

(or higher) for a measure of internal homogeneity and external variability. A procedure that enables one to assess the suitability of administrative units for spatial analyses is expected to be the main methodological outcome of the work. Several sets of functional regions, based on the 2011 census data for the Czech Republic are expected to be the applied outcome of the paper. If functional regions have higher values for internal homogeneity and external variability, the administrative division would have certain insufficiencies regarding the principles of spatial efficiency and equity. If these values were comparable, administrative regions could be considered as being defined according to spatial functionality and as suitable for spatial analyses.

Given these introductory remarks, the main objective of the paper is to evaluate the efficiencies and suitability of administrative regions for spatial analysis. In order to fulfil this objective, two steps have to be carried out. First, as a tool for further analyses and the introductory objectives, functional regions based on daily travel-to-work flows have to be defined. Several sets of optimised functional regional systems and approximated functional regional systems are produced to serve as bases for comparison. Second, the analyses of regional variability of a set of selected variables for each regional system will be carried out. This paper will use the territory of the Czech Republic as the study area, and will analyse selected socio-economic characteristics in terms of their regional variability, in two administrative systems ('districts', and 'Areas of Municipalities with Extended Powers': AMEP) and five functional regional systems (three of them consisting of optimised ["natural"] functional regions, and two of which consist of approximated functional regions, with the latter approximation taking into consideration the number of districts and AMEPs).

The remainder of the paper is organised in the following way. Section 2 presents some necessary theoretical background regarding the issue of functional regions, administrative regions and regional variability. Section 3 describes the methods that are applied in the paper regarding the delineation of functional regions and the measurement of regional variability. Section 4 presents the results and necessary comments. In the conclusion the paper returns to the objectives and hypothesis.

2. Theoretical background

The paper theoretically builds on three presuppositions that are interlinked, the common denominator being that some units for spatial analysis are needed and are sought, and this has its practical purposes (such as the reporting of statistical data in general (e.g. the Census) and for planning purposes – the two most general). Statistical data in geography have mostly an aggregated character; they are composed of attributes referring to an individual, who has a position in geographic space. The first presupposition is that there are two options with regard to the units of spatial analysis: either the existing administrative regions can be used, or regions based on particular criteria have to be defined. With reference to the latter, this paper has opted to use functional regions based on daily travel-to-work flows, and to compare them to the existing administrative units. The second presupposition is based on the general fact that all geographical regions which consist of some arbitrary basic spatial units face the general problem of how many regions there should be optimally, and how these regions should be composed from basic spatial units (MAUP; see section 2.2). Finally, the third presupposition is based on the belief or the requirement that the spatial uncertainty stemming from the MAUP be reduced as much as possible. Thus, there are two types of regional systems (administrative and functional) and the need to decide on their suitability for other purposes is paramount: in order to achieve such suitability, the analysis of regional variability within both systems appears to be a convenient procedure.

2.1 Functional and administrative regions

A functional region is regarded in this paper as it is in our preceding research (e.g. Klapka et al., 2013a, 2014; Halás et al., 2015). It is a general concept that has to meet only the condition of the self-containment of region-organising horizontal interactions or flows. This means that these horizontal functional relations should be maximised within a region and minimised across its boundaries, so that the principles of internal cohesiveness and external separation regarding spatial interactions are met (see for instance, Smart 1974; Karlsson and Olsson, 2006; Farmer and Fotheringham, 2011). Sometimes functional regions can be seen as nodal regions, i.e. regions defined and identified by the core-periphery dichotomy. Such nodal regions, however, also very often fulfil the condition of self-containment and they can be regarded as a more specific concept, a subset for a functional region (see Klapka et al., 2013a). As the interactions come from human activities, functional regions can be seen as representative spatial images or imprints for relevant aspects of the (aggregated) spatial behaviour of individuals (Halás et al., 2015). The delineation of functional regions is mostly based on the analysis of statistical data, particularly daily travel-to-work flows (e.g. Goodman, 1970; Casado-Díaz and Coombes, 2011). These flows represent a residence-workplace daily rhythm of spatial behaviour and as such are the most frequent regular movements for a large part of the population (Hanson and Pratt, 1992; Heldt Cassel et al., 2013; Halás et al., 2015).

Administrative regions are usually strictly defined on the basis of rigorous rules and criteria and are used for normative purposes. One can assume that it should be of the utmost importance that they reflect an existing geographical reality (spatial behaviour of individuals, spatial patterns of their movements). If this is done, the inhabitants of respective areas will find their administrative region, particularly its centre with all the necessary offices and public services, localised in a space which they frequently use in their daily rhythms. Their ties to such regions exist objectively and are considerably strong. All this can also result in the strengthening of their emotional ties to a space. If such a spatial pattern and design is achieved, other geographical factors and characteristics can reflect and follow this arrangement, such as transport infrastructure, the distribution of public transport lines in space and time, etc.

Apart from the above-mentioned functional relationships, the construction of administrative regions also takes into account other auxiliary criteria, such as historical precedents, the existence of natural borders and barriers, and the spatial distribution of national and other population groups, inter alia. This is not always the case, however, and in some cases administrative systems are not well designed for political reasons or just because the rules or norms are unsuitable or they are designed on purpose (as is the case of Slovakia: see for example Buček, 2002, 2005; or Romania: Suciu, 2002). The delineation of administrative regions can be negatively affected by several risks, which have potentially opposite effects to the delineation of administrative regions compared to definitions using a functional approach. For example, such risks are political influence, nationalistic motives, economic motives, etc. In this respect the risk of gerrymandering is among the first to arise (see for example, Bunge, 1966; Johnston, 2002; Moore, 2002; Suciu, 2002; Apollonio et al., 2009). As administrative regions also frequently serve as statistical areas, their unsuitable delineation can distort statistical spatial analyses in many cases – and in statistically unknowable ways.

In theory, the definition of administrative regions should respect three basic principles with regard to a space: spatial efficiency, spatial equity, and spatial stability (Bezák, 1997, who builds upon the concepts put forward by Goodall, 1987; Michniak, 2003; Halás and Klapka, 2012; Klapka, et al., 2014). The principle of spatial efficiency states that the administrative geography of a territory should reflect the population distribution and its spatial behaviours (particularly spatial movements) to the greatest possible extent. Here is a clear connection to the concept of a functional region. The principle of spatial equity is based on the assumption that administrative centres should be equally accessible from the most peripheral parts of each administrative region. Finally, the principle of spatial stability requires that the administrative geography (e.g. boundaries of administrative units) of a territory should be stable over time.

The principle of spatial efficiency can sometimes be in contrast to the principle of spatial equity, because large regional centres usually tend to form much larger hinterlands than smaller regional centres. In this case, it is necessary to balance the opposite demands of the two principles. If functional regions are to be used as administrative regions, the principle of spatial equity should prevail. This requirement can be secured in concrete functional regionalisation tasks by relativising the interaction data, for instance by the use of Smart's interaction measure (Smart, 1974; Casado-Díaz, 2000; Klapka, et al., 2014). Similarly the principle of spatial stability can be in contrast to the two above-mentioned principles. This is the case when a biased administrative division does not respect natural patterns of settlement and regional systems and the interactions occurring in them. It is also appropriate to note that regions defined according to daily travel-to-work flows can change over time (Ozkul, 2014). Therefore, a compromise between the principle of spatial stability on the one hand and the principles of spatial efficiency and equity on the other should be reached in legitimate cases, and revisions to administrative divisions should be made only in the most necessary cases.

2.2 Considerations on the assessment of relations between spatial distribution patterns and regional variability

A geographic space is non-homogeneous, both in vertical and horizontal directions. This inherent quality of space forms the basis for the study of spatial distribution patterns and regional variability. There is also an inherent temporal dimension. The assessment and analysis of such variability, however, relies to a considerable extent on the character and availability of relevant statistical information. Geographers often work with data that are spatially referenced and aggregated. The reason is twofold. First, secondary data are reported for some kind of arbitrary spatial units (e.g. census tracts, municipalities etc.) and they do not have the character of unique objects (statistical individuals). Second, it is useful to report primary or individual statistical data for certain kinds of spatial units, otherwise the analyses of their spatial distribution would be impossible or methodologically incorrect. In both cases, however, there is a possibility of spatial bias, which can compromise the spatial analyses and the interpretation of the results, because there is an almost infinite number of ways to aggregate individual pieces of statistical information into spatial units, zones and regions, and it has to be decided or estimated which spatial design better follows the spatial functionality principle and is thus more suitable.

This is referred to as a modifiable areal unit problem (MAUP); it has been identified by Gehlke and Biehl (1934), extended by Yule and Kendall (1950) and discussed thoroughly by Openshaw (1984), Fotheringham and Wong (1991), Unwin (1996) and Grasland, et al. (2006). MAUP consists of two issues demanding attention: the first concerns the number of spatial units and is referred to as a scale effect (Openshaw, 1984); and second concerns the issue of alternative aggregations at the same or similar scales and is referred to as a zoning or aggregation effect (Openshaw, 1984).

There are ways to tackle the MAUP, or, more precisely, how to choose from various solutions to spatial designs, in terms of which one is more suitable for a given purpose and which one is less suitable (see for example, suggestions made already by Openshaw, 1977, 1984; Fotheringham and Wong, 1991). This paper does not tackle the problem fully in the first place, because the objective is not to define functional regions using quantitative methods so that their internal homogeneity and external variability is maximised. Consideration of MAUP cannot be avoided, however, because the paper compares the internal homogeneity and external variability of the existing normative administrative regions to the optimised and approximated functional regions, defined on the basis of daily movements of the population. Three types of regional variability measures can be generally applied in this respect: inter-regional variability, intra-regional variability (internal homogeneity), and relative regional variability inequality (for methods, see section 3).

Another inherent quality of space, the horizontal distance (either absolute or relative) between geographical locations, raises the question of whether neighbourhood matters or not, when assessing the spatial distribution of geographical phenomena. It is generally agreed that it does (Goodchild, 1986: 3), which means that the values for a certain characteristic in one location, in one spatial unit, are affected by the values of this characteristic in neighbouring locations and neighbouring spatial units (see for instance Cliff and Ord, 1973; Goodchild, 1986; Anselin, 1995; Getis, 2008). This spatial dependency is considered to be an inherent feature of spatial data and reflects such basics as Tobler's 'first law of geography' (Tobler, 1970: 236; 2004) and the role of distance in the probability of contacts between geographical locations (distance-decay functions). Spatial dependency can be measured by spatial autocorrelation statistics, which can be expressed both by global and local indices (see for instance, Anselin, 1995; Spurná, 2008; Netrdová and Nosek, 2009 in the Czech literature). While the global indices enable us to quantify the extent of spatial clustering of similar values in a space with one value, the results of local indices can be depicted on a map and used to identify spatial clusters and outliers. In the context of this paper, the global statistic of spatial autocorrelation, Moran's I (Cliff and Ord, 1973; Anselin, 1988), is important for the selection of studied characteristics according to their different level of spatial concentration. LISA analysis, the local statistic of spatial autocorrelation, is interesting for its comparisons of how defined regional systems conform to the actual spatial patterns of selected geographical characteristics.

3. Methods and data

3.1 Functional regions

A detailed overview of methods for the definition of functional regions is beyond the scope of this paper. Relevant discussions can be found for example in Coombes (2000), van der Laan and Schalke (2001), Flórez-Revuelta, et al. (2008); Casado-Díaz and Coombes (2011), Farmer and Fotheringham (2011), and in our own earlier papers (Klapka et al., 2013b, 2014; Halás et al., 2014, 2015). This paper favours the use of the so-called rule-based, or multistage approach to functional regionalisation that was introduced to this field of study by Smart (1974) and later extended at the Centre for Urban and Regional Development Studies (CURDS) in Newcastle, UK (Coombes, et al., 1982, 1986). In this paper, the third variant of the CURDS regionalisation algorithm (Coombes and Bond, 2008; Coombes, 2010) is applied using the constraint function proposed and used by Halás et al. (2015), and which has already been tested practically (Halás et al. 2014; Klapka et al., 2014), but only on the 2001 census data and using the second variant of the CURDS algorithm (Coombes et al., 1986). This is the first time the third variant of the CURDS regionalisation algorithm has been applied to the territory of the Czech Republic. The method identifies as many functional regions as possible, according to the criteria set by the regionalisation algorithm.

The identification of functional regions is based on the analysis of spatial patterns of daily travel-to-work flows using the 2011 population census. These data have been stored in a large and sparse 6,251 × 6,251 non-symmetrised (flows: $t_{ij} \neq t_{ji}$) matrix, for the municipalities of the Czech Republic that served as basic spatial units for all analyses presented in this paper. It is very important to note that the diagonal of the matrix included intra-unit flows (t_{ii} – in fact it is the number of employed residents working locally).

A crucial role in the regionalisation algorithm is played by the constraint function. It sets the minimal size and self-containment criteria for the resulting functional regions and it also comprises the trade-off between the two parameters. The trade-off means that smaller regions have to reach a higher level of self-containment, while, in contrast, larger regions are allowed to manifest a lower level of self-containment. The constraint function is in the form of a continuous curve and its shape is determined by five parameters (see below), four of which can be easily estimated. The notation of the constraint function is:

$$\frac{T_{jj}}{\left(\sum_k T_{jk} + \sum_k T_{kj}\right) - T_{jj}} - \frac{\alpha(\beta_4 - \beta_3)(\beta_2 - \beta_1)}{\left|\max\left(\beta_3 + 1; \sum_k T_{jk}\right) - \beta_3\right|} \geq \beta_1 \qquad (1)$$

where β_1, β_2, β_3, β_4 are limits of the trade-off between the size and self-containment of a region (β_1, β_2 are lower and upper limits of the self-containment; β_3, β_4 are lower and upper limits of the size), and α determines the measure of the deflection of the trade-off part of the function ($\alpha = 0.09$ in this paper). For remaining expressions see the notation of the Smart's measure (2) below.

As proposed by Halás et al. (2015), the constraint function can be used for the identification of the relatively optimal number of resulting functional regions through the estimation of four beta parameters. The analysis starts with loose values for these parameters, which produce a larger number of functional regions, and which provide the initial spatial pattern (in this paper $\beta_1 = 0.5$, $\beta_2 = 0.55$, $\beta_3 = 2{,}000$, $\beta_4 = 10{,}000$). These regions can be plotted on a graph according to the self-containment and size variables. The graph also contains the constraint function and the regions appear in its upper right sector. If there is a considerable gap in the field of points, a new constraint function can be inserted and the values for the new beta parameters can be estimated. This step can be repeated several times and thus it can provide several variants of the optimised regional system (3 in the case of this paper). Of course, this operation can be used for the identification of a given number of regions, i.e. approximated regional systems can be defined in this way (2 in the case of this paper). The parameters for all regional systems, including their total self-containment[1], are presented in Table 2. Detailed theoretical and methodological discussion of the constraint function and its use is provided in Halás et al. (2015), and operations with the constraint function were applied by Halás et al. (2014) and Klapka et al. (2014).

The interaction measure used for the expression of the strength of the relationship between two basic spatial units (or a basic spatial unit and a "proto" region) was recommended, but not used, by Smart (1974). This measure is currently the most frequently used for the type of research tasks presented in this paper (see for example, Casado-Díaz and Coombes, 2011). It is mathematically the most correct way for the relativisation and symmetrisation of two-dimensional interaction data. This measure levels the size differences between the regions and thus it is the most suitable compromise between the principles of spatial efficiency and equity.

The notation of the Smart's measure is

$$\frac{T_{ij}^2}{\sum_k T_{ik} \sum_k T_{kj}} + \frac{T_{ji}^2}{\sum_k T_{jk} \sum_k T_{ki}} \qquad (2)$$

where T_{ij} is a value for a flow from the municipality i to the municipality j, T_{ji} is a value for a flow from the municipality j to the municipality i, and k is the total number of basic spatial units (municipalities) in the system.

Finally, the procedure for the identification of functional regions of the Czech Republic consists of the following steps:

1. all basic spatial units are ranked in descending order according to the values of the constraint function and are considered to be so-called "proto" functional regions;
2. if all regions equal or exceed the value of the β_1 parameter in the constraint function, the procedure stops, otherwise it proceeds to the next step;
3. the "proto" functional region with the lowest rank according to the value of the constraint function is dissolved into its constituent basic spatial units (municipalities) and these are ranked in descending order according to the constraint function;

[1] The total self-containment of the regional system is calculated as $\frac{\sum_j T_{jj}}{\sum_j \sum_k T_{jk}}$.

4. the constituent basic spatial unit with the highest rank is amalgamated with the "proto" functional region that it is most strongly related to according to the interaction measure (see further); and
5. after each amalgamation the values for the constraint function are recalculated and the procedure returns to the first step.

3.2 Spatial distribution patterns and regional variability

There are different types of geographical characteristics in a spatial and regional context, which have different spatial patterns and are influenced by different spatial and regional processes. The basic typology of the possible nature of characteristics based on spatial and regional concentration is shown in Table 1. Three of four types of characteristics can be found in reality, the characteristic with high regional concentration and low spatial concentration does not exist because a high regional concentration always implies a level of spatial concentration. We have analysed 17 characteristics from the 2011 census at the municipal level through both global and local spatial autocorrelation statistics and the regional decomposition of variability.

According to this typology, 4 geographical characteristics have been selected for analysis in this paper, in terms of their distinctive spatial distribution patterns and relative regional variability. Two characteristics with high spatial concentration are closely connected to the data used for the definition of functional regions, i.e. with the economic activity of the population: unemployment rate, and employment rate in agriculture. While the unemployment rate exhibits a relatively high regional concentration at all hierarchical regional levels, employment in agriculture is a specific characteristic influenced more by physical conditions than a regional structure based on socioeconomic relations. The two remaining characteristics with low regional concentration differ in their spatial concentration and do not manifest such a close connection for methodological reasons: average years in education, and the age preference index. The basic typology of the data used including their definitions is presented in Table 1. All the data for municipalities were obtained from the 2011 census.

For the purposes of comparison between the sets of the existing normative administrative regions and the optimised and approximated functional regions, a minor adjustment had to be made. As the four largest cities of the Czech Republic (Prague, Brno, Ostrava, and Plzeň) have their own normative administrative units, these cities are treated separately from their functional regions in all five sets of optimised and approximated regional systems in the parts of the paper dealing with the assessment of the regional variability of the four above-mentioned geographical characteristics.

The basic spatial patterns of the characteristics studied are introduced using local spatial autocorrelation, specifically LISA cluster maps (local indicators of spatial association) (Anselin, 1995). Based on the LISA methodology, we can categorize the municipalities with significant local spatial clustering into four categories. If a municipality, as well as its surrounding (geographically close) municipalities, has an above-mean value and the relationship is statistically significant, a cluster (hot spot or high-high type in this case) is formed. Besides hot spots, there are cold spots (low-low clusters), high-low (high values surrounded by low values), and low-high (low values surrounded by high values) outliers. If the relationship between the close municipalities is not significant according to tests based on the comparison between observed and expected values for the local Moran's I statistics and the computation of z-scores, then no clusters or outliers are identified.

In spatial autocorrelation analysis it is important to operationalize geographical proximity using the matrix of spatial weights. In this paper, the distance-based spatial weight matrices are not chosen arbitrarily, but with respect to analyses of global spatial autocorrelation. Firstly, for each variable, Moran's I (Cliff and Ord, 1973; Anselin, 1988) is calculated for a series of distances. Then the LISA cluster maps are constructed using the spatial weight matrix with the maximum z-score. With regard to the definition of regions, the highest values of z-score identify the level (geographical distance) at which the process operates most significantly. Thus, selected geographical characteristics can be attributed to specific regional levels. By using a z-score which reflects

	Regional Concentration HIGH	**Regional Concentration LOW**
Spatial Concentration HIGH	SPATIALLY dependent and bounded in REGIONS Concentrations in regions *– Unemployment rate – based on a labour market delimitation, which highly corresponds with regional levels in the Czech Republic*	SPATIALLY dependent with no relation to REGIONS Concentrations across regional borders *– Employment in agriculture – determined to a large extent by physical geography* *– Average years in education – concentrated in larger settlements*
Spatial Concentration LOW		Both SPATIALLY and REGIONALLY independent No concentrations *– Age preference index – as a demographic characteristic relatively regularly distributed in space*

Tab. 1: General "spatial and regional" typology of characteristics used in the analysis
Source: Nosek and Netrdová (2014) – modified
Notes: (1) The unemployment rate is computed as the ratio of unemployed to the economically active population; (2) Employment in agriculture as a ratio of employed in agriculture to the total number of the employed population; (3) The average years in education as a weighted mean of the ratio of educated people at different stages in their education and the number of years needed to achieve this level of education; (4) The age preference index as a ratio of the population older than 64 years to the population younger than 15 years.

the intensity of spatial clustering for the identification of the optimal spatial weight matrix, the final LISA cluster map with the highest significance shows the largest clusters for each characteristic.

Sets of normative and functional regional systems are compared through:

1. a measure of regional variability (differences between regional means);
2. a measure of relative regional variability (importance of the regional level compared with the overall inter-municipal variability); and
3. a measure of the internal homogeneity of regions (variability within regions).

For further description of different concepts of regional variability, see Nosek and Netrdová (2014).

Regional variability is measured using standard variability measures such as the coefficient of variation, and the Theil index. All these measures are analysed both in unweighted and weighted forms. The unweighted measures treat all regions the same, no matter how large they are in terms of their size. The weighted measures take some measure of size into account (see note below Table 3). Similarly, homogeneity (intra-regional variability) is measured by both weighted and unweighted coefficients of variation.

Relative regional variability is measured by the Theil index decomposition. Of the standard variability measures, the Theil index is scale independent and decomposable, similar to the variance (Cowell and Jenkins, 1995; Shorrocks and Wan, 2005). The main purpose of the Theil index decomposition is to calculate both inter-regional (between regional) variability (T_B) and intra-regional (within regional) variability (T_W). By comparing inter-regional variability (T_B) with the overall variability ($T_B + T_W$), the importance of respective regional levels can be quantified. These results are skewed to some extent, however, by stochastic variability, which appears irrespective of the design of regional patterns. Thus, a geographical standardization is introduced, which can filter out the stochastic component and isolate the contextual component (for details including formulas, see Novotný and Nosek, 2012). This filtering and isolation is used also in this paper.

4. Results

Basic statistical characteristics for five regional schemes are presented in Table 2: for three variants of optimised functional regions (functional regions according to daily travel-to-work flows – FRD); and for two variants of approximated functional regions (AFRD). Regional system AFRD 1 approximates the number of AMEPs, and regional system AFRD 2 approximates the number of districts in the Czech Republic. Delimitation of regions for regional systems is presented in Figures 1–5. For a comparison with the results from the 2001 census, see Klapka et al. (2014).

The overall spatial distribution of four selected characteristics regarding various manifestations of the neighbourhood effect is presented in Figure 6. Types of spatial autocorrelations are laid over the mean variant of the optimised functional regional system (FRD 2). The unemployment rate shows clusters of low unemployment in a belt stretching from south-western Bohemia through central Bohemia to north-eastern Bohemia. Clusters of high unemployment are particularly concentrated in problematic regions of north-western Bohemia and peripheral areas of Moravia and Silesia. Employment in agriculture presents a high degree of clustering, but without relation to the borders

Attribute for regional system		FRD 1	FRD 2	FRD 3	AFRD 1	AFRD 2
β_1 value		0.60	0.60	0.65	0.56	0.65
β_2 value		0.65	0.65	0.70	0.70	0.80
β_3 value		7,500	6,000	11,500	2,500	7,500
β_4 value		15,000	100,000	30,000	25,000	120,000
Self-containment of regional system		0.908	0.916	0.926	0.896	0.930
No. of regions		142	125	95	201	80
Self-containment	Mean	0.802	0.820	0.841	0.776	0.857
	Median	0.809	0.824	0.857	0.778	0.861
	Coeff. of variation	0.097	0.080	0.076	0.094	0.060
Economically active population	Mean	28,434	32,217	42,501	20,087	50,585
	Median	16,843	19,149	27,217	10,029	34,973
	Coeff. of variation	2.016	1.955	1.715	2.420	1.541
Population	Mean	74,381	84,497	111,181	52,548	132,027
	Median	46,989	54,368	76,305	29,290	95,159
	Coeff. of variation	2.037	1.761	1.547	2.185	1.386
Area km^2	Mean	555.39	630.93	830.17	392.37	985.83
	Median	463.81	504.63	734.01	343.17	849.91
	Coeff. of variation	0.585	0.579	0.568	0.600	0.471

Tab. 2: Attributes for variants of regional system
Source: authors´ computations

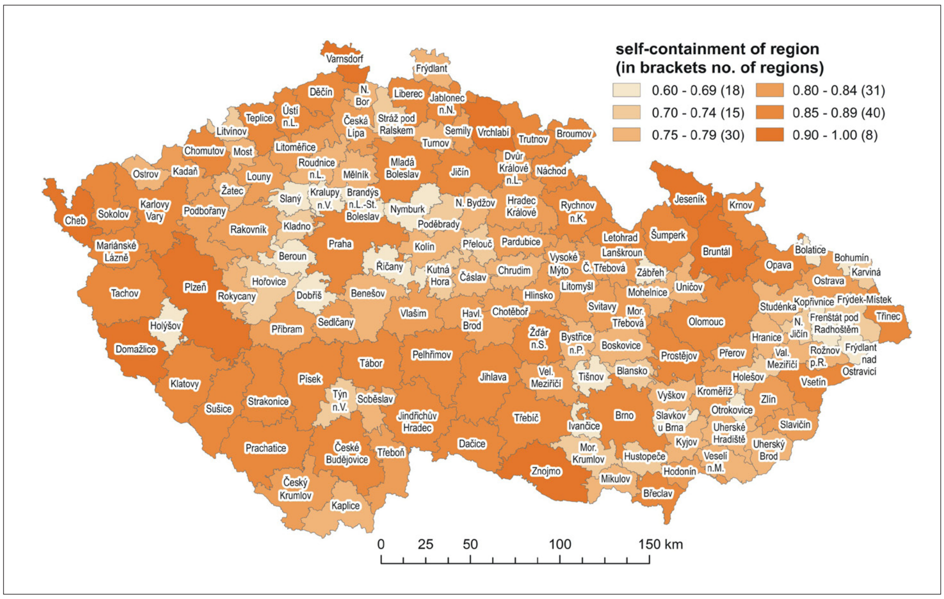


Fig. 1: Regional system FRD 1. Source: authors' elaboration

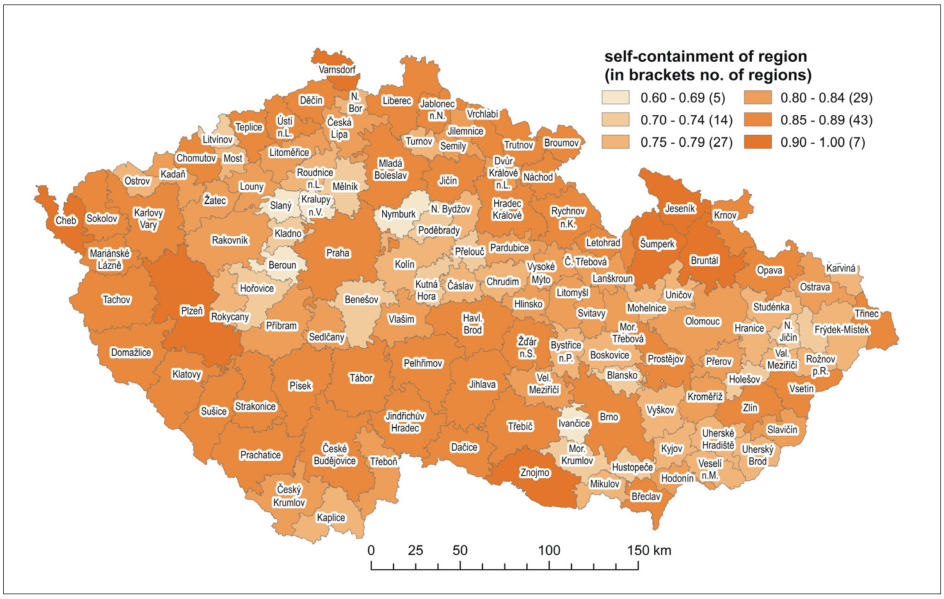


Fig. 2: Regional system FRD 2. Source: authors' elaboration

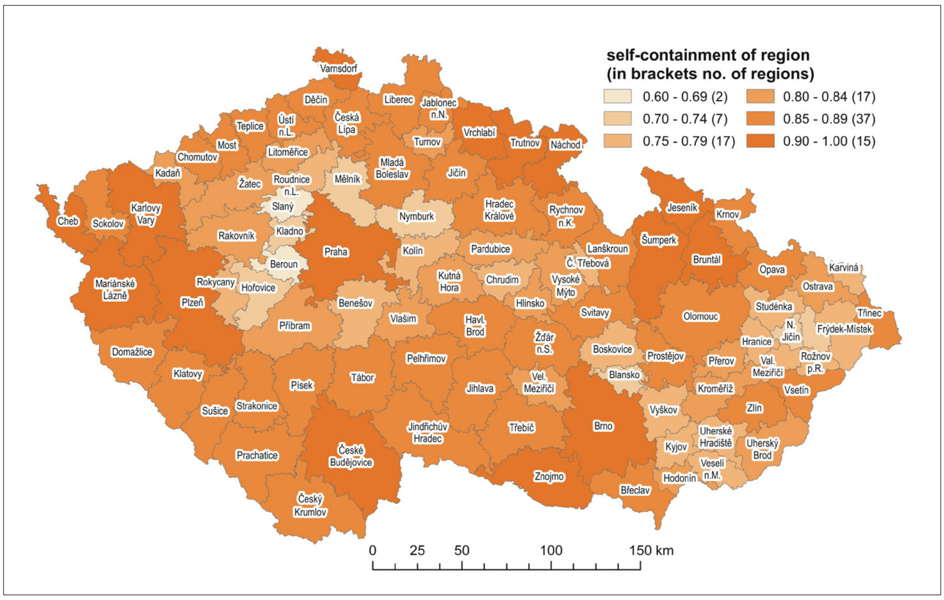


Fig. 3: Regional system FRD 3. Source: authors' elaboration

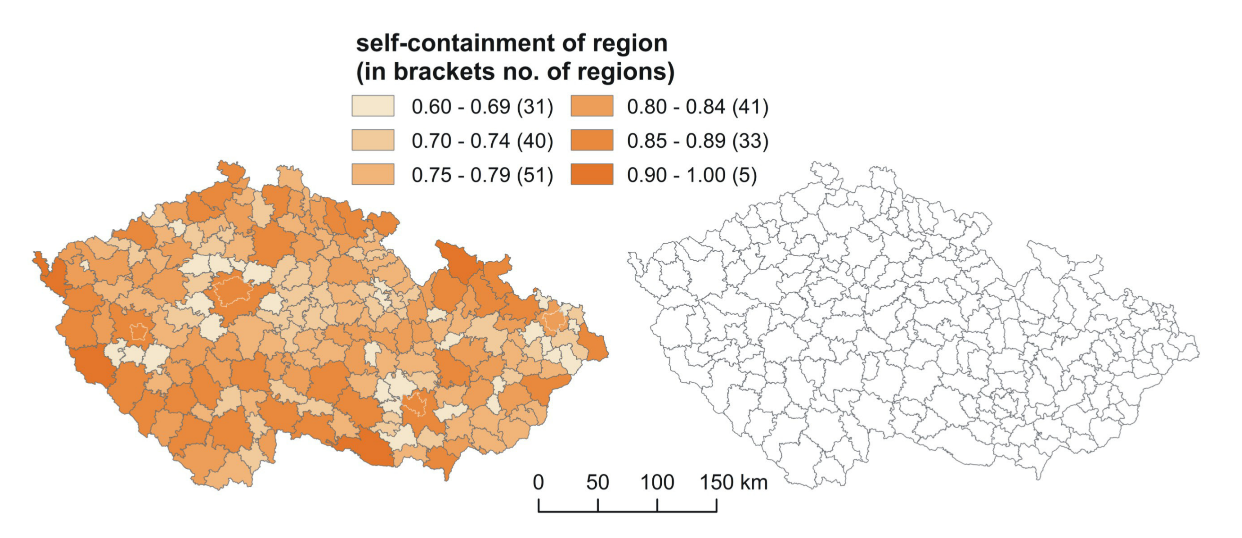


Fig. 4: Approximated regional system AFRD 1 (left) and AMEPs (right)
Source: authors' elaboration

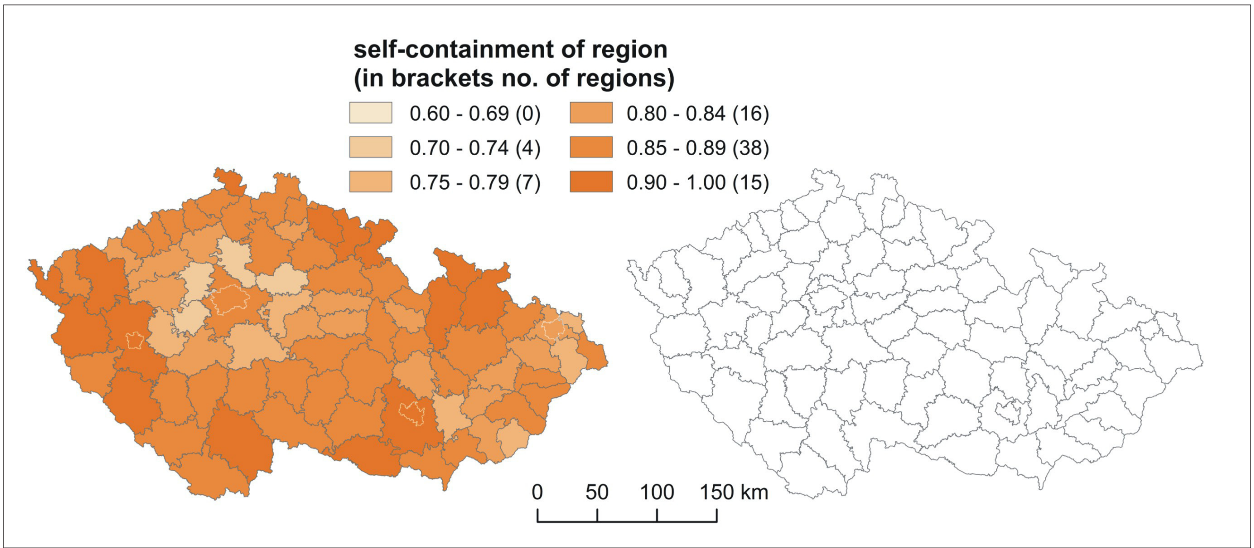


Fig. 5: Approximated regional system AFRD 2 (left) and districts (right)
Source: authors' elaboration

of micro regions and even those of meso regions. The average years in education cluster positively in the hinterland of large university cities. Finally the age preference index, as the least complex characteristic, clusters the least in spatial terms. The most relevant regions for this characteristic should have their centres approximately 40 km apart. Out of the four selected characteristics, the unemployment patterns in LISA cluster maps best approximate the borders of functional regions. Relatively cohesive clusters within functional regions result from the fact that unemployment is a characteristic directly related to the interaction data used for the construction of functional regions – i.e. daily travel-to-work flows.

Tables 3, 4, and 5 show different statistics measuring homogeneity, the importance of respective regional levels for overall variability, the regional variability for different regional systems (two administrative systems, two approximated regional systems, and three optimised regional systems), and the four selected socio-geographical characteristics. In accordance with the main objective of the paper, special attention is paid to the differences (and similarities) between administrative systems and functional regional systems. Intra-regional variability (homogeneity) is measured by the coefficient of variation, separately for each regional unit. This statistic was calculated in both unweighted and weighted form in order to eliminate the effect of different population sizes of units. The minimum, maximum, and mean values of the coefficient of variation presented in Table 3 show the level of differences between municipalities in each regional unit for a particular regional system.

Employment in agriculture has the highest values for intra-regional variability of all regional systems. In contrast, the average number of years in education has the lowest values. These results fully correspond with the spatial patterns of the characteristics studied and presented in Figure 6, particularly in regards to the homogeneity of spatial clusters of high or low values (i.e. the presence of spatial outliers), and the spatial relationship between clusters and regional boundaries. For example, agriculture is primarily not affected by the socio-economic regions, but by differences between rural and urban areas and by physical geographical conditions. In general, the values of inter-regional variability indicate no differences between administrative and functional regional systems. The only

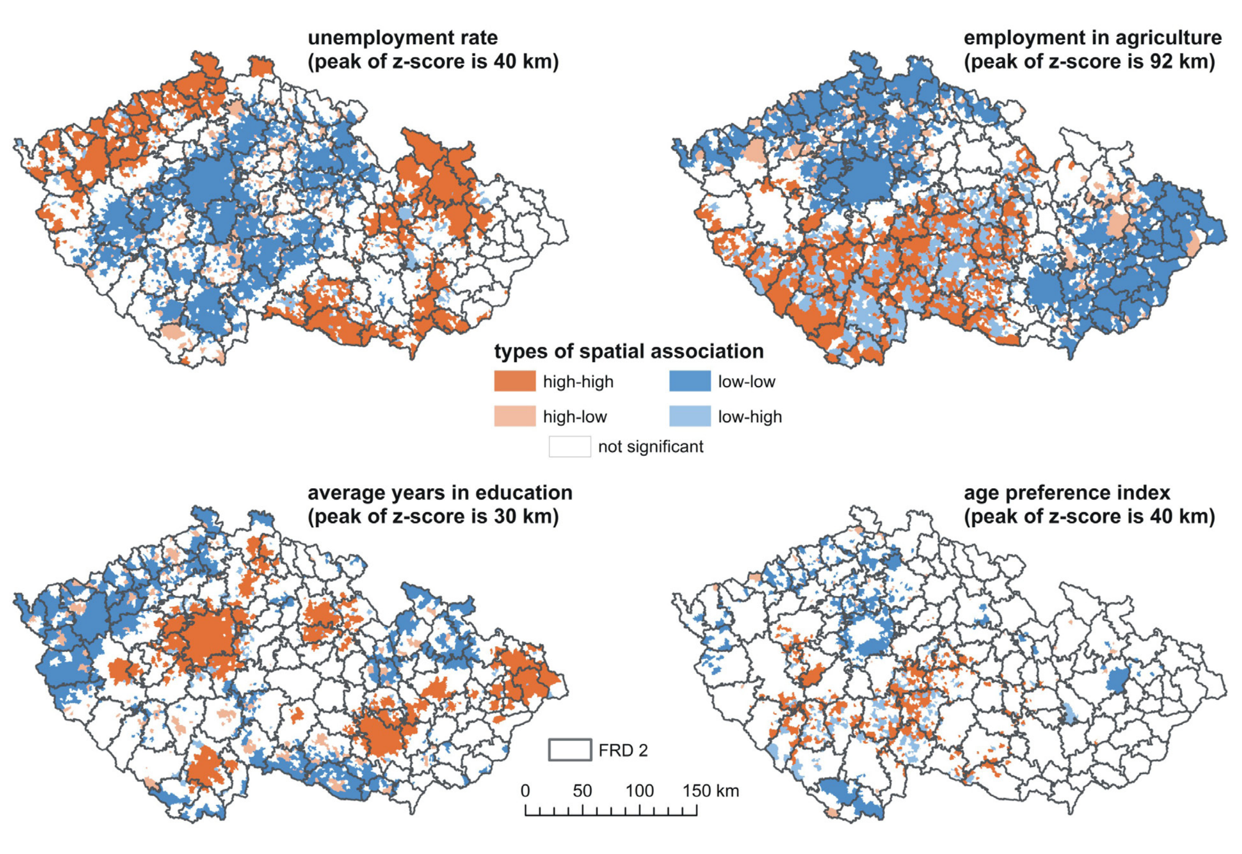


Fig. 6: LISA cluster maps for studied characteristics. Source: authors' elaboration
Notes: LISA cluster maps are constructed using a distance-based spatial weight matrix with the cut-off equal to the maximum z-score for respective characteristics. The significance level is 5%.

logical dependence is on the number of units in each regional system: the more units there are, the lower the measure of intra-regional variability.

Table 4 presents values for inter-regional variability using the coefficient of variation. The same results were reached using the Theil index as another measure of interregional variability. The values show that not only the intraregional, but also the interregional variability reaches maximum values for the employment in agriculture and minimum values for the average number of years in education. The comparison of different regional systems again shows neither significant differences between administrative and functional systems, nor the influence of the number of units.

Table 5 presents the share the interregional component of the Theil index has of the total variability, when its intraregional component can be easily derived as an algebraic complement to 100%. Unlike previous results, these calculations bring new and unexpected information about the structure of interregional variability. The unemployment rate and the average years in education have the highest interregional component of the overall variability. In the case of the unemployment rate, it documents the effect of local labour markets (i.e. the functional regions used in this paper) on the spatial pattern of this characteristic. However, even this characteristic with its close relationship with functional regions does not show any major differences when compared with administrative regions.

All measures of intra-regional, relative regional and inter-regional variability for selected socio-geographical characteristics show very similar results for all seven sets of regional systems; only the number of regions, i.e. the scale effect of MAUP, plays some role in this respect. All of the measures of variability are primarily affected by the number of regions; the zoning effect of MAUP has a marginal role with minimum effects as documented by the comparison of administrative and functional regional systems with similar numbers of units. One reason for this is that all characteristics studied are influenced by and operate on a micro-regional level, as demonstrated by the spatial autocorrelation analysis. It can be expected that more distinct regional variability should occur at higher or lower hierarchical levels. In this work, however, only the structure as a whole was analysed, without regard to local differences. It could be interesting to compare the regional delimitation and regional variability of some particular administrative and functional regions for a broader set of characteristics.

5. Conclusion

The third variant of the CURDS regionalization algorithm, using the original constraint function proposed by Halás et al. (2015), has proved to be a suitable method for the definition of functional regions, and has produced relevant results. This variant uses the latest knowledge of operations using the constraint function and the regions are delineated without the unnecessary effects of further constraints, such as normative identification of regional cores and normative determination of size and self-containment of the resulting regions. The paper analysed seven regional systems in the Czech Republic at the micro-regional level. Two of them were represented by existing administrative divisions: districts and areas of municipalities with extended powers

Area type			Coefficient of variation unweighted				Coefficient of variation weighted			
			UNEMP	AGRI	EDU	AGE	UNEMP	AGRI	EDU	AGE
Czech Republic			0.47	0.86	0.04	0.74	0.33	1.36	0.06	0.27
Administrative regions	AMEP (206 units)	Minimum	0.09	0.32	0.00	0.04	0.05	0.30	0.00	0.02
		Maximum	0.72	1.67	0.12	1.57	0.42	1.87	0.07	0.79
		mean	0.34	0.69	0.03	0.42	0.20	0.85	0.03	0.24
	districts (77 units)	Minimum	0.20	0.46	0.02	0.10	0.07	0.64	0.01	0.04
		Maximum	0.62	1.57	0.07	1.34	0.39	1.83	0.05	0.57
		mean	0.37	0.77	0.04	0.52	0.22	0.99	0.03	0.26
Functional regions	AFRD 1 (205 units)	Minimum	0.09	0.35	0.01	0.03	0.06	0.33	0.01	0.02
		Maximum	0.64	1.48	0.11	1.67	0.44	2.01	0.07	0.83
		mean	0.35	0.69	0.03	0.44	0.21	0.88	0.03	0.25
	AFRD 2 (84 units)	Minimum	0.19	0.44	0.02	0.10	0.10	0.65	0.02	0.13
		Maximum	0.62	1.37	0.07	1.50	0.39	1.85	0.05	0.62
		mean	0.37	0.75	0.04	0.50	0.23	0.99	0.03	0.26
	FRD 1 (146 units)	Minimum	0.09	0.34	0.01	0.03	0.07	0.43	0.01	0.02
		Maximum	0.62	1.51	0.07	1.62	0.44	1.85	0.05	0.81
		mean	0.35	0.71	0.03	0.47	0.21	0.92	0.03	0.25
	FRD 2 (129 units)	Minimum	0.14	0.34	0.02	0.10	0.07	0.44	0.01	0.10
		Maximum	0.62	1.44	0.07	1.61	0.44	1.99	0.05	0.70
		mean	0.35	0.72	0.04	0.47	0.22	0.93	0.03	0.25
	FRD 3 (99 units)	Minimum	0.14	0.44	0.02	0.10	0.10	0.51	0.01	0.11
		Maximum	0.61	1.44	0.07	1.50	0.38	1.85	0.05	0.62
		mean	0.36	0.74	0.04	0.49	0.22	0.96	0.03	0.25
difference between AMEP and AFRD 1 means			0.01	0.00	0.00	0.03	0.01	0.02	0.00	0.01
difference between districts and AFRD 2 means			0.00	– 0.02	0.00	– 0.02	0.00	0.00	0.00	– 0.01

Tab. 3: Intra-regional variability for administrative and functional regional systems. Source: authors' computations
Notes: (1) in this and all following tables: UNEMP = unemployment rate; AGRI = employment in agriculture; EDU = average years in education; AGE = age preference index; (2) the weights used are (a) economically active population for UNEMP, (b) population for AGRI, EDU, and AGE.

Area type	Coefficient of variation unweighted				Coefficient of variation weighted			
	UNEMP	AGRI	EDU	AGE	UNEMP	AGRI	EDU	AGE
AMEP (206 units)	0.26	0.59	0.03	0.14	0.28	0.80	0.05	0.14
districts (77 units)	0.25	0.55	0.03	0.12	0.26	0.72	0.05	0.13
AFRD 1 (205 units)	0.25	0.59	0.03	0.13	0.28	0.81	0.05	0.14
AFRD 2 (84 units)	0.24	0.56	0.03	0.11	0.26	0.72	0.05	0.13
FRD 1 (146 units)	0.24	0.55	0.03	0.12	0.27	0.76	0.05	0.14
FRD 2 (129 units)	0.24	0.54	0.03	0.12	0.26	0.76	0.05	0.14
FRD 3 (99 units)	0.24	0.55	0.03	0.11	0.26	0.73	0.05	0.13
difference between AMEP and AFRD 1 means	0.01	0.00	0.00	0.00	0.00	– 0.01	0.00	0.00
difference between district and AFRD 2 means	0.01	– 0.01	0.00	0.01	0.00	0.00	0.00	0.00

Tab. 4: Inter-regional variability of the coefficient of variation for administrative and functional regional systems
Source: authors' computations

Area type	Inter-regional component (%)			
	UNEMP	AGRI	EDU	AGE
	Theil	Theil	Theil	Theil
AMEP (206 units)	70.42	50.73	75.83	32.07
districts (77 units)	62.12	43.38	68.18	23.58
AFRD 1 (205 units)	69.74	51.15	75.70	32.33
AFRD 2 (84 units)	61.74	44.17	68.43	25.09
FRD 1 (146 units)	65.97	47.98	72.62	28.42
FRD 2 (129 units)	65.15	47.19	71.71	28.00
FRD 3 (99 units)	62.95	45.29	69.06	25.91
difference between AMEP and AFRD 1 means	0.68	− 0.43	0.14	− 0.26
difference between district and AFRD 2 means	0.38	− 0.78	− 0.26	− 1.51

Tab. 5: Inter-regional component for administrative and functional regional systems
Source: authors' computations

(AMEPs). Five spatial schemes were based on the concept of a functional region, which particularly favours the self-containment of regions. Three of these spatial patterns were considered to consist of optimised functional regions, while two consisted of approximated functional regions, where the approximation took into account the number of administrative units, i.e. districts and AMEPs. The regional variability of four selected socio-geographical characteristics for the seven regional systems was analysed in order to fulfil the main objective of the paper, which was the evaluation of the efficiency and suitability of agglomerated areal units for the purpose of spatial and regional analysis.

The results of the spatial analyses indicated that there are no significant differences between administrative and functional regional systems with respect to the measurement of regional variability in the Czech Republic, at least for the chosen characteristics. Regarding the modifiable areal unit problem (MAUP), the agglomeration of basic spatial units (municipalities) into administrative or functional regions does not manifest any significant deviations within the set of seven regional systems. It has been shown that the number of regions is significant (the issue of scale) and that the statistical information presented in the tables changes gradually with a decreasing number of regions, without any shift in the direction of this change (with a decreasing number of regions the inter-regional variability and internal homogeneity increases). When the issue of aggregation (zoning) is taken into account, for the two pairs of regional systems with approximately the same number of regions, the results of all three kinds of analyses also did not show any significant differences within each pair.

The three variants of optimised functional regional systems, however, have the advantage of capturing the natural distribution of daily movements of a considerable part of the population, and thus for purely scientific and local view purposes they should be preferred to administrative regions. Moreover, these sets of functional regions are not manually adjusted, for instance with regard to the contiguity of regions. Thus, these regional systems offer further possibilities for spatial analyses between the level of AMEPs and the level of districts, such as a local view of the differences in the delineation of individual regions.

Finally, it can be generally concluded that the two analysed administrative systems of the Czech Republic (districts and AMEPS) do not differ significantly from regional systems which consist of functional regions with similar numbers of units, according to the measurement of regional variability. Therefore, administrative regional systems can be regarded as efficient enough and suitable for geographic, regional and spatial analysis. On the other hand, however, there are local differences between administrative and functional regional systems, particularly in the hinterlands of large cities. The outcomes of this project offer general conclusions not only for the Czech Republic, but also for other countries and regions. This generalisation is that functional regions are very suitable areal units for spatial analysis, regarding the labour market in particular. Given that the results of this analysis, however, do not differ to a great extent from the analysis carried out for administrative regions (at the same hierarchical level), there is no crucial reason to modify the administrative division in a more significant way. In this case it is more suitable to follow the principle of spatial stability, i.e. to support the stability of the current administrative divisions, including the operation of its institutions, over time.

Acknowledgement

This work was supported by the Czech Science Foundation (GAČR) under Grant No. 16-13502S: "Spatial uncertainty and fuzzy regional systems: identification, analysis and implications of pulsating functional regions" (sections of this article dealing with functional regions and functional regionalisation – participants Klapka and Halás), and by the Czech Science Foundation (GAČR) under Grant No. P404-12-P340: "Spatial aspects of social differentiation in the Czech Republic and neighbouring countries with special attention to boundary effects" (sections dealing with spatial differentiation and spatial analysis – participants Netrdová and Nosek).

References:

ANSELIN, L. (1988): Spatial Econometrics: Methods and Models. Dordrecht, Kluwer Academic Publishers.

ANSELIN, L. (1995): Local Indicators of Spatial Association – LISA. Geographical Analysis, 27(2): 93–115.

APOLLONIO, N., BECKER, R. I., LARI, I., RICCA, F., SIMEONE, B. (2009): Bicolored graph partitioning, or: gerrymandering at its worst. Discrete Applied Mathematics, 157(17): 3601–3614.

BEZÁK, A. (1997): Priestorová organizácia spoločnosti a územno-správne členenie štátu. Acta Universitatis Matthiae Belii, Geografické štúdie, 3: 6–13.

BUČEK, J. (2002): Regionalization in the Slovak Republic – from administrative to political regions. In: Regionalization for development and accession to the European Union: A comparative perspective. Budapest (pp. 141–178). Local Government and Public Service Reform Initiative.

BUČEK, J. (2005): The role of local government in local economic development – Slovakia. In: Local government and economic development. Budapest, Local Government and Public Service Reform Initiative (pp. 461–542). Open Society Institute.

BUNGE, W. (1966): Gerrymandering, geography, and grouping. Geographical Review, 56(2): 256–263.

CASADO-DÍAZ, J. M. (2000): Local labour market areas in Spain: a case study. Regional Studies, 34(9): 843–856.

CASADO-DÍAZ, J. M., COOMBES, M. (2011): The delineation of 21st century local labour markets areas: a critical review and a research agenda. Boletín de la Asociación de Geógrafos Espańoles, 57: 7–32.

CLIFF, A. D., ORD, J. K. (1973): Spatial Autocorrelation. London, Pion.

COOMBES, M. (2010): Defining labour market areas by analysing commuting data: innovative methods in the 2007 review of Travel-to-Work Areas. In: Stillwell, J., et al. [eds.]: Technologies for migration and commuting analysis: spatial interaction data applications (pp. 227– 241). Hershey, IGI Global.

COOMBES, M. G. (2000): Defining locality boundaries with synthetic data. Environment and Planning A, 32(8): 1499–1518.

COOMBES, M. G., BOND, S. (2008): Travel-to-Work Areas: the 2007 review. London, Office for National Statistics.

COOMBES, M. G., DIXON, J. S., GODDARD, J. B., OPENSHAW, S., TAYLOR, P. J. (1982): Functional regions for the population census of Great Britain. In: Herbert, D. T., Johnston, R. J. [eds.]: Geography and the Urban Environment. Progress in Research and Applications 5 (63–112). Chichester, John Wiley and Sons Ltd.

COOMBES, M. G., GREEN, A. E., OPENSHAW, S. (1986): An Efficient Algorithm to Generate Official Statistical Reporting Areas: The Case of the 1984 Travel-to-Work Areas Revision in Britain. The Journal of the Operational Research Society, 37(10): 943–953.

COWELL, F. A., JENKINS, S. P. (1995): How Much Inequality Can We Explain? A Methodology and an Application to the United States. Economic Journal, 105(429): 421–430.

DZIEWOŃSKI, K. (1967): Concepts and terms in the field of economic regionalization. In: Macka, M. [ed.]: Economic regionalization (pp. 25–36). Praha, Academia.

FARMER, C. J. Q., FOTHERINGHAM, A. S. (2011): Network-based functional regions. Environment and Planning A, 43(11): 2723–2741.

FLÓREZ-REVUELTA, F., CASADO-DÍAZ, J. M., 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.

FOTHERINGHAM, A. S., WONG, D. W. S. (1991): The modifiable areal unit problem in multivariate statistical analysis. Environment and Planning A 23(7): 1025–1044.

GEHLKE, C. E., BIEHL, K. (1934): Certain effects of grouping upon the size of the correlation coefficient in census tract material. Journal of the American Statistical Association, 29(185): 169–170.

GETIS, A. (2008): A History of the Concept of Spatial Autocorrelation: A Geographer's Perspective. Geographical Analysis, 40(3): 297–309.

GOODALL, B. (1987): The Penguin dictionary of human geography. Harmondsworth, Penguin.

GOODCHILD, M. F. (1986): Spatial Autocorrelation. CATMOG 47. Norwich, GeoBooks.

GOODMAN, J. F. B. (1970): The definition and analysis of local labour markets: some empirical problems. British Journal of Industrial Relations, 8(2): 179–196.

GRASLAND, C. et al. (2006): ESPON 3.4.3 The Modifiable Areas Unit Problem: Final Report [online]. [cit.14.04.2016]. Available at: https://www.espon.eu/export/sites/default/Documents/Projects/ESPON2006Projects/StudiesScientificSupportProjects/MAUP/espon343_maup_final_version2_nov_2006.pdf

HAGGETT, P. (1965): Locational Analysis in Human Geography. London, Arnold.

HALÁS, M., KLAPKA, P. (2012): Contribution to regional division of Slovakia based on the application of the Reilly's model. Hungarian geographical bulletin, 61(3): 237–255.

HALÁS, M., KLAPKA, P., BEDNÁŘ, M., TONEV, P. (2015): An alternative definition and use for the constraint function for rule-based methods of functional regionalisation. Environment and Planning A, 47(5): 1175–1191.

HALÁS, M., KLAPKA, P., BLEHA, B., BEDNÁŘ, M. (2014): Funkčné regióny na Slovensku podľa denných tokov do zamestnania. Geografický časopis, 66(2): 89–114.

HANSON, S., PRATT, G. (1992): Dynamic dependencies: a geographic investigation of local labor markets. Economic Geography, 68(4): 373–405.

HELDT CASSEL, S., MACUCHOVA, Z., RUDHOLM, N., RYDELL, A. (2013): Willingness to commute long distance among job seekers in Dalarna, Sweden. Journal of transport geography, 28: 49–55.

JOHNSTON, R. (2002): Manipulating maps and winning elections: measuring the impact of malapportionment and gerrymandering. Political Geography 21(1): 1–31.

KARLSSON, C., OLSSON, M. (2006): The identification of functional regions: theory, methods, and applications. The Annals of Regional Science, 40(1): 1–18.

KLAPKA, P., HALÁS, M., ERLEBACH, M., TONEV, P., BEDNÁŘ, M. (2014): A multistage agglomerative approach for defining functional regions of the Czech Republic: the use of 2001 commuting data. Moravian Geographical Reports, 22(4): 2–13.

KLAPKA, P., HALÁS, M., TONEV, P. (2013a): Functional regions: concept and types. In 16th International Colloquium on Regional Sciences, Conference Proceedings (pp. 94–101). Brno, Masaryk University.

KLAPKA, P., HALÁS, M., TONEV, P., BEDNÁŘ, M. (2013b): Functional regions of the Czech Republic: comparison of simple and advanced methods of regional taxonomy. Acta Universitatis Palackianae Olomucensis, Facultas Rerum Naturalium, Geographica, 44(1): 45–57.

MICHNIAK, D. (2003): Dostupnosť okresných miest na Slovensku. Geografický časopis, 55(1): 21–39.

MOORE, T. (2002): Comments on Ron Johnston's "Manipulating maps and winning elections: measuring the impact of malapportionment and gerrymandering". Political Geography, 21(1): 33–38.

NETRDOVÁ, P., NOSEK, V. (2009): Přístupy k měření významu geografického rozměru společenských nerovností. Geografie, 114(1): 52–65.

NOSEK, V., NETRDOVÁ, P. (2014): Measuring Spatial Aspects of Variability. Comparing Spatial Autocorrelation with Regional Decomposition in International Unemployment Research. Historical Social Research, 39(2): 292–314.

NOVOTNÝ, J., NOSEK, V. (2012): Comparison of regional inequality in unemployment among four Central European countries: an inferential approach. Letters in Spatial and Resource Sciences, 5(2): 95–101.

OPENSHAW, S. (1977): A Geographical Solution to Scale and Aggregation Problems in Region-Building, Partitioning and Spatial Modelling. Transactions of the Institute of British Geographers, New Series, 2(4): 459–472.

OPENSHAW, S. (1984): The Modifiable Areal Unit Problem. CATMOG 38. Norwich, GeoBooks.

OZKUL, B. D. (2014): Changing home-to-work travel in England and Wales. Regional Studies, Regional Science, 1(1): 32–39.

SHORROCKS, A. F., WAN, G. (2005): Spatial Decomposition of Inequality. Journal of Economic Geography, 5(2): 59–81.

SMART, M. W. (1974): Labour market areas: uses and definition. Progress in Planning, 2(4): 239–353.

SPURNÁ, P. (2008): Prostorová autokorelace – všudypřítomný jev při analýze prostorových dat? Sociologický časopis/Czech Sociological Review, 44(4): 767–787.

SUCIU, M. (2002): Local government and decentralisation in Eastern Europe. Case study on Romania. Revista Romana de Geografie Politica, 4(1–2): 33–55.

TOBLER, W. (1970): A computer movie simulating urban growth in the Detroit region. Economic Geography, 46, Supplement: Proceedings. International Geographical Union, Commission on quantitative methods: 234–240.

TOBLER, W. (2004): On the first law of geography: a reply. Annals of the Association of American Geographers, 94(2): 304–310.

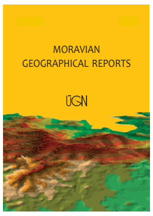
UNWIN, D. J. (1996): GIS, spatial analysis and spatial statistics. Progress in Human Geography, 20(4): 540–551.

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(2): 201–221.

YULE, G. U., KENDALL, M. G. (1950): An introduction to the theory of statistics. London, Griffin.

Please cite this article as:

KLAPKA, P., HALÁS, M., NETRDOVÁ, P., NOSEK, V. (2016): The efficiency of areal units in spatial analysis: Assessing the performance of functional and administrative regions. Moravian Geographical Reports, 24(2): 47–59. Doi: 10.1515/mgr-2016-0010.



MORAVIAN GEOGRAPHICAL REPORTS

Institute of Geonics, The Czech Academy of Sciences

journal homepage: **http://www.geonika.cz/mgr.html**

doi: **10.1515/mgr-2016-0011**

DE GRUYTER OPEN

Functional regions in gravity models and accessibility measures

Michael OLSSON [a] *

Abstract

Accessibility measures are useful for studies in Economic Geography. For example, accessibility to potential customers can be used in a study of firm behaviour. In such a study, it would be relevant to consider where potential customers live. This can be accomplished by splitting the accessibility measure into three parts: accessibility within the municipality, in other municipalities within the functional region, and in other regions. Many studies have proved this to be a very useful way to incorporate the spatial structure of the economy into economic studies. This paper deals with the issue of finding the distance-friction parameters needed to calculate such accessibility measures. There is a particular distance-friction parameter for interaction within the municipality, between municipalities within the functional region, and between regions. One way to find the distance-friction parameters is to solve a constrained gravity model, in which the functional regions are used as constraints. Both the models and the optimisation procedures in matrix form, and the Matlab programs used in the research are presented. The spatial constraints are gradually introduced into the models, which empowers the researcher to make such adjustments on their own. The data set used is available for downloading, and the reader can then try the models before creating a data set of their own.

Key words: *spatial interaction; commuting; gravity model; entropy; constrained optimisation; Matlab; Sweden*

Article history: *Received 28 June 2015; Accepted 29 April 2016; Published 30 June 2016*

1. Introduction

On the global level, the use of specialisation and scale economies increase overall production. Individuals as well as regions specialise in producing only a part of what they consume. With more goods and services available, society has the potential to create a better life for the population. Transportation of both production factors and products are essential factors in this complex system.

It is important to consider and take into account that economic activity has a location, since spatial interaction in most cases declines with distance. Geurs and van Wee (2004) present and review accessibility measures: Hansen (1959) was one of the first to use the accessibility concept. Johansson, Klaesson and Olsson (2002, 2003) suggest that it is useful to split the accessibility measure into parts, and the idea of accessibility measures on three different spatial levels has been widely adopted. For example, it matters to a firm, with a store in a municipality, if a potential customer lives within the municipality, in another municipality within the functional region, or in another region. The firm can calculate accessibility to potential customers within the municipality, in other municipalities within the functional region, and in other regions. It can be valuable to split the accessibility in this way, since they are likely to be of unequal importance to the firm.

Many studies, mostly Swedish, have used the results from our earlier studies (Johansson, Klaesson and Olsson, 2002, 2003). It has been used to study many different activities: for example, Andersson and Ejermo (2005) study knowledge sources and the innovativeness of corporations; Gråsjö (2006) studies spatial spillovers of knowledge production; Karlsson and Olsson (2006) study how to define functional regions; Johansson and Karlsson (2007) study R&D and export diversity; Andersson and Gråsjö (2009) study representations of space in empirical models; Olsson (2012) studies the work at the public employment offices; Backman (2013) studies human capital and firm productivity; Larsson and Öner (2014) study retail location; and Larsson (2014) studies the density-wage relationship. Gråsjö and Karlsson (2015) is a nice review that contains additional papers. Gråsjö and Karlsson (2013: 1) write "*However, it is a general method and there is no reason why the method does not apply for other countries*".

In order to calculate accessibility at three different spatial levels, the corresponding distance-friction parameters are needed. The main purpose of this paper is to enable you to

[a] Regional Economic Activity and Development, Enterprises for the Future, School of Business, University of Skövde, Sweden (*corresponding author: M. Olsson, e-mail: *michael.olsson@his.se*)

calculate the distance-friction parameters for the country you are interested in. The procedures are illustrated, and you will learn how to solve such models in detail using Matlab. In this paper, three models are stated in matrix form. This makes it easier to connect the text to the computer program. The ambition is to make it easy to look at the mathematical formulation and find almost the same in the program. In order to reduce the threshold, a data set is available for downloading. With the data and programs available, you can run the programs and check all the results. In this paper, the first model is gradually improved by incorporation of additional spatial constraints. There are several advantages with this approach. It makes the presentation cleaner and easier to grasp. Moreover, it enables you to make your own changes in the programs. In the future, you may want to estimate another version of the third model, or you may have a data set structured differently. After reading this paper you can handle such issues with ease. At least, that is the intention. The models are gradually made more complex, by adding constraints, to better reflect reality. It is also a purpose of the paper to present a comparison of the predictive power of the models. The third model has relatively many constraints, and performs better.

2. Commuting

Most workers have a relatively short commute, and it is rare to find a worker with a really long commute. This tendency is illustrated in Fig. 1. In this paper, municipalities are used as the spatial unit of analysis. The municipalities are more or less related to each other, however, and this relatedness across municipalities is captured using functional regions. It is possible to form functional regions using several approaches. The basic idea is that a functional region is built from municipalities with a relatively high level of interaction. In this paper the local labour market definition of a functional region is used. A local labour market consists of the municipalities that are tightly connected by commuting. A local labour market has a self-sufficient centre and surrounding municipalities. The surrounding municipalities are added to the core municipality, or to a municipality connected to the core, using one-way commuting. You find details of the procedure and maps of the Swedish local labour markets from Statistics Sweden (2015). An alternative to local labour markets is to create commuting zones using two-way commuting. Obviously, it is also possible to make other considerations. Karlsson and Olsson (2006) present local labour markets and some other methods and alternatives. The exact version of the functional region is not that important. The results will be similar if another version is picked. The basic reason is that most municipalities would be aggregated to the same functional region, independent of approach.

The commuting pattern gradually changes with time, and the area under the curve in Fig. 1 gradually shifts to the right with increased mobility. Not much happens to the pattern during a short period of time, but the pattern may change significantly if you observe a longer period. In Sweden, the daily average mobility of persons has increased from half a kilometer in the year 1900 to 45 kilometers in the year 1999 (Andersson and Strömquist, 1988; SIKA, 2000). The Swedish Institute for Transport and Communications Analysis (SIKA) has been replaced by the government agency Transport Analysis, and they estimate that the 2011 mobility is 44 kilometers (Transport Analysis, 2013). This change is also readily seen in the

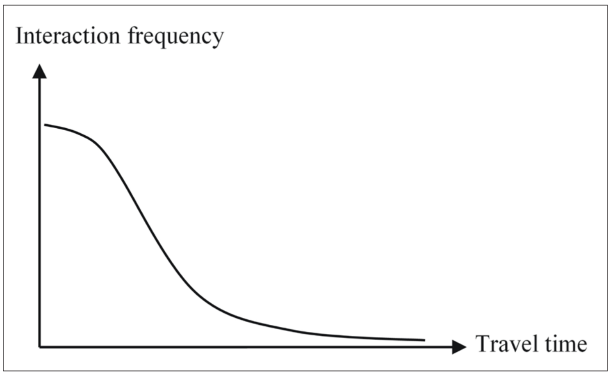


Fig. 1: Interaction declines with distance
Source: author´s elaboration

number of functional regions. With a long-term perspective, the number of functional regions has declined. This means that the models capture the spatial structure at a point in time. The overall tendencies will be the same for example ten years later, but the models should be rerun once in a while with the then present spatial structure. Moreover, it is possible to form functional regions per category of workers. In some studies, one may want to investigate educational, occupational, and/or gender differences.

The country consists of n municipalities, and a worker commutes from the home municipality, $i = 1, 2, \ldots, n$, to the work-place municipality, $j = 1, 2, \ldots, n$. The observed commuting information is collected in the $(n \times n)$ commuting matrix, $\mathbf{c} = \{c_{ij}\}$. A solution to a model would give the estimated commuting matrix, $\tilde{\mathbf{c}}$. There is also a corresponding $(n \times n)$ commuting-time matrix, $\mathbf{t} = \{t_{ij}\}$. Let us define a $(1 \times n)$ unit row vector, $\mathbf{u}$. The $(n \times 1)$ vector with the number of workers that lives in the municipalities equals the row sum of the commuting matrix, $\mathbf{o} = \mathbf{cu}'$, and the $(1 \times n)$ vector with the number of jobs in the municipalities equals the column sum of the commuting matrix, $\mathbf{d} = \mathbf{uc}$. The existing spatial structure is captured in the form of matrices. Three dummy variables are used to classify that a commute may end within the home municipality, in another municipality within the same functional region, or in another region. When a commute ends within the home municipality $k_{ij} = 1$, otherwise $k_{ij} = 0$. If a commute ends in another municipality within the home region $l_{ij} = 1$, otherwise $l_{ij} = 0$. If the commute ends in another region $m_{ij} = 1$, otherwise $m_{ij} = 0$. This information is collected in the $(n \times n)$ regional dummy matrices $\mathbf{k}$, $\mathbf{l}$, and $\mathbf{m}$, respectively. In this study, only links with a commuting time less than 150 minutes are included. This means that commuting on the other links, $(\mathbf{c}(\mathbf{t} > 150) = 0)$, are ignored. To identify all links that are included in this study zones are created, and collected into the $(n \times n)$ zone matrix, $\mathbf{z}$. In this matrix $z_{ij} = 1$ if $t_{ij} \leq 150$, otherwise $z_{ij} = 0$.

3. Data

In Table 1 you find the first five and last five rows in the Excel file used as input. The data set is in an Excel file that you have to download to run the Matlab programs. It is available from the following address: www.his.se/commuting. Nevertheless, it is useful to illustrate the structure of the data in this paper. At that time (1998), Sweden was separated in 289 municipalities. Hence, there are 83,521 commuting links. Each link has its own row in the Excel file. For each link, the data contains information whether the commute is within a municipality, between municipalities within a

Row	From	To	k_{ij}	l_{ij}	m_{ij}	t_{ij}	c_{ij}
1	1	1	1	0	0	4.71	6,904
2	1	2	0	1	0	11.52	141
3	1	3	0	1	0	24.31	28
4	1	4	0	1	0	30.98	11
5	1	5	0	1	0	16.10	479
…	…	…	…	…	…	…	…
83,517	289	285	0	0	1	209.64	225
83,518	289	286	0	0	1	236.85	3
83,519	289	287	0	0	1	216.61	19
83,520	289	288	0	0	1	219.08	2
83,521	289	289	1	0	0	67.13	10,549

Tab. 1: An excerpt from the data file, but the file only contains the white part
Source: Statistics Sweden and the Swedish Road Administration; author´s calculation

region, or between regions, the commuting time, and the number of commuters. The Excel file only contains the white part of Table 1.

The data has several sources. The commuting information originates from the Labor Statistics based on administrative sources (RAMS) from Statistics Sweden. Also, the information regarding the spatial structure originates from Statistics Sweden. The commuting times come from The Swedish Road Administration.

In Tab. 2, descriptive statistics for the number of commuters per link are presented. In Tab. 3, you find descriptive statistics for the commuting time per link. These numbers are calculated using only the active links with a commute shorter than 150 minutes. In Tab. 2, you also find the number of active links, the number of links with zero commuters, and the total number of links, given that the commuting time is shorter than 150 minutes. The number of commuters and the commuting times are clearly different for commutes within a municipality, between municipalities within a region, and between regions. In this paper, commuting flows are separated into commuting within a municipality, between municipalities within a region, and between regions. This separation is based on that these commuting flows differ. The null hypotheses, that the relative commuting frequencies, c_{ij} / o_i, for commuting within a municipality, between municipalities in a region, and commuting between regions are equal, have been tested and they are rejected.

4. Models, Matlab programs and results

Some spatial-interaction models are linear when written in logarithmic form. Fischer and Wang (2011) present the drawbacks related to the use of ordinary least squares to

Measure	Within municipality	Within region	Between regions	Sum
Min	761	1	1	–
Median	4,394	51	3	–
Mean	9,732	392	22	–
Max	266,980	19,647	6,050	–
Std. dev.	20,595	1,340	105	–
# active links*	289	2,087	9,911	12,287
# zero links*	0	81	8,413	8,494
# links*	289	2,168	18,324	20,781

Tab. 2: Descriptive statistics for commuting per link per commuting type
*Note: * Only commuting with $t < 150$ included*
Source: Statistics Sweden and the Swedish Road Administration; author´s calculation

Measure	Within municipality	Within region	Between regions
Min	3.4	6.3	12.3
Median	13.0	31.4	89.6
Mean	17.3	33.8	90.4
Max	89.8	96.3	150.0
Std. dev.	13.6	15.7	33.9

Tab. 3: Descriptive statistics for commuting time (minutes) per link per commuting type
Source: Statistics Sweden and the Swedish Road Administration; author´s calculation

estimate such a model. So, even though it may be tempting to estimate the logarithmic form of a spatial-interaction model using ordinary least squares, it should be avoided. In models of commuting it is preferred that the observed number of a) jobs in a municipality, and b) workers that live in a municipality (i.e. the data) both to be exactly equal to the estimates produced from the model. Olsson (2002) writes that constrained models have the advantage in that "(by construction) the model outcome is consistent with actual in- and out-commuting." In addition, we often want to include other constraints (e.g. time constraints). You can model the individual's choice to commute, or the aggregate commuting pattern. A gravity model of the aggregated commuting pattern relates interaction to an origin weight function, a destination weight function, and a distance deterrence function (Sen and Smith, 2011). The aggregate commuting function derived from maximising entropy is equivalent in form to the one derived from a logit model of individual (discrete) choice (Anas, 1983; Mattsson, 1984). So, studying the commuting pattern by maximising entropy a) produces a solution similar in form to the one that follows from individuals choosing their commute, and also b) enforces structure to the model via constraints.

In this paper, the aggregate commuting pattern is modelled by maximising entropy. In this section, you find three models of commuting. The first model has only two constraints. The point of this model is not that it will replicate the commuting pattern well, but that this model contains the essence of the following models. Each model is fully presented, i.e. the program used to estimate the model is described and the results are presented, before the next model is introduced. The first model is the base to which spatial structure (e.g. functional region, origin and destination constraints) is gradually incorporated. This is straight forward given an understanding of the first model and the Matlab program used to solve it. In the following models, the ideas presented in the first model are just extended. The second model has six constraints, and the third model has 582 unique constraints. The ambition is to incorporate spatial constraints into the model and to better replicate the pattern illustrated in Figure 1. If you want to get a preview of what is ahead, you can compare Fig. 1 to Fig. 10. It is the third model that is the best, since it enforces many more constraints. The first two models are just used to get to the third model, in the easiest possible manner.

By construction, Model 3 does a better job replicating the commuting pattern. In Model 3, the balancing factors (i.e. constraint multipliers) for where workers live (i.e. the origin constraints) and work (i.e. the destination constraints) captures the spatial surrounding of locations. All workers in Sweden are included in the data, but if all workers and all firms could redo their choices, many choices would change. The observed commuting data is but one realisation out of many possible. It is an aggregate observation in time of the (random) discrete choices made by individuals and firms. One consequence of this randomness is that there are spatial dependencies, e.g. if relatively many from a municipality commute on one link, it follows that relatively fewer commute on the other links. This would be seen as deviations from the estimated pattern.

4.1 Model 1

The observed population equals the sum of all commuters, $p = \sum_i \sum_j c_{ij} = \mathbf{ucu}'$. The Hadamard product sign, $\circ$, is used for entrywise multiplication of matrices. The observed total commuting time equals $r = \sum_i \sum_j c_{ij} t_{ij} = \mathbf{u(c \circ t)u}'$. In the first model, two constraints enforce that the estimated population, $\tilde{p} = \mathbf{u\tilde{c}u}'$, equals the observed population, and that the estimated total commuting time, $\tilde{r} = \mathbf{u(\tilde{c} \circ t)u}'$, equals the observed total commuting time. In constrained gravity models the objective is to maximise the system entropy, $\sum_i \sum_j c_{ij} \ln(c_{ij}) - c_{ij} = -\mathbf{u}(\tilde{\mathbf{c}} \circ \ln(\tilde{\mathbf{c}}) - \tilde{\mathbf{c}})\mathbf{u}'$, subject to the constraints. Therefore, the primal formulation of the problem is to max $L(\tilde{\mathbf{c}}, \delta, \gamma)$, where the Lagrangian function is $L(\tilde{\mathbf{c}}, \delta, \gamma) = -\mathbf{u}(\tilde{\mathbf{c}} \circ \ln(\tilde{\mathbf{c}}) - \tilde{\mathbf{c}})\mathbf{u}' + \delta(\mathbf{u\tilde{c}u}' - p) + \gamma(r - \mathbf{u}(\tilde{\mathbf{c}} \circ \mathbf{t})\mathbf{u}')$.

Let us call the Lagrangian multipliers, δ and γ, the proximity-preference parameter and the distance-friction parameter, respectively. In this model, the proximity-preference parameter is a fixed factor for all commutes, and does not really reveal any preference for proximity. But, the name will make more sense in the following models. The Lagrangian written in this form highlights the constraints. But, to get to the dual formulation of the problem it is easier to use $L(\tilde{\mathbf{c}}, \delta, \gamma) = \mathbf{u}(\tilde{\mathbf{c}} \circ \ln(\tilde{\mathbf{c}}) + \tilde{\mathbf{c}} + \delta\tilde{\mathbf{c}} - \gamma\tilde{\mathbf{c}} \circ \mathbf{t})\mathbf{u}' - \delta p + \gamma r$. We can rewrite $\partial L / \partial\tilde{\mathbf{c}} = -\ln(\tilde{\mathbf{c}}) + \delta - \gamma t = 0$ as $\tilde{\mathbf{c}} = \exp(\delta\mathbf{u}'\mathbf{u} - \gamma\mathbf{t})$. Hence, commuting on a particular link equals $c_{ij} = \exp(\delta - \gamma t_{ij})$. Inserting this in the primal form gives the dual form: $\min D(\delta, \gamma)$, where $D(\delta, \gamma) = \mathbf{u} \exp(\delta\mathbf{u}'\,\mathbf{u} - \gamma\mathbf{t})\mathbf{u}' - \delta p + \gamma r$.

The Newton-Raphson iterative procedure is used to find the optimum, and you find a description of the procedure in Appendix 1 (see link to Supplementary material at the end of the article). The iterative procedure needs some parameter start values. Reasonable start values must fulfill one of the constraints, and here the population constraint is used, $\mathbf{u\tilde{c}u}' - p = 0$. If $\gamma_0 = 0$ it follows that $\delta_0 = \ln(p / (\mathbf{uzu}'))$. In this study all links where $t_{ij} > 150$ are ignored. This reduces the number of links from $\mathbf{u(k + l + m)u}'$ which is 83,521 to $\mathbf{uzu}'$ which is 20,781. With p = 3,847,782 the start value is equal to $\delta_0 = 5.2212$. Now, it is time to iterate from the start values towards the solution. The start values imply that commuting is not affected by commuting time. Hence, estimated commuting on links with long commuting time is bigger than observed commuting. Therefore, the estimated commuting flows use more time than is allowed. This implies that the distance friction parameter has to be raised. Raising the distance friction reduces estimated commuting flows, which leads to that too few persons work. It gives that the proximity-preference parameter has to be raised. And, this is sequentially repeated until the solution is found. If a constraint is violated in the opposite direction, the parameter estimate is adjusted accordingly.

As said, it is most likely that the estimated commuting flows do not fulfill the constraint on commuting time, $r - \mathbf{u}(\tilde{\mathbf{c}} \circ \mathbf{t})\mathbf{u}' = 0$, at the start. The distance-friction parameter estimates are adjusted using the Newton-Raphson procedure. The partial derivatives are $\partial D / \partial\gamma = -\mathbf{u}(\tilde{\mathbf{c}} \circ \mathbf{t})\mathbf{u}' + r = r - \tilde{r}$ and $\partial^2 D / \partial\gamma^2 = \mathbf{u}(\tilde{\mathbf{c}} \circ \mathbf{t} \circ \mathbf{t})\mathbf{u}' = \tilde{s}$, which leads to the following adjustment scheme $\gamma_{(n+1)} = \gamma_n - \rho(r - \tilde{r}_n) / \tilde{s}_n$. It is important to recalculate the commuting flows, before adjusting the proximity-preference parameter. The derivatives are $\partial D / \partial\delta = \mathbf{u\tilde{c}u}' - p = \tilde{p} - p$ and $\partial^2 D / \partial\delta^2 = \mathbf{u\tilde{c}u}' = \tilde{p}$, which leads to the following adjustment scheme, $\delta_{(n+1)} = \delta_n - \rho(\tilde{p}_n - p) / \tilde{p}_n$. In the first model, $\rho = 1$. It is important to recalculate the commuting flows, before starting over again. The program iterates until all constraints are fulfilled with extreme accuracy, since the run time is short.

4.1.1 The Matlab program

Now it is time to look at the Matlab program for Model 1. To make the reading easier, the program is included in

Appendix 2 (see Supplementary material). The structure of the first program is maintained in the models to come. First, the data file is read. In this section of the program **u**, **t**, and **c** are declared, and filled with values from the data. Then, the a priori information is calculated from the data, and the start values are set. In this part of the program, r, p, **z**, and the start values are calculated as described in the text above. Here, the estimated commuting flows using the parameter start values are calculated. The parameter start values and the value of the dual function are saved. This is done to later illustrate convergence. In the main iterative part of the program, each parameter is adjusted in relation to the constraint deviation. First, the distance-friction parameter is adjusted. Second, the proximity-preference parameter is adjusted. After each parameter adjustment the estimated commuting flows are recalculated. The new parameter values and the value of the dual function are saved. The end part of the program creates graphs, and saves the results to an Excel file. The Model 1 program is adjusted in the following models to incorporate additional spatial information.

The mathematical notation in the program is for the most part as in the text, so it should be easy to follow. However, there are four minor exceptions. In the text the Hadamard product sign ∘ is used for entrywise multiplication of matrices. In Matlab .* multiply two matrices entrywise. The other three types of exceptions are illustrated by example. The proximity-preference parameter is delta in the program and δ in the text. The travel-time matrix is **t** in the text and t in the program. In the text $\tilde{p}$ refers to the estimated working population, while p_tilde is used in the program.

The program is published on the following web address: *www.his.se/commuting*. This means that you do not have to retype the code to run the program, you can just use the published file. In order to run the program for the first model you must save the data and the program to your computer. It is recommended that you first save the Excel file to your Matlab folder. In the next step, you save the program file containing the first program into the same Matlab folder. Then start Matlab and run the program.

4.1.2 Results

In Figure 2 you find the estimated distance-friction parameter per iteration. In Figure 3 you find the estimated proximity-preference parameter per iteration. To keep the first Matlab program as simple as possible the value of the dual function and the parameter values are collected per iteration in the published program. The start values of the distance-friction parameter and the proximity-preference parameter is zero and 5.2212, respectively. This gives the start point (5.2212,0) in Figure 4. In Figure 4, the thick line illustrates the path from the start point to the solution. The value of the dual function per iteration is presented in Figure 5. After about 15 iterations neither the parameters nor the value of the dual function change more than marginally. The model converges at the solution, where the distance-friction parameter is 0.1197 and the proximity-preference parameter is 9.809.

However, nothing prevents us from saving all information during the approach to the solution. By doing some small adjustments in the Matlab program, it is possible to save the parameter values and the value of the dual function at every parameter adjustment, rather than per iteration. From the start point (5.2212,0), the distance friction parameter is adjusted to 0.0075, leading to the point (5.2212,0.0075) in Figure 4. Then the estimated commuting flows are recalculated and the proximity-preference parameter is adjusted to 6.1519, leading to the point (6.1519,0.0075). This ends the first iteration, and is seen as the first step from the start point following the thin line in Figure 4. Hence, iterating and saving results in this way gives a set of steps to the solution. It is of course also an option to just save the final solution values.

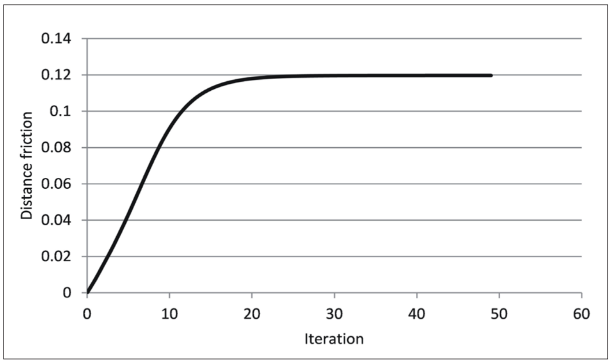


Fig. 2: Distance-friction parameter convergence
Source: author´s elaboration

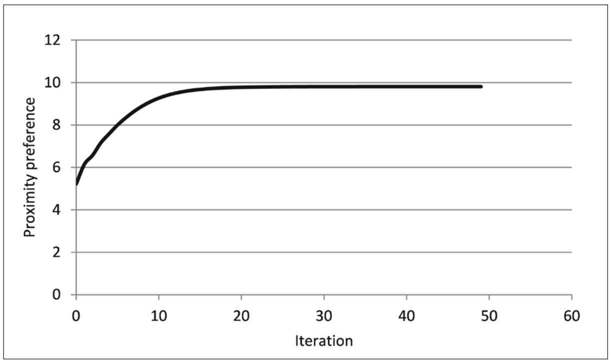


Fig. 3: Proximity-preference parameter convergence
Source: author´s elaboration

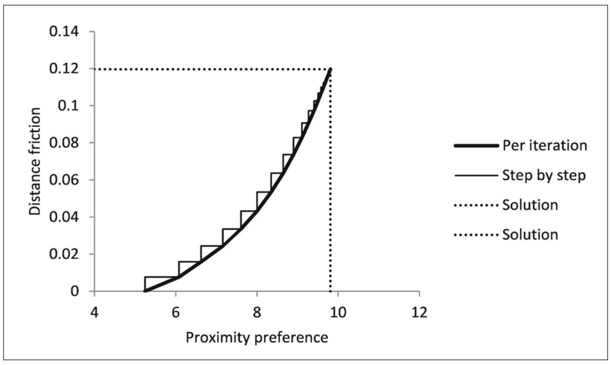


Fig. 4: The two ways to the solution
Source: author´s elaboration

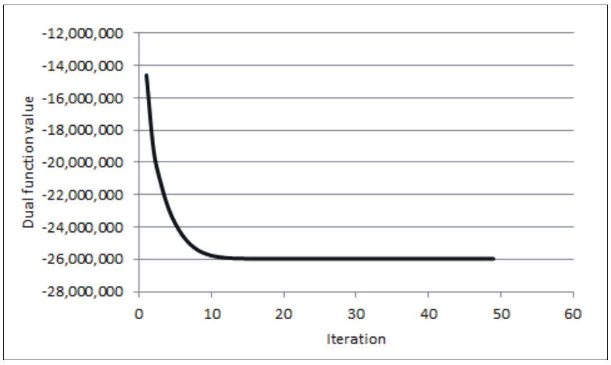


Fig. 5: The value of the dual function per iteration
Source: author´s elaboration

4.2 Model 2

Model 2 has six constraints, and they are similar to the two constraints present in Model 1. The idea is to incorporate more spatial information into the model. Some persons work within their home municipality, while others commute to another municipality within their home region, and some even commute to another region. Model 2 has three constraints replacing the Model 1 constraint enforcing that the estimated working population is equal to the observed population. The observed number of commuters within a municipality is equal to $p_1 = \mathbf{u}(\mathbf{k} \circ \mathbf{c})\mathbf{u}'$. The observed number of commuters between municipalities within the home region is $p_2 = \mathbf{u}(\mathbf{l} \circ \mathbf{c})\mathbf{u}'$. The observed number of commuters between regions is $p_3 = \mathbf{u}(\mathbf{m} \circ \mathbf{c})\mathbf{u}'$. They are collected in the column vector **p**. In this study the working population is divided such that p_1 = 2,812,614, p_2 = 817,802 and p_3 = 217,366. To each constraint there is a proximity-preference parameter, all collected in the column vector δ. In Model 2, three constraints replace the constraint regarding total commuting time present in Model 1. The observed total commuting time for commutes within a municipality equals $r_1 = \mathbf{u}(\mathbf{k} \circ \mathbf{c} \circ \mathbf{t})\mathbf{u}'$. The observed total commuting time for commutes between municipalities within the home region is $r_2 = \mathbf{u}(\mathbf{l} \circ \mathbf{c} \circ \mathbf{t})\mathbf{u}'$. The observed total commuting time for commutes between regions is $r_3 = \mathbf{u}(\mathbf{m} \circ \mathbf{c} \circ \mathbf{t})\mathbf{u}'$. They are collected in the column vector **r**. To each time constraint there is a distance-friction parameter, and they are collected in the column vector γ. This model is like splitting Model 1 into three completely separate parts. The primal form of the problem is max $\mathrm{L}(\tilde{\mathbf{c}}, \delta, \gamma)$, where $\mathrm{L}(\tilde{\mathbf{c}}, \delta, \gamma) = \sum_{s=0}^{6} \mathrm{L}_s$ and the Lagrangian parts L_s are defined in (1)–(7).

$$\mathrm{L}_0 = -\,\mathbf{u}(\tilde{\mathbf{c}} \circ \ln(\tilde{\mathbf{c}}) - \tilde{\mathbf{c}})\mathbf{u}' \quad (1)$$

$$\mathrm{L}_1 = \delta_1(\mathbf{u}(\mathbf{k} \circ \tilde{\mathbf{c}})\mathbf{u}' - p_1) \quad (2)$$

$$\mathrm{L}_2 = \delta_2(\mathbf{u}(\mathbf{l} \circ \tilde{\mathbf{c}})\mathbf{u}' - p_2) \quad (3)$$

$$\mathrm{L}_3 = \delta_3(\mathbf{u}(\mathbf{m} \circ \tilde{\mathbf{c}})\mathbf{u}' - p_3) \quad (4)$$

$$\mathrm{L}_4 = \gamma_1(r_1 - \mathbf{u}(\mathbf{k} \circ \tilde{\mathbf{c}} \circ \mathbf{t})\mathbf{u}') \quad (5)$$

$$\mathrm{L}_5 = \gamma_2(r_2 - \mathbf{u}(\mathbf{l} \circ \tilde{\mathbf{c}} \circ \mathbf{t})\mathbf{u}') \quad (6)$$

$$\mathrm{L}_6 = \gamma_3(r_3 - \mathbf{u}(\mathbf{m} \circ \tilde{\mathbf{c}} \circ \mathbf{t})\mathbf{u}') \quad (7)$$

You find the three constraints for the number of commuters in (2)–(4) and the three constraints on total commuting time in (5)–(7). This is similar to the earlier model, and the adjustment process to find the six Lagrangian multipliers is therefore straight forward. The derivative of the Lagrangian with respect to commuting gives the estimated commuting matrix $\tilde{\mathbf{c}} = \exp(\delta_1\mathbf{k} + \delta_2\mathbf{l} + \delta_3\mathbf{m} - (\gamma_1\mathbf{k} + \gamma_2\mathbf{l} + \gamma_3\mathbf{m}) \circ \mathbf{t})$. By inserting this into the Lagrangian we get the dual formulation of the problem, min $\mathrm{D}(\delta,\gamma)$, where $\mathrm{D}(\delta,\gamma) = \mathbf{u} \exp(\delta_1\mathbf{k} + \delta_2\mathbf{l} + \delta_3\mathbf{m} - (\gamma_1\mathbf{k} + \gamma_2\mathbf{l} + \gamma_3\mathbf{m}) \circ \mathbf{t})\, \mathbf{u}' - \delta'\mathbf{p} + \gamma'\mathbf{r}$.

To find reasonable start values, assume that all distance-friction parameters are zero and choose to enforce the three constraints regarding the number of commuters within the home municipality, between municipalities within the home region, and between the regions. Then the start values for the proximity preferences are $\delta_1 = \ln(p_1 / \mathbf{u}(\mathbf{k} \circ \mathbf{z})\mathbf{u}'))$, $\delta_2 = \ln(p_2 / (\mathbf{u}(\mathbf{l} \circ \mathbf{z})\mathbf{u}'))$, and $\delta_3 = \ln(p_3 / (\mathbf{u}(\mathbf{m} \circ \mathbf{z})\mathbf{u}'))$, respectively. If you compare these start values to the start value in Model 1 you see the similarity. Collect the derivatives $\tilde{s}_1 = \mathbf{u}(\mathbf{k} \circ \mathbf{t} \circ \mathbf{t} \circ \tilde{\mathbf{c}})\mathbf{u}'$, $\tilde{s}_2 = \mathbf{u}(\mathbf{l} \circ \mathbf{t} \circ \mathbf{t} \circ \tilde{\mathbf{c}})\mathbf{u}'$, and $\tilde{s}_3 = \mathbf{u}(\mathbf{m} \circ \mathbf{t} \circ \mathbf{t} \circ \tilde{\mathbf{c}})\mathbf{u}'$ in the column vector $\tilde{\mathbf{s}}$. Then the friction vector is adjusted using $\gamma_{(n+1)} = \gamma_n - \rho(\tilde{\mathbf{r}}_n - \mathbf{r})./\,\tilde{\mathbf{s}}_n$, where ./ is the symbol for piecewise division. The estimated commuting flows are recalculated before adjusting the proximity-preferences using $\delta_{(n+1)} = \delta_n - \rho(\tilde{\mathbf{p}}_n - \mathbf{p})./\,\tilde{\mathbf{p}}_n$. Also in Model 2 $\rho = 1$. Before iterating, the estimated commuting flows are recalculated once more.

4.2.1 The Matlab program

You find the program for Model 2 in Appendix 3 (see Supplementary material), and it is also available for downloading at www.his.se/commuting. The overall structure of the program is the same as for Model 1. However, Model 2 uses more spatial information. Therefore the **k**, **l**, and **m** matrices are also read from the Excel file. With them the new necessary vectors **p** and **r** are calculated. In the main part of the program, the parameters are adjusted. First, the distance-friction vector is adjusted in relation to the relevant constraint deviation. In this part **s** is calculated. Second, the proximity-preference vector is adjusted. This is the same as the adjustment procedure used in Model 1. A comment on notation: In the text for example $\tilde{s}_2$ refers to the second value in $\tilde{\mathbf{s}}$. In the Matlab program s_tilde(2) does that job. This is the principle used for any vector or matrix.

4.2.2 Results

In Figure 6 you find the distance-friction parameters per iteration. In Figure 7 you find the proximity-preference parameters per iteration. At the start the distance-friction parameters are set to zero, and the proximity-preferences are 9.1832, 5.9328, and 2.4734, for commuting within a municipality (i.e. local), commuting between municipalities within a region (i.e. regional) and between regions, respectively. The solution for the distance-friction parameters are 0.0294, 0.1027, and 0.0483. The solutions for the proximity-preference parameters are 9.6335, 8.5289, and 6.1309.

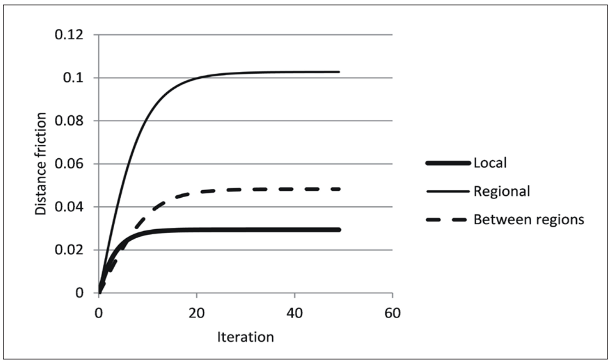


Fig. 6: Distance-friction parameter convergence
Source: author´s elaboration

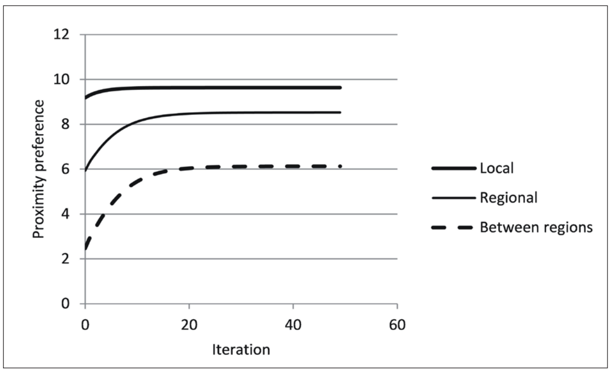


Fig. 7: Proximity-preference parameter convergence
Source: author´s elaboration

In Figure 8 you find the proximity preference and distance friction pairs from the start to the solution. Here the convergence process starts from points along the x-axis. In Fig. 9, you find the value of the dual function per iteration.

The solution is found after about 15 iterations. Then nothing much happens to the parameters and the value of the dual function. At the solution, the value of the dual function is smaller for Model 2 compared to the value for Model 1. This is expected, since Model 2 enforces more constraints. The first two models are only presented as the way to the final model. However, we can compare the results from Model 1 and Model 2 anyway. Both the proximity-preference parameter and the distance-friction parameter are higher in Model 1. Model 1 replicates the commuting pattern (Fig. 1) with one exponential function. The proximity-preference parameter is related to the intersection with the y-axis. The distance-friction parameter is related to the decline of commuting as commuting time is increased. This is illustrated in in Figure 10 by the dotted line. Model 2 replicates the pattern using three exponential functions. This is seen in Figure 10 as one solid line for commuting within the home municipality, one solid line for commuting between municipalities within the home region, and one solid line for commuting between regions. Model 3 has 582 unique constraints. In that model, the pattern is replicated using 20,781 (out of maximum 83,521) exponential functions.

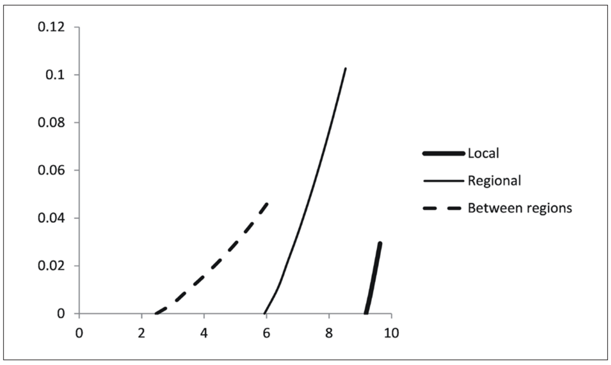


Fig. 8: The paths to the solution
Source: author´s elaboration

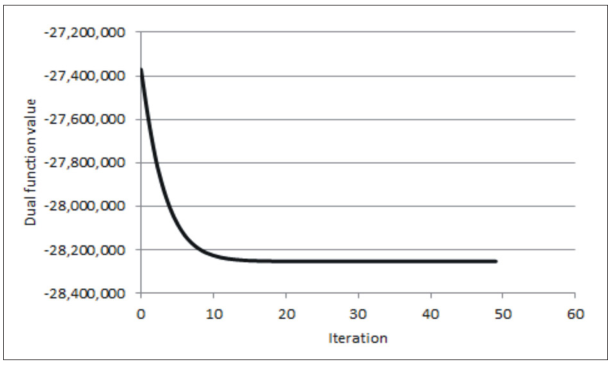


Fig. 9: The value of the dual function per iteration
Source: author´s elaboration

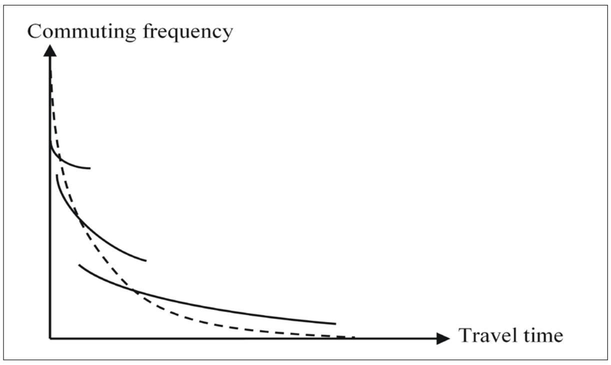


Fig. 10: The principal result of Model 1 and Model 2
Source: author´s elaboration

4.3 Model 3

Model 2 has three distance-friction parameters and three proximity-preference parameters. Those parameters (and constraints) are also present in Model 3, but in addition Model 3 also has commuting origin- and destination constraints. The third model is max $L(\tilde{\mathbf{c}}, \alpha, \beta, \delta, \gamma) = \sum_{s=0}^{8} L_s$, where the Lagrangian parts, L_s, are found in (8)–(16). Model 2 has three constraints for the amount of commuting, and they are included in the same way in Model 3, (11)–(13). Model 2 has three time constraints, and they are included in the same way in Model 3, (14)–(16). Model 3 in addition enforces that the estimated number of workers that live in each municipality is equal to the observed number, $\mathbf{o} = \mathbf{cu}' = \tilde{\mathbf{c}}\mathbf{u}'$. This adds 289 origin constraints, (9). However, only 288 origin constraints provide new information. The three constraints on the number of commuters together enforce that the estimated number of commuters is equal to the observed working population. This makes the 289th origin constraint redundant, since it will be enforced by the other constraints. To each origin constraint there is a Lagrangian multiplier which is called a push factor. They are collected in the column vector α. Because of programming convenience all 289 destination constraints are used, but one origin is used as base, here $\alpha_1 = 0$. Model 3 also enforces that the estimated number of jobs in each municipality is equal to the observed number of jobs, $\mathbf{d} = \mathbf{uc} = \mathbf{u}\tilde{\mathbf{c}}$. This adds 289 destination constraints, (10). As for the origin constraints, one of the destination constraints is redundant, since only 288 destination constraints provide information. To each destination constraint there is a Lagrangian multiplier which is called a pull factor. They are collected in the row vector β. Because of programming convenience all 289 destination constraints are used, but one pull factor is used as base, here $\beta_1 = 0$. The complete model now has 582 constraints. This is the setup in the Matlab program in Appendix 4 (see Supplementary material).

$$L_0 = -\mathbf{u}(\tilde{\mathbf{c}} \circ \ln(\tilde{\mathbf{c}}) - \tilde{\mathbf{c}})\mathbf{u}' \tag{8}$$

$$L_1 = \mathbf{u}(\alpha \circ (\tilde{\mathbf{c}}\mathbf{u}' - \mathbf{o})) \tag{9}$$

$$L_2 = (\beta \circ (\mathbf{u}\tilde{\mathbf{c}} - \mathbf{d}))\mathbf{u}' \tag{10}$$

$$L_3 = \delta_1(\mathbf{u}(\mathbf{k} \circ \tilde{\mathbf{c}})\mathbf{u}' - p_1) \tag{11}$$

$$L_4 = \delta_2(\mathbf{u}(\mathbf{l} \circ \tilde{\mathbf{c}})\mathbf{u}' - p_2) \tag{12}$$

$$L_5 = \delta_3(\mathbf{u}(\mathbf{m} \circ \tilde{\mathbf{c}})\mathbf{u}' - p_3) \tag{13}$$

$$L_6 = \gamma_1(r_1 - \mathbf{u}(\mathbf{k} \circ \tilde{\mathbf{c}} \circ \mathbf{t})\mathbf{u}') \tag{14}$$

$$L_7 = \gamma_2(r_2 - \mathbf{u}(\mathbf{l} \circ \tilde{\mathbf{c}} \circ \mathbf{t})\mathbf{u}') \tag{15}$$

$$L_8 = \gamma_3(r_3 - \mathbf{u}(\mathbf{m} \circ \tilde{\mathbf{c}} \circ \mathbf{t})\mathbf{u}') \tag{16}$$

The program needs reasonable start values. It is assumed that all distance-friction parameters, push- and pull factors are zero. The start values for the proximity preferences are $\delta_1 = \ln(p_1 / \mathbf{u}(\mathbf{k} \circ \mathbf{z})\mathbf{u}'))$, $\delta_2 = \ln(p_2 / (\mathbf{u}(\mathbf{l} \circ \mathbf{z})\mathbf{u}'))$, and $\delta_3 = \ln(p_3 / (\mathbf{u}(\mathbf{m} \circ \mathbf{z})\mathbf{u}'))$, respectively, which is exactly the same as is used in Model 2.

The partial derivative of the Lagrangian with respect to commuting gives the estimated commuting matrix $\tilde{\mathbf{c}} = \exp(\alpha\mathbf{u} + \mathbf{u}'\beta + \delta_1\mathbf{k} + \delta_2\mathbf{l} + \delta_3\mathbf{m} - (\gamma_1\mathbf{k} + \gamma_2\mathbf{l} + \gamma_3\mathbf{m}) \circ \mathbf{t})$ Inserting this into the Lagrangian gives the dual form, $\min D(\alpha,\beta,\gamma,\delta)$, where $D(\alpha,\beta,\gamma,\delta) = \mathbf{u} \exp(\alpha\mathbf{u} + \mathbf{u}'\beta + \delta_1\mathbf{k} + \delta_2\mathbf{l} + \delta_3\mathbf{m} - (\gamma_1\mathbf{k} + \gamma_2\mathbf{l} + \gamma_3\mathbf{m}) \circ \mathbf{t}) \mathbf{u}' - \alpha'\mathbf{o} - \beta\mathbf{d}' - \delta'\mathbf{p} + \gamma'\mathbf{r}$.

Model 3 has four groups of parameters, and each group is adjusted separately. In the program the push factors are adjusted first. The origin constraints are constraints on the number of commuters. In that way they are similar to the three constraints on the number of commuters within the home municipality, between municipalities within the home region, and between regions. Therefore, how to adjust the push factors are easily inferred. The push factors are adjusted using $\alpha_{(n+1)} = \alpha_n - \rho(\tilde{\mathbf{o}}_n - \mathbf{o})./\,\tilde{\mathbf{o}}_n$. After recalculating the estimated commuting flows the pull factors are adjusted, $\beta_{(n+1)} = \beta_n - \rho(\tilde{\mathbf{d}}_n - \mathbf{d})./\,\tilde{\mathbf{d}}_n$, in a similar way. The estimated commuting flows are recalculated before the distance-friction vector is adjusted. At the end of the iteration, the proximity-preference vector is adjusted and the estimated commuting flows are recalculated once more. The distance-friction vector and the proximity-preference vector are adjusted as described above in Model 2. Compared to the previous models, Model 3 is more complex. In Model 3 the number of constraints is larger, and the constraints are interwoven. For that reason the relative adjustment factor is reduced for convergence, $\rho = 0.2$. It is possible to rerun the program for other adjustment factors, and trace the way to the solution in each case.

4.3.1 The Matlab program

You find the Matlab program for Model 3 in Appendix 4 (see Supplementary material). This program is also available to download from www.his.se/commuting. The program has grown to include the adjustment of the push and pull factors. In the main part of the program, the parameters are adjusted. First, the push factors are adjusted. Second, the pull factors are adjusted. Third, the distance-friction parameter vector is adjusted. Fourth, the proximity-preference parameter vector is adjusted. After a set of parameters has been adjusted, the estimated commuting flows are recalculated. In the program it is convenient to keep all 289 push factors and 289 pull factors. Hence, all 289 factors are adjusted using the same procedure, but then one of each factor is set to zero.

4.3.2 Results

At the solution, the distance-friction parameter for commuting within a municipality is 0.0248, the distance-friction parameter for commuting between municipalities within a region is 0.0958, and the distance-friction parameter for commuting between regions is 0.0514. You find the convergence process for the distance-friction parameters in Figure 11.

At the solution, the proximity-preference parameter for commuting within a municipality is 8.5147, the proximity-preference parameter for commuting between municipalities within a region is 7.4679, and the proximity-preference parameter or commuting between regions is 5.4938. The convergence processes for the proximity-preference parameters are illustrated in Figure 12. In Figure 13 the proximity-preference parameter and distance-friction parameter pairs from the start (along the x-axis) to the solution are illustrated. In the background, the 288 push and 288 pull factors are adjusted as well.

You find the value of the dual function per iteration in Fig. 14. Little happens to the parameter values and value of the dual function after 200 iterations. However, by iterating more the solution is pinpointed. The program is set to do 500 iterations. The solution value of the dual function

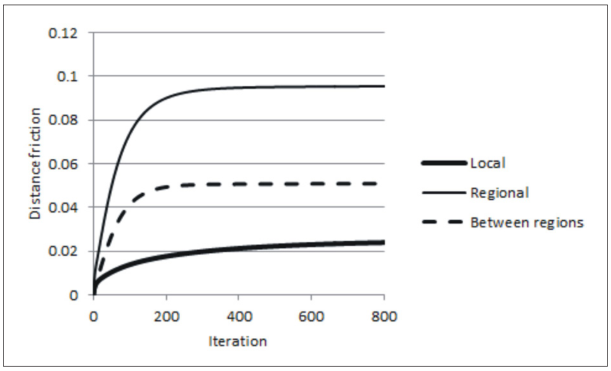


Fig. 11: Distance-friction parameter convergence
Source: author´s elaboration

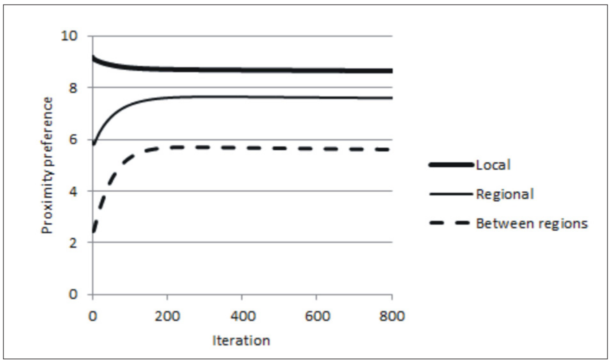


Fig. 12: Proximity-preference parameter convergence
Source: author´s elaboration

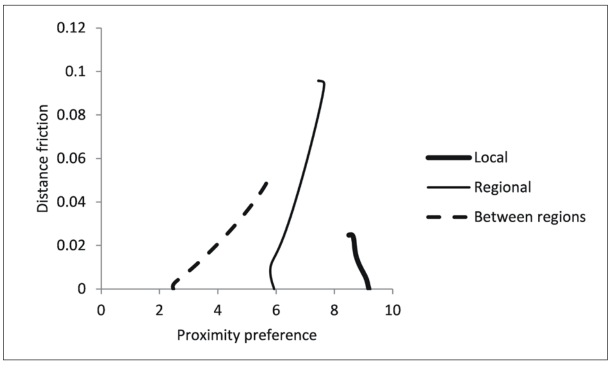


Fig. 13: The paths to the solution
Source: author´s elaboration

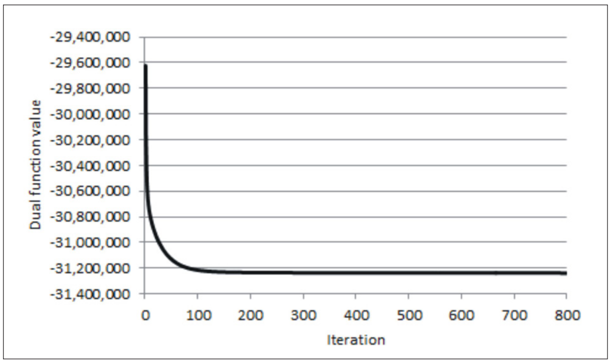


Fig. 14: The value of the dual function per iteration
Source: author´s elaboration

is smaller for Model 3. This is as expected, since Model 3 enforces many additional constraints.

4.4 Model 3 alternatives

In Model 3 the first municipality is used as base case, and hence both α_1 and β_1 are set to zero. Then, every other parameter is estimated in relation to them. Of course, any other municipality could serve as base. Using another municipality as base implies setting two other parameters to zero. This would give other solution push factors, pull factors, and proximity-preference parameters. It is their combined effect that is interesting. The combined effect remains the same in all cases. Moreover, the distance-friction vector is the same in all cases.

Another alternative would be to for example set β_1 and δ_3 to zero. Then the model has 289 origin constraints, and 289 push factors. To find start values all pull factors, all distance-friction parameters, and the two proximity-preference parameters are set to zero. Then it follows that $\alpha_0 = \ln(\mathbf{o}/289)$ in the case that all destinations are included for all origins. In this study the commuting time from a municipality should be smaller than 150 minutes in order for a destination to be included in the commuting zone. Then the start values are $\alpha_0 = \ln(\mathbf{o}./(\mathbf{zu}'))$. Obviously, you could have chosen another of the pull factors and push factors instead of β_1 and another of the proximity preferences instead of δ_3. This would give other push factors, pull factors, and proximity preference parameters. However, their combined effect is the same, nevertheless. The proximity-preference vector differs between model set ups, however the proximity-preference parameter differences are maintained in all set ups. Moreover, the distance-friction vector is the same in all cases.

Sometimes you see studies that set no parameter value to zero, i.e. all constraints are used, even though in Model 3 two constraints contain no new information. This works, since the parameters are estimated in relation to each other. In such a case, there are several sets of feasible start values to choose from. The solution push factors, pull factors, and proximity-preference parameters change with start values. However, this procedure gives the same parameter estimates for distance friction and proximity-preference parameter differences. Still, it is not good practice to include constraints with no information. Under such circumstances one needs to be careful when interpreting the results. When that is done properly, you find that the results are the same as you get if you only use constraints with real information.

4.5 Model comparisons

In this paper, three models of commuting have been presented. The idea was to start from a simple model, and gradually add spatial constraints to the model to better capture reality. It is interesting to see how well the models estimate the observed commuting pattern.

In Tables 4–6, you find descriptive statistics for (c.)□c □ for Model 1–3, respectively. For example, the median value for observed commuting as a share of estimated commuting within a municipality is 1.7, for Model 1 (Tab. 4). The corresponding median values for Model 2 and Model 3 are 0.5 (Tab. 5) and 1.0 (Tab. 6), respectively. Note that the median gets closer and closer to one. This is the case also for commuting between municipalities within a region, and for commuting between regions. It is also the case that the standard deviation is smaller in Model 3 than in Model 1. Model 3 has many more constraints and therefore performs better. This is also seen by that the means converge.

In Table 6, the standard deviation is relatively large for commuting between regions. One possible explanation for this is that there are some links that deviate from the pattern due to commuting by other means than car, i.e. train. Such flows are not accurately captured in this model.

5. Discussion and conclusion

Given the third model, you can create a version of the model by altering the set-up. You could for example just use one commuting time constraint instead of three. That means that you estimate only one distance-friction parameter. You could also remove the constraints for the number of commuters within a municipality, between municipalities within a region, and between regions. That means that you estimate no proximity-preference parameters. In such a version of the model, you must allow all 289 push- and pull-factors to adjust using the described procedure. The resulting distance-friction parameter is 0.1406. For this version of the model the median number of $\mathbf{c}/\tilde{\mathbf{c}}$ is 2.0, 0.5, 31.0 for commuting within a municipality, between municipalities within a region, and between regions, respectively. This can be compared to the corresponding numbers for Model 3 in Table 6.

As expected, Model 3 outperforms a version of the model using less constraints. It is also possible to alter Model 3 in other ways. Model 3 uses 20,781 links out of the maximum 83,521 links, and some of those links are not active (Table 2). It is straight forward to change the code such that only the 12,287 active links out of the 20,781 links

Measure	Within municipality	Within region	Between regions
Min	0.2	0.0	0.0
Median	1.7	0.1	8.9
Mean	62.0	0.7	1,298
Max	6,472	59.4	655,370
Std. dev.	447.6	2.7	11,985

Tab. 4: Descriptive statistics for the predictive performance of Model 1. Source: author´s calculation

Measure	Within municipality	Within region	Between regions
Min	0.1	0.0	0.0
Median	0.5	0.3	0.7
Mean	1.0	1.2	3.3
Max	21.6	90.5	687.8
Std. dev.	1.8	3.9	16.7

Tab. 5: Descriptive statistics for the predictive performance of Model 2. Source: author´s calculation

Measure	Within municipality	Within region	Between regions
Min	0.8	0.0	0.0
Median	1.0	0.6	0.8
Mean	1.1	0.9	2.6
Max	2.8	31.4	661.1
Std. dev.	0.2	1.5	10.5

Tab. 6: Descriptive statistics for the predictive performance of Model 3. Source: author´s calculation

are used ($\tilde{\mathbf{c}}(\mathbf{c} = = 0) = 0$). The resulting estimates for the distance-friction parameters do not differ much between these two versions of Model 3. Nothing prevents us from adding more constraints to Model 3. You could add, for example, housing expenditure and income constraints. Such models and results are discussed in Olsson (2015).

Hopefully, this paper has stimulated you into modelling spatial interaction. When a model like Model 3 has been solved, you have a set of distance-friction parameters. With them you can calculate accessibility measures to incorporate spatial aspects into different types of studies (e.g. the literature presented in the Introduction). Let us assume that we want to look at the accessibility to workers. This would be one important variable to consider when studying, for example, how easy it is to find someone to fill a vacancy. Johansson et al. (2002, 2003) suggest that one separates the total accessibility into three parts: accessibility within the municipality, accessibility in other municipalities within the region, and accessibility in other regions. Such spatial decomposition of the total accessibility is useful in empirical studies, since they likely are of unequal importance. In this study, the commuting pattern was in focus. But not all persons work, e.g. the unemployed, the retired, students, etc. The non-working part of the population also interacts spatially. Although not part of this study, such spatial interactions are also interesting to model. Moreover, those individuals are often included in the accessibility measures (e.g. as potential workers or customers, depending on the focus of the study). Such a spatial decomposition, moreover, does not acknowledge that competition also varies across locations. Geurs and van Wee (2004) identified several ways to introduce competition aspects into the accessibility measures, and one way would be to use the balancing factors of the solution to the gravity model (from α and β).

References:

ANAS, A. (1983): Discrete choice theory, information theory and the multinomial logit and gravity models. Transportation Research Part B: Methodological, 17(1): 13–23.

ANDERSSON, M., EJERMO, O. (2005): How does accessibility to knowledge sources affect the innovativeness of corporations? – Evidence from Sweden. The Annals of Regional Science, 39(4): 741–765.

ANDERSSON, M., GRÅSJÖ, U. (2009): Spatial dependence and the representation of space in empirical models. The Annals of Regional Science, 43(1): 159–180.

ANDERSSON, Å., STRÖMQUIST, U. (1988): K-samhällets framtid. Värnamo, Prisma.

BACKMAN, M. (2013): Human capital in firms and regions: Impact on firm productivity. Papers in Regional Science, 93(3): 557–575.

FISCHER, M. M, WANG, J. (2011): Spatial Data Analysis. Models, Methods and Techniques. Heidelberg, Springer.

GEURS, K. T., VAN WEE, B. (2004): Accessibility evaluation of land-use and transport strategies: Review and research directions. Journal of Transport Geography, 12(2): 127–140.

GRÅSJÖ, U. (2006): Spatial spillovers of knowledge production – An accessibility approach. JIBS Dissertation Series No. 034. Jönköping, Jönköping International Business School.

GRÅSJÖ, U., KARLSSON, C. (2013): Accessibility: a useful analytical and empirical tool in spatial economics – experiences from Sweden [online]. CESIS Electron Work Paper Series No. 314. Available at: https://static.sys.kth.se/itm/wp/cesis/cesiswp314.pdf

GRÅSJÖ, U., KARLSSON, C. (2015): Accessibility: an underused analytical and empirical tool in spatial economics. In: Condeço-Melhorado, A., Reggiani, A., Gutiérrez, J. [eds.]: Accessibility and Spatial Interaction (pp. 211–236). Cheltenham, Edward Elgar.

HANSEN, W. G. (1959): How accessibility shapes land use. Journal of the American Institute of planners, 25(2): 73–76.

JOHANSSON, B., KLAESSON, J., OLSSON, M. (2002): Time distances and labor market integration. Papers in Regional Science, 81(3): 305–327.

JOHANSSON, B., KLAESSON, J., OLSSON, M. (2003): Commuters' non-linear response to time distances. Journal of Geographical Systems, 5(3): 315–329.

JOHANSSON, S., KARLSSON, C. (2007): R&D accessibility and regional export diversity. The Annals of Regional Science, 41(3): 501–523.

KARLSSON, C., OLSSON, M. (2006): The identification of functional regions: theory, methods and applications. The Annals of Regional Science, 40(1): 1–18.

LARSSON, J. P. (2014): The neighborhood or the region? Reassessing the density-wage relationship using geo-coded data. The Annals of Regional Science, 52(2): 367–384.

LARSSON, J. P., ÖNER, Ö. (2014): Location and co-location in retail: a probabilistic approach using geo-coded data for metropolitan retail markets. The Annals of Regional Science, 52(2): 385–408.

MATTSSON, L-G. (1984): Equivalence between Welfare and Entropy Approaches to Residential Location. Regional Science and Urban Economics, 14(2): 147–173.

OLSSON, M. (2002): Studies of Commuting and Labour Market Integration. JIBS Dissertation Series No. 16. Jönköping, Jönköping International Business School.

OLSSON, M. (2012): Free versus monitored job search in Sweden. In: Karlsson, C., Johansson, B., Stough, R. R. [eds.]: The regional economics of knowledge and talent, Local advantage in a global context (pp. 194–209). Cheltenham, Edward Elgar.

OLSSON, M. (2015): The Swedish Commuting Pattern. A Constrained Gravity Model with Housing Expenditure and Income Constraints. In: Bernhard, I. [ed.]: Regional Development in an International Context. Regional, National, Cross Border and International Factors for Growth and Development (pp. 479–486). Trollhättan, University West.

PERSSON, H. (1986): Algorithm 12: Solving the entropy maximization problem with equality and inequality constraints. Environment and Planning A, 18(12): 1665–1676.

SEN, A., SMITH T. E. (2011): Gravity models of spatial interaction behavior. Heidelberg, Springer.

SIKA (2000): Transports and communications, Yearbook 2000/2001. Värnamo, The SIKA-institute.

STATISTICS SWEDEN (2015): Local Labor Markets (in Swedish) [online]. Available at: http://www.scb.se/sv_/Hitta-statistik/Statistik-efter-amne/Arbetsmarknad/Sysselsattning-forvarvsarbete-och-arbetstider/Registerbaserad-arbetsmarknadsstatistik-RAMS/7899/Lokala-arbetsmarknader-LA/

TRANSPORT ANALYSIS (2013): RVU_Sverige_2011.xls [online]. Available at: http://www.trafa.se/sv/Statistik/Resvanor/

Supplementary material: http://www.his.se/commuting/

Please cite this article as:

OLSSON, M. (2016): Functional regions in gravity models and accessibility measures. Moravian Geographical Reports, 24(2): 60–70. Doi: 10.1515/mgr-2016-0011.

MORAVIAN GEOGRAPHICAL REPORTS

Aims and Scope of the Journal

Moravian Geographical Reports [MGR] is an international peer-reviewed journal, which has been published in English continuously since 1993 by the Institute of Geonics, Academy of Sciences of the Czech Republic, through its Department of Environmental Geography. It receives and evaluates articles contributed by geographers and by researchers who specialize in related disciplines, including the geosciences and geo-ecology, with a distinct regional orientation, broadly for countries in Europe. The title of the journal celebrates its origins in the historic land of Moravia in the eastern half of the Czech Republic. The emphasis for MGR is on the role of 'regions' and 'localities' in a globalized society, given the geographic scale at which they are evaluated. Several inter-related questions are stressed: problems of regional economies and society; society in an urban or rural context; regional perspectives on the influence of human activities on landscapes and environments; the relationships between localities and macro-economic structures in rapidly changing socio-political and environmental conditions; environmental impacts of technical processes on bio-physical landscapes; and physical-geographic processes in landscape evolution, including the evaluation of hazards, such as floods. Theoretical questions in geography are also addressed, especially the relations between physical and human geography in their regional dimensions.

The MGR journal publishes the following types of papers:

Original scientific papers are the backbone of individual journal issues. These contributions from geography and regionally-oriented results of empirical research in various disciplines, normally have theoretical and methodological sections and must be anchored in the international literature. We recommend following the classical structure of a research paper: introduction, including objectives; theoretical and methodological bases for the work; empirical elaboration of the project; evaluation of results and discussion; conclusions and references. With the exception of purely theoretical papers, each contribution should contain colour graphic enclosures such as maps, charts, diagrams, photographs, etc. Some of the photographs may be placed on the second, third or fourth cover pages of the journal. For papers on regional issues, a simple map indicating the geographical location of the study region should be provided. Any grant(s) received to support the research work should be acknowledged. Major scientific papers include an Abstract (up to 150 words) and 3 to 6 keywords. The length of the text should be in the range of 6,000 – 8,000 words (the word count does not include the abstract, tables, figures, and references), plus a maximum of 3 pages of enclosures (tables, figures). The number of graphic enclosures can be increased by one page provided the text is shortened by 500 words below the maximum allowable length (per graphic page). All scientific papers are subject to the peer-review process by at least two independent reviewers appointed by the Editorial Board.

Scientific communications are published to inform the public of continuing research projects, scientific hypotheses or findings. This section is also used for scientific discussions that contest or refine scientific opinions, including viewpoints and/or comments that critique recently- published papers. The maximum text length for these scientific communications is 4,000 words. Some contributions may be reviewed at the discretion of the Editorial Board.

MGR also publishes **Invited reviews** of major monographs from geography and related disciplines published as books or atlases. Reviews are supplied exclusively on request from the Editorial board. The review must contain a complete citation of the reviewed work and its maximum text length is 2,000 words. Graphics are not expected for the reviews section.

Detailed Guide for Authors is available at the journal webpage:

http://www.geonika.cz/EN/research/MGR_instructAuthors.pdf