Functional geographies through the R package LabourMarketAreas

Keywords: R package LabourMarketAreas, functional geographies

National Statistical Institutes are requested to provide increasingly more detailed information reflecting the underlying structure of the society at which policy decisions need to be focused. There is a growing request not only for higher geographical detail related to administrative boundaries, but also for meaningful statistics for the same geographical areas. Labour Market Areas (LMAa) have long being recognised as relevant for assessing the effectiveness of local policy decisions in labour-related matters. Am LMA is a functional geographic area defined for purposes of compiling, reporting and evaluating employment, unemployment, workforce availability and related topics. LMAs are generally intended as areas built on the basis of commuting to work data so that the majority of the labour force lives and works within their boundaries.

The LMA delineation is a complex process that deals with commuting data availability, clustering algorithms and spatial statistics concepts. Istat has developed the R package LabourMarketAreas to ease the LMA production process. This paper illustrates the main features of the R package LabourMarketAreas.

1. INTRODUCTION

Labour market areas (LMAs) are sub-regional geographical areas where the bulk of the labour force lives and works, and where establishments can find the main part of the labour force necessary to occupy the offered jobs.

The LMA delineation is a complex process that deals with commuting data availability, clustering algorithms and spatial statistics concepts. The necessary input for the delineation of the LMAs is a commuting flow matrix and the topology of the basic territorial units. The latter are used to establish the necessary contiguity inside the LMAs. Traditional decennial population censuses fully based on field enumeration allow to choose a precise definition of the commuters (those who regularly leave the residence to go to work and come back in the evening) and to compile a complete origin-destination matrix of travel to work flows.

LMAs represent a functional geography i.e. a geography based on relationships between elementary territorial units: these can be municipalities, census output areas, provinces, etc. In what follows such elementary unit is called community. A LMA is an aggregation of communities; such aggregation is made in order to maximize the internal relationships amongst communities. The relationships are represented by commuting flows (travel-to-work flows).

LMAs are clusters comprising two or more communities. To be called LMAs, such clusters should satisfy a set of principles already outlined principles in Eurostat (1992) and recognized in the scientific literature as of primary importance. These principles comprise objectives of LMA (to be statistically defined areas each representing a labour market); constraints (being a partition of the whole country, each part formed by contiguous communities); criteria (autonomy, homogeneity, coherence and conformity) and finally the need to be flexible in order to cope with elementary units that could be of very different sizes. In the R package LabourMarketAreas, Ichim et al. (2018), the clusters are identified by means of a rule-based iterative agglomerative algorithm. This method represents an evolution of the classical methodology of the "Travel-To-Work-Areas", defined in Coombes et al. (1986). The initial TTWA algorithm has been used also in Italy (Istat, 1997 and Istat and IRPET, 1989) to create the LMAs for the years 1981, 1991 and 2001. The 2011 version of the Italian LMAs has been developed using the algorithm described in this paper (Istat, 2014).

2. METHODS

By means of a modular structure, the R package LabourMarketAreas takes into account each stage of the LMA delineation process. The implementation consists of a series of functions; each function addresses a specific stage of the LMAs delineation process. The core function of the package is *findClusters*. This function implements a greedy algorithm which may be used to find the territorial partition representing the LMAs. Starting from the basic communities, the algorithm iteratively aggregates them until all clusters satisfy a pre-defined validity criterion, see (Coombes et al. 1986).

Four parameters, called minSZ, minSC, tarSZ and tarSC may be valued by users. These parameters identify the thresholds on the size (minSZ and tarSZ), in terms of number of residents, and on the level of self-containment (minSC and tarSC) required in order for a cluster to be considered an LMA. Consequently, enough flexibility is given in order to control the output in terms of autonomy and heterogeneity, see (Franconi et al. 2017).

The validity condition establishes the criterion that should be met by a cluster in order to be considered an LMA. The condition is operatively defined through a function expressing the trade-off between dimension (in terms of occupied persons), SZ, and the self-containment, SC, of the cluster. In the R package LabourMarketAreas the following validity function, f_v, is implemented:

$$f_v(SZ,SC) = \left[1 - \left(1 - \frac{minSC}{tarSC}\right) \cdot \max\left(\frac{tarSZ - SZ}{tarSZ - minSZ}, 0\right)\right] \left[\frac{\min(SC, tarSC)}{tarSC}\right]$$

Consequently, the validity condition states that a cluster with size SZ_c and self-containment SC_c is a proper LMA if $f_v(SZ_c, SC_c) \geq \frac{minSC}{tarSC}$.

The algorithm is an aggregative one therefore it is necessary to define a measure that establishes to which cluster the selected community ought to be aggregated in order to maximize the interaction of the communities inside the cluster. The measure of interaction between community *h* and cluster *k*, L_{hk} , (Smart, 1974) is called cohesion measure. It is defined by means of the sums of inflow, f_{hk}, and outflow, f_{kh}, between the involved entities standardized w.r.t. the total number of residents, R_h, and total number of workers, W_h:

$$L_{hk} = \left[\frac{(f_{hk})^2}{(R_h W_k)}\right] + \left[\frac{(f_{kh})^2}{(R_k W_h)}\right]$$

In brief, at each iteration, the algorithm assigns the community C having the least validity to the cluster P showing the greatest cohesion (link) with C. If the validity of the cluster P is not increased, the community C is assign to a reserve list. The algorithm converges when each cluster has a validity greater than 1. The communities in the reserve list are assigned to the existing clusters by relaxing the validity condition.

The implemented LMA delineation process consists of four subsequent stages. Table 1 lists them and the corresponding tools in the package. In the first stage, the rule-based algorithm is applied by using the user-defined parameters. A territorial partition characterised by some core figures such as size of LMAs or flows between LMAs represents the main output of the algorithm. When community geographical coordinates are available in geospatial vector format, tools to deploy this information at LMA level are included in the package in order to visualise them.

Table 1 - Stages of LMA production process and tools in LabourMarketAreas

Stage	Action	Tool name
1	Identification of a partition by applying the rule-based algorithm	findClusters AssignLmaName CreateLMAShape
2	Quality assessment and compliance to core LMA defining principles	StatClusterData StatReserveList
3	Validation of the spatial contiguity constraints (fine tuning)	FindIsolated FindContig FineTuning
4	Sensitivity analysis and visualisation	PlotLmaCommunity CompareLMAsStat

Quality assessment (second stage) is performed by analysing statistical indicators provided by the package. Such statistics describe areas, flows and the entire partition; they allow testing the compliance of the solution to LMA principles. Examples of area level indicators include the level of self-containment, the number of attracting communities and commuters living in the area measuring, respectively, autonomy, relevance and homogeneity characteristics.

As the greedy algorithm is based solely on commuting flows, some areas may include communities which are not spatially contiguous. Table 1 lists the R package LabourMarketAreas tools which identify non-contiguous areas and re-assign communities accordingly. This semi-automatic fine tuning procedure is based on both visualisation and strength of links between geographical entities.

Finally, in the last stage, a sensitivity analysis can be performed by setting different groups of parameters in the algorithm. In this way the multivariate optimisation problem of finding the most appropriate partition can be addressed.

The choice of developing free and open source software has many advantages, including the obvious cost reduction. The usage of shared methodology and software tool for LMA identification is a major contribution to achieving harmonisation and comparability at EU level. Other National Statistical Institutes (NSIs) have used the package to produce LMAs (e.g. Denmark) see Soares (2017) or Tosheva et al. (2018) for example. In Figure 1, the adoption of LMAs in official statistics across Europe is represented.

3. CONCLUSIONS AND FURTHER WORK

The paper briefly describes the main salient features of the R package LabourMarketAreas which is released on the CRAN web-site. In the full paper, other issues related to the LMAs production process might be further discussed: a) availability and quality of commuting matrix in presence of mixed-mode data acquisition or source models, b) management of extreme communities, c) different ways to manage the reserve list, d) complete automatisation of the fine tuning process, etc.

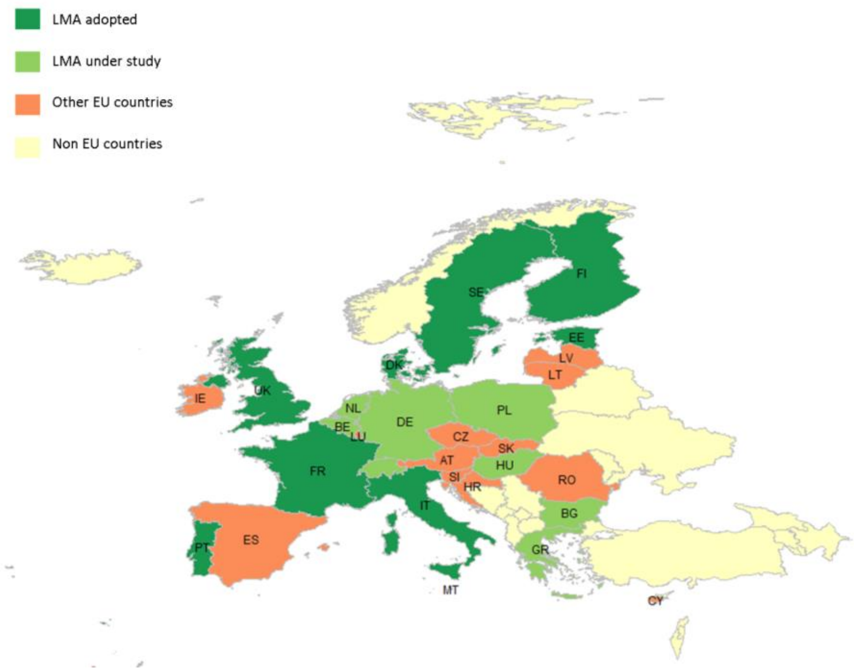

Figure 1. LMAs across Europe.

REFERENCES

[1] Eurostat (1992) Study on employment zones (E/LOC/20) Office for Official Publications of the European Communities, Luxembourg.

[2] Ichim D, Franconi L, D'Aló M, van den Heuvel G. Package LabourMarketAreas v. 3.2: Identification, Tuning, Visualisation and Analysis of Labour Market Areas; 2018.

[3] M.G. Coombes, A.E.Green, S. Openshow, 1986, An efficient algorithm to generate official statistics report areas: the case of the 1984 Travel-to-Work Areas in Britain. The Journal of Operational Research Society. 37(10): 943–953.

[4] Istat and IRPET, 1989. I Mercati Locali del Lavoro. Franco Angeli. Milano. Italia.

[5] Istat. La nuova geografia dei sistemi locali. Roma: Istat; 2015. Available from: http://www.istat.it/it/files/2015/10/La-nuova-geografia-dei-sistemi-locali.pdf.

[6] L. Franconi, D. Ichim, M. D'Alò (2017). Labour Market Areas for territorial policies: Tools for a European approach. Statistical Journal of the IAOS. 33 (3), 581-584.

[7] M. Smart,1974. Labour market areas: uses and definitions. Progress in Planning. 2: 239–353.

[8] Soares E, Figueiredo R, Vale F. Defining Labour Market Areas and its relevance from a statistical perspective: the Portuguese case. SJIAOS 2017; 2017; 33(3): 615–625.

[9] V. Angelova-Tosheva, Teodora Brandmuller, D. Buono, The method of harmonised Labour Market Areas in Europe, European Conference on Quality in Official Statistics, 2018.