Multi-scale spatial analysis of household car ownership using distance-based Moran's eigenvector maps: Case study in Loire-Atlantique (France)

Pierre Hankach, Pascal Gastineau, Pierre Olivier Vandanjon

Post print version of a paper published in the Journal of Transport Geography

https://doi.org/10.1016/j.jtrangeo.2021.103223

Multi-scale spatial analysis of household car ownership using distance-based Moran's eigenvector maps: Case study in Loire-Atlantique (France)

Pierre Hankach[†], Pascal Gastineau[⋆], and Pierre-Olivier Vandanjon[⋆]

[†]MAST-LAMES, Univ Gustave Eiffel, IFSTTAR, Bouguenais, France
[⋆]AME-SPLOTT, Univ Gustave Eiffel, IFSTTAR, Bouguenais, France

January 2022

Abstract

Analyzing spatial structures of transportation data at various scales can be of prime interest to transportation planning and governance. In recent years, multi-scale spatial analysis methods have been developed and used in fields like ecology and geography, but only a few studies have applied these methods to transportation data. However, such methods can provide an efficient exploratory tool for: identifying those scales at which transportation data vary spatially; modeling the spatial structures at each scale; and determining the processes at work that explain these spatial structures. This paper describes and demonstrates how a multi-scale spatial analysis method, namely distance-based Moran's eigenvector maps (dbMEM), can be applied to study the spatial layout of car ownership. For this analysis, we rely on aggregated census data for small statistical areas within France's Loire-Atlantique administrative region. At first, 176 spatial vectors representing spatial patterns with a positive autocorrelation are constructed. Among the 176 vectors, only 23 significant ones are retained after performing a regression with car ownership as the dependent variable. Next, we divide these spatial vectors into three sub-models representing three spatial scales: broad scale, medium scale, and fine scale. Lastly, we identify a set of sociodemographic factors capable of explaining the spatial variation at each scale, i.e.: the broad-scale variation is mainly explained by population density, couples with children and income variables; the medium scale by couples with children, share of individuals in the 25-54 year age range and income; and the fine scale by couples with children and income variables.

Keywords— distance-based Moran's eigenvector maps; car ownership; spatial structure; spatial vectors; spatial scales; multi-scale spatial analysis; sociodemographic factors; variation partitioning

1 Introduction

Transportation is highly correlated with spatial organization at the local, regional and global levels. Transportation data may display recognizable spatial structures at many scales. For our purposes herein, scale refers to the unit of measurement, e.g. broad scale is used to describe phenomena with large extents, while fine scale for those with small extents. Analyzing such spatial structures at each scale could play a very important role in understanding the specific transportation-related context and can therefore be of prime interest to policymakers for transportation planning and governance.

Multi-scale spatial analysis can prove to be an efficient exploratory tool for: 1) identifying those scales at which transportation data vary spatially, ranging from the broadest, encompassing the entire studied area, down to the finest scale; 2) modeling the spatial structures at each scale; and 3) determining, at each scale, the processes at work to create these spatial structures, i.e. in explaining the share of transportation data variation at each scale that can be attributed to certain explanatory variables (e.g. sociodemographic) available for the analysis. Different processes are often at work at these various scales to shape spatial structures.

In recent years, Moran's eigenvector maps (MEM) have been developed and applied to analyse spatial data. MEM were developed in two distinct fields: statistical geography (Griffith 1996, 2000*a*,*b*, Boots & Tiefelsdorf 2000), and quantitative community ecology (Borcard & Legendre 2002, Dray et al. 2006, Legendre & Legendre 2012). They were applied 1) to filter the effect of spatial autocorrelation out of model residuals; 2) as a multi-scale spatial analysis method to analyze scale-dependent spatial structures. Multi-scale spatial analysis using Moran's eigenvector maps has yet to be applied more widely to transportation problems (Wang et al. 2013).

This paper describes and demonstrates how distance-based Moran's eigenvector maps (Borcard & Legendre 2002, Dray et al. 2006), a multi-scale spatial analysis method, can be applied to study the spatial layout of household car ownership patterns. Distance-based Moran's eigenvector maps analysis uses space as an explicit predictor. It is based on sets of variables describing spatial structures explicitly, derived from the coordinates of the geographic sites or from neighbourhood relationships among sites. This approach offers a detailed and precise description of spatial structures. Since space is fundamental to transportation planning and governance, analyzing transportation data with this method can prove to be determinant.

Car ownership has received considerable attention in recent years due to its important role in transportation and land use planning. Song & Wang (2017) noted the necessity of understanding the spatial patterns of household car ownership. Various studies carried out have found that geography influences car ownership and moreover that the number of vehicles depends on household location. For example, suburbanization and car ownership have been shown to be closely correlated (AURAN 2006).

Most previous studies on the determinants of household car ownership rates (see Section 2) were conducted at the household level (i.e. disaggregated data). The advantage of this approach is to place the vehicle acquisition decisions made by households in their historical context, e.g. vehicle acquisition or disposal may be tied to changes in household characteristics. However, these studies targeted a sample of the population and were unable to provide a comprehensive view of the study territory. Since disaggregated data are costly to collect, the geographic coverage of surveys remains uneven, making local variations sometimes difficult to identify.

Few works have thus far studied household car ownership by relying on statistical areas as their basic units. A statistical area is a polygon, representing a proportion of a 2-dimensional map, with which one or more aggregated variables can be associated, e.g. car ownership rate, sociodemographic characteristics. This paper analyzes the spatial structures of car ownership rates by using small statistical areas from census data. Patterns of statistical areas are identified at various scales and attributed to a number of explanatory variables. A major advantage of this approach is the lower cost required to construct the database compared with a disaggregated approach. Data can easily be obtained from a number of sources, either public or commercial. For example, periodic population censuses are common and readily available sources for such aggregated data.

In order to represent spatial structures in dbMEM (distance-based Moran's eigenvector maps) analysis, spatial variables are constructed. For this step, the dbMEM method computes a matrix of geographic distances between statistical areas. After truncation (as explained in Section 3), a principal coordinate analysis of the resulting matrix is performed, yielding eigenvalues and eigenvectors (also called spatial eigenfunctions). Taken together, these eigenfunctions depict the multi-scale distance relationships between statistical areas; they model spatial relationships in decreasing order of spatial scale (i.e. the first dbMEM eigenvector with the largest eigenvalue corresponds to the broadest spatial scale). Furthermore, the eigenvectors possess some interesting properties (e.g. orthogonality) that increase their desirability as spatial explanatory variables.

Following this construction step, the variation in car ownership rates is analyzed with respect to the spatial structures represented by the eigenvectors. For this purpose, the eigenvectors are used in generalized linear models as explanatory variables for the car ownership variable. Each significant eigenvector identified by regression explains part of the variation in car ownership. Since the eigenvectors are orthogonal by design, significant ones identified by regression can be grouped into sub-models representing several categories of spatial scales: broad, medium, etc. The subsequent step of the analysis consists of studying, at each scale, the influence of the sociodemographic variables on variation of car ownership rate. The last step of multi-scale spatial analysis consists in partitioning the deviance respectively explained by the sociodemographic variables and the spatial variables through variation partitioning.

The remainder of this paper will be organized as follows. Section 2 presents the existing literature, while Section 3 describes distance-based Moran's eigenvector maps for spatial analysis and other theoretical aspects. Section 4 displays the results of the car ownership study conducted in France's Loire-Atlantique administrative region. These results are then discussed in Section 5, and the final section offers concluding remarks.

2 Literature review

Car ownership has received a good deal of attention in recent years by virtue of its important role played in transportation and land use planning, as well as its relationship with energy, environmental and health issues (Whelan 2007). Moreover, car ownership is considered by some authors as a key mediating link between residential location and choice of travel mode (Ding et al. 2018). Lavery et al. (2013) reports that car owners tend to have lower modality, which helps explain why their use of other modes is low. As such, it constitutes a key variable for urban planning policy-making. According to Van Acker & Witlox (2010), urban planners should focus not only on influencing car use directly but also on indirect measures through car ownership, given that once a car has been purchased, it tends to be used more often.

The main determinants of vehicle ownership usually include variables affecting travel demand (e.g. individual and household attributes) and variables affecting transportation supply (e.g. accessibility and characteristics of the built residential environment (Zegras 2010, Cao & Cao 2014). Table 1, based on Anowar et al. (2014) comprehensive survey of car ownership in addition to more recent papers (Cao et al. 2019, Moeckel & Yang 2016, Jiang et al. 2017, Ding & Cao 2019, Ma et al. 2018), provides the main variables used in previous studies to explain car ownership. These variables are classified into four categories: household demographics, individual attributes, environment and accessibility attributes.

Household	Individual	Built Environment	Accessibility
• family composition • income • education • number of driving licences • number of children • residence type	• age • bicycle possession • socio-professional category • income • education	• density • urban form	• transport speed • public transport availability • distance to city center • travel time

Table 1: Main determinants of vehicle ownership

Most studies in the literature investigate the determinants of car ownership at a disaggregated (individual or household) level; they rely on surveying a sample of individuals in order to correlate the acquisition or disposal of a vehicle with changes in the characteristics of the household, environment or life events (Clark et al. 2016). Disaggregated level studies include Bhat & Pulugurta (1998), Whelan (2007), Potoglou & Kanaroglou (2008), Anastasopoulos et al. (2012) and Oakil et al. (2016). Only a small number of these studies have actually taken into account the spatial dimension, e.g. Adjemian et al. (2010), Paleti et al. (2013) and Clark et al. (2016). Of these, the study of Páez et al. (2013) is the only one to provide evidence that all spatial variability has been properly accounted for by the model.

Few studies have thus far examined vehicle ownership rates at the statistical area level. However, as explained above, a major advantage of this approach is its reliance on more easily accessible data, i.e. obtained from readily available governmental sources and less expensive to collect than survey data.

Geographically Weighted Regression (GWR) (Brunsdon et al. 1998, Clark 2007, Clark & Finley 2010) is one of the approaches employed to study car ownership at the statistical area level. This approach integrates the spatial autocorrelation of regression coefficients in analyzing the spatial distribution of car ownership. GWR fits a series of local models to the data, thus providing a series of locally varying parameter estimates (rather than explicitly accounting for correlations in the residuals). In their seminal paper, using census data for the County of Tyne and Wear in North East England, Brunsdon et al. (1998) showed that car ownership rates are correlated with both social class and male unemployment. Clark (2007) estimated the spatially varying coefficient of the income parameter and hence produced the local income elasticity. Clark & Finley (2010) highlighted the influence of income and population density on the UK car ownership rate at the electoral ward scale.

In another approach, Clark (2007) and Clark & Finley (2010) estimated a spatial error model (SEM) that accounts for spatial dependence in the error terms rather than in the dependent variable. In addition, in Clark & Finley (2010), a hierarchical Bayesian spatial model was proposed to handle residual spatial autocorrelation. Morton et al. (2018) investigated the determinants of car fleet composition (i.e. the dieselization rate) recorded in Northern Ireland through a Spatial Durbin Error Model (SDEM), which accounts for spatial dependence among the error terms as well as the exogenous interaction effect. In considering 890 contiguous statistical areas containing a mean of 2,000 residents, these authors found that the availability of relatively inexpensive diesel fuel in the Republic of Ireland affected the diesel car ownership rate in Northern Ireland. Clark & Rey (2017) explored the local dynamics of household vehicle ownership in the UK, by comparing a classical Markov chain approach with a spatial Markov chain analysis. They showed that the transitions in relative vehicle ownership are influenced by the neighborhood context.

While the above approaches take neighborhood influence into account in car ownership models, none is well suited for an exhaustive identification of spatial structures at various scales. It is crucial to understand however the spatial patterns (including global and local spatial clustering) of car ownership, as Song & Wang (2017) pointed out. This exploratory process can be of major interest to decision-makers while also being valuable to modelers in generating new hypotheses on explanatory variables.

Few methods have been proposed to model spatial structures at all scales using space as an explicit variable. Moran eigenvector spatial filtering (MESF) is one such method that has been proposed for this purpose. This method computes spatial eigenvectors as explicit spatial predictors. Interestingly, MESF was developed independently and nearly simultaneously in two distinct fields, i.e. statistical geography (Griffith 1996, 2000*b*) and quantitative community ecology (Borcard & Legendre 2002). Griffith (1996, 2000*b*) set the goal of filtering the effect of spatial autocorrelation out of model residuals, in transferring this component to a model's conditional mean (i.e. intercept), whereas Borcard & Legendre (2002) sought to explicitly model the multi-scale nature of univariate or multivariate response data. Dray et al. (2006) formalized the theory originally proposed in Borcard & Legendre (2002) as distance based Moran's eigenvector maps (dbMEM).

Griffith & Peres-Neto (2006) showed equivalencies of and differences between these two implementations. They advocated that the two methods can be unified under the class of spatial eigenfunction maps. They also highlighted that one important advantage of these two approaches over any other spatial approach is that they provide a flexible tool that allows the full range of general and generalized linear modeling theory to be applied to ecological and geographical problems in the presence of nonzero spatial autocorrelation.

Software implementations of MESF method were proposed in both quantitative community ecology and statistical geography fields. Dray et al. (2018) implemented Moran's Eigenvector Maps and related methods for the spatial multi-scale analysis of ecological data in the R package adespatial.[1] Murakami & Griffith (2019) proposed efficient algorithms for MESF which are implemented in the R package spmoran (Moran Eigenvector-Based Scalable Spatial Additive Mixed Models).[2] A pedagogical synthesis of spatial (geo-)statistical and spatial econometric methods with R is presented in Yoshida & Murakami (2020).

The dbMEM method has already been successfully used in several ecological applications including: Borcard et al. (2004), Brind'Amour et al. (2005), Legendre et al. (2005), Laliberté et al. (2009), Brind'Amour et al. (2018), Gáspár et al. (2019), Wagner et al. (2017), Cilleros et al. (2017), Brice et al. (2016), Jin et al. (2020), Chen et al. (2020), Santos et al. (2020), Pollice et al. (2020) and Taddeo et al. (2021).

To the best of our knowledge, the MESF method has only very rarely been used to study transportation-related problems. Moniruzzaman & Páez (2012) investigated the transit shares for the city of Hamilton (Canada) by means of a logistic regression model for proportions and using a spatial filtering approach to control for spatial autocorrelation. Wang et al. (2013) applied MESF to land use data and highlighted the relevance of this method for analyzing both land use and transportation data. Griffith (2009) demonstrated the use of MESF to study 2002 German journey-to-work flows among 439 German administrative units. Griffith (2011) used MESF method for the visualization of spatial autocorrelation in the journey-to-work dataset of Pennsylvania, USA. Yet MESF has never been applied to study car ownership. The following section provides a detailed description of the MESF method as formalized in Dray et al. (2006), i.e. the dbMEM multi-scale spatial analysis method.

3 Distance-based Moran's eigenvector maps for multi-scale spatial analysis

dbMEM analysis identifies spatial patterns across the entire range of perceptible scales with a given dataset. This method is based on a computation of the principal coordinates of a matrix of geographic neighbors among the geographic sites (Borcard & Legendre 2002); it can be successfully applied on various spatial designs: linear (transect) and two-dimensional (surface), regular or irregular geographic schemes. This section describes both the construction of dbMEM eigenfunctions and multi-scale spatial analysis.

3.1 Construction of dbMEM spatial vectors

As stated above, dbMEM vectors are obtained by performing a principal coordinate analysis of a truncated matrix of Euclidean distances between defined geographic units. These units are given sites (e.g. statistical areas) for which population statistics are available. Notably, the share of car ownership and other variables are associated with

[1] https://cran.r-project.org/web/packages/adespatial/index.html
[2] https://cran.r-project.org/web/packages/spmoran/index.html

each of these units. dbMEM spatial vectors are constructed according to the four following steps:

1. A 2-dimensional matrix of Euclidean distances (**D**) between geographic unit barycenters is built using the geographic coordinates of the relevant points.
2. A truncated connectivity matrix (**W**) is then constructed. During this operation, geographic distances are considered to belong to one of two groups: small or large distances. The truncation step is performed according to the following rule:

$$w_{ij} = d_{ij} \text{ if } d_{ij} \leq t \quad (1)$$

$$\text{and } w_{ij} = 4t \text{ if } d_{ij} > t \quad (2)$$

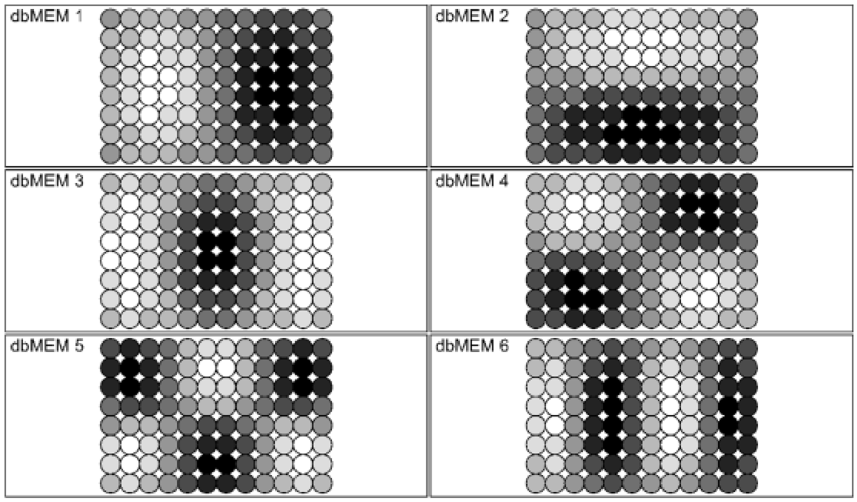
   The value of threshold, t, is chosen. This parameter derives from the observation that MEM eigenvectors display variation across the full set of sites under study if the sites form a connected graph in the truncated matrix. Furthermore, a large truncation value implies a loss of the finest spatial structures. The most commonly applied solution calls for computing the minimum spanning tree of a single-linkage clustering of the site coordinates and then retaining the largest edge value. Hence, threshold t is computed as follows:

   - A minimum spanning tree (MST) linking all points (i.e. site barycenters) in the study is created.
   - The length of the largest edge in the chain forming the MST is determined.
   - t is set equal to the length of the largest edge in the MST.

   Disconnected pairs are identified in the matrix D truncated by distances equal to $4t$. The value $4t$ is chosen here because computer simulations have shown little change to numerical results when using larger values (Borcard & Legendre 2002).
3. The diagonal values of the distance matrix, which were originally zeros, are replaced by the value $4t$; this change on the diagonal of the truncated matrix indicates that a site is not connected to itself.
4. The principal coordinate analysis (PCoA) of the truncated matrix is computed, thereby yielding $(n - 1)$ non-zero eigenvalues and their corresponding eigenvectors (Gower 1966).

3.2 dbMEM spatial vector interpretation

Eigenvectors produced by PCoA decomposition are readily available tools for space partitioning. The components of each eigenvector, associated with corresponding geographic units, form an unique spatial pattern; this can apply to any spatial unit design, but spatial vectors are easier to interpret in the case of regular designs. As an illustration, Figure 1, extracted from Legendre & Legendre (2012), shows maps of the first six eigenvectors produced after performing a dbMEM principal coordinate decomposition for a regular 12 x 8 grid. Shades of gray represent the values in each eigenvector, from white (largest negative value) to black (largest positive value). Spatial patterns, corresponding to the eigenvectors, can indeed be clearly identified.



Note: shades of gray represent the values in each eigenvector, from white (largest negative value) to black (largest positive value).

Figure 1: Maps of the first six eigenvectors (extracted from Legendre & Legendre (2012)) produced after performing a dbMEM principal coordinate decomposition for a regular 12 x 8 grid

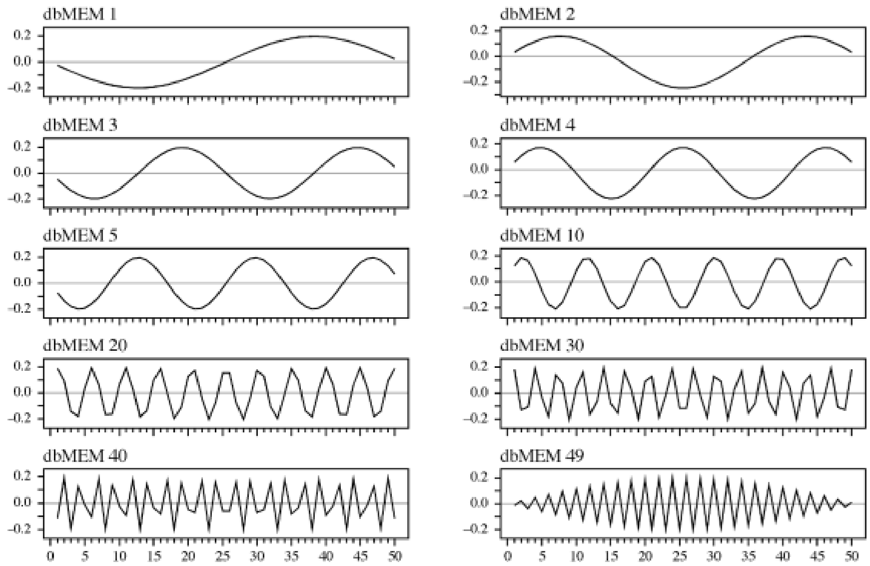
3.2.1 Scale representation

dbMEM variables model the spatial relationships among sites in decreasing order of spatial scale (Borcard & Legendre 2002). The first dbMEM eigenvector (with the largest eigenvalue) corresponds to the broadest spatial scale, indicating the spatial extent of the entire study area (large-scale variation), while the last dbMEM eigenvector (with the smallest eigenvalue) corresponds to the finest spatial scale (small-scale variation).

To illustrate the spatial scales associated with dbMEM variables, let us consider a one-dimensional transect with n equally-spaced sites. Since the design is regular, the dbMEM eigenvectors representing the spatial variation resemble sine waves. Figure 2, extracted from Legendre & Legendre (2012), shows the results for a 50-point transect. The complete sine wave of the first eigenvector has a wavelength of 51 units. The following eigenvectors form sine waves of shorter wavelengths, with the last eigenvector having a wavelength of 2.04. With irregularly-spaced designs (whether one-dimensional or two-dimensional), the dbMEM vectors lose the regularity of their shapes (at times complicating the scale assessment), yet the eigenvectors modeling broader-scaled and finer-scaled phenomena can still be distinguished.

3.2.2 Spatial autocorrelation and dbMEM spatial vectors

Spatial autocorrelation is a measure of similarity between the value of a variable at one location and the value(s) of the same variable at one or more proximal locations.



Note: Abscissa, from left to right: sites 1 to 50. Ordinates: values along the dbMEM eigenfunctions.

Figure 2: Graphs (extracted from Legendre & Legendre (2012)) of 10 of the 49 dbMEM eigenfunctions representing the spatial variation along a transect with 50 equally-spaced points

A widely used tool for measuring spatial autocorrelation is Moran's coefficient (Moran 1948). In matrix form, Moran's coefficient can be formulated as follows:

$$MC = \frac{n}{\sum_i \sum_j w_{ij}} \frac{x'Wx}{x'x} \tag{3}$$

where x is a vector $(n \times 1)$ of mean-centered values of a geo-referenced variable, and W a spatial weights matrix of dimensions $(n \times n)$ with elements w_{ij}. The elements of the spatial weights matrix assume non-zero values if locations i and j are deemed to be spatially proximate in some sense, and 0 otherwise.

Dray et al. (2006) showed that the eigenvalues of the dbMEM spatial eigenvectors are equal to Moran's I coefficients of spatial correlation (Eq. 3) computed for these same eigenvectors, divided by a constant. A positive eigenvalue therefore corresponds to a positive spatial autocorrelation, i.e. the Moran index is positive. On the contrary, a negative eigenvalue corresponds to a negative spatial autocorrelation at short range. In the case of a linear transect with equally spaced points (example shown in Figure 2), roughly half the eigenvectors have a positive Moran's I and model a positive spatial correlation, while the other half has a negative Moran's I and models a negative spatial correlation at short range.

This paper only considers eigenvectors with positive eigenvalues and Moran's I. The focus in fact lies on the positive spatial autocorrelation of car ownership in the spatial analysis. Most previous studies with dbMEM have considered positive spatial autocorrelation (Legendre & Legendre 2012). However, eigenvectors with negative Moran I remain available when investigating negative spatial autocorrelation is needed.

3.3 Multi-scale spatial analysis

3.3.1 Significant spatial vectors

The number of dbMEM eigenvectors created by the principal coordinate analysis tends to be high. However, not all these vectors are significant and contribute to an explanation of the dependent variable (e.g. car ownership). Therefore, an appropriate selection method is needed to reduce their number (Borcard et al. 2004). Since dbMEM eigenvectors are orthogonal by design so as not to correlate with each other, they can be used in linear models as explanatory variables of the dependent variable. The model can thus be made more parsimonious, by means of retaining just the significant dbMEM eigenvectors, using an appropriate selection method (e.g. stepwise eigenvectors selection). The dbMEM eigenvectors can be used as explanatory variables in any model of general and generalized linear modeling theory (Griffith & Peres-Neto 2006).

3.3.2 Spatial scales

Given that dbMEM eigenvectors are orthogonal to one another, they can be combined into sub-models, corresponding to different spatial scales. Furthermore, any sub-model containing a subset of dbMEM is also independent of any other sub-model containing another subset. The spatial scales corresponding to these sub-models vary from large to very fine. For example, if dbMEM eigenvectors are grouped into three sub-models, then the sub-model representing the largest spatial scale is to be characterized by eigenvectors with the largest positive eigenvalues. Similarly, the sub-model representing the medium spatial scale is to be characterized by eigenvectors with intermediate positive eigenvalues and the fine scale by eigenvectors with small eigenvalues.

These sub-models are defined arbitrarily, yet a few procedures can still be followed for this selection step (Borcard et al. 2018):

- predefine spatial scale limits, using the sizes of patterns corresponding to the dbMEM variables;
- identify groups of eigenvectors by examining a scalogram showing in the ordinate the regression coefficients (or the Pearson correlations or the absolute values of the *t*-statistics associated with the regression coefficients) of the dbMEM eigenvectors ordered along the abscissa in decreasing eigenvalues;
- draw maps of the significant dbMEM variables and group them visually according to the scales of the patterns they represent.

3.3.3 Variation of dependent data

Another step of multi-scale spatial analysis consists in identifying factors that explain the variation of dependent data:

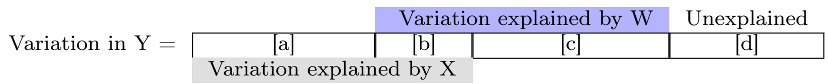


Figure 3: Diagram describing the partitioning of the variation of a response variable Y by two sets of explanatory variables (X and W)
.

- variation of dependent data (e.g. car ownership) at each scale can be attributed to the sociodemographic variables available for the analysis (e.g. sociodemographic variables identified in Section 2);
- variation of dependent data can be represented simultaneously with respect to spatial scales and sociodemographic variables.

Variation partitioning analysis (Mood 1971, Borcard et al. 1992, Peres-Neto & Jackson 2001) constitutes a comprehensive approach to represent the variation of dependent data. It consists of apportioning the variation of a variable, between two or more explanatory datasets pertaining to the different classes. Variation partitioning is commonly used in the field of ecology as an exploratory tool and to give insights into processes structuring communities (Gebrehiwot et al. 2020, Checon & Amaral 2017, Cao et al. 2019, Zbinden & Matthews 2017).

In a partitioning by two explanatory matrices (sets) X and W, both explain some variation of the response data. When the explanatory datasets are not orthogonal to one another, some amount of variation is explained jointly by the two sets. Consequently, the variation explained by all variables taken together is less than the sum of the variations explained by the various subsets. Figure 3 illustrates the partitioning of the explanatory power of different explanatory matrices (X and W) in relation to the same response matrix (Y).

Each explanatory matrix uniquely explains a portion of the variation in Y (X (resp. W) explains partition [a] (resp. [c])). In addition, both matrices explain another portion of the variation in Y (partition [b]), i.e. the explanatory matrices are redundant in this partition. The larger this fraction, the more multicollinearity present in the model. The residual variation in matrix Y (partition [d]) is explained by neither X nor W.

4 Study of car ownership in a French administrative region

This section will demonstrate the use of dbMEM multi-scale spatial analysis on car ownership rate data. The procedure presented in Section 3 is applied to the Loire-Atlantique Department (France).[3] The dataset will be described before conducting a dbMEM analysis of the dependent variable (i.e. car ownership), then dividing the significant eigenfunctions into sub-models and interpreting these sub-models using explanatory variables. As a final step, dbMEM eigenfunctions will be applied within the framework of variation partitioning.

[3]In France's administrative divisions, the department is one of the three administrative jurisdictions of government below the national level, positioned between regions and municipalities.

4.1 Data considerations

For this analysis, we are using data of the Loire-Atlantique Department from the 2016 census and the FiLoSoFi (localised disposable income system) database. These datasets have been collected respectively by the French National Institute of Statistics and Economic Studies (INSEE) and the French fiscal administration. The Loire-Atlantique department is located in northwestern France and has a population of roughly 1,400,000. The main city Nantes is France's 6^{th} largest (with a population of about 310,000).

The statistical areas used in this study are known as IRIS units (acronym for "aggregated units for statistical information"); they represent the basic units for disseminating sub-municipal data in France. All municipalities with over 10,000 inhabitants and a large proportion of those with populations between 5,000 and 10,000 are divided into several IRIS units. The municipalities not divided into multiple IRIS units constitute an IRIS unit in themselves. Within a residential IRIS, the population lies between 1,500 and 5,000. Both census and FiLoSoFi datasets are associated with IRIS units. Variables in the census database include household characteristics (e.g. family composition and car ownership) and individual ones (e.g. profession, age and gender). In Appendix B, Table B.1 provides a detailed description of census variables. Household median disposable annual income variable is given in the FiLoSoFi database.

Car ownership rate is one of the variables contained in the census database, meaning that the car ownership rate can be computed for each IRIS. The mean value of this variable for the study area is 87.41 percent, with a standard deviation of 10.92 percent and a median value of 91.79 percent. Households are more likely to own a car if they live in the Nantes outskirts. Figure 4 shows the proportion of households in each IRIS possessing at least one car.

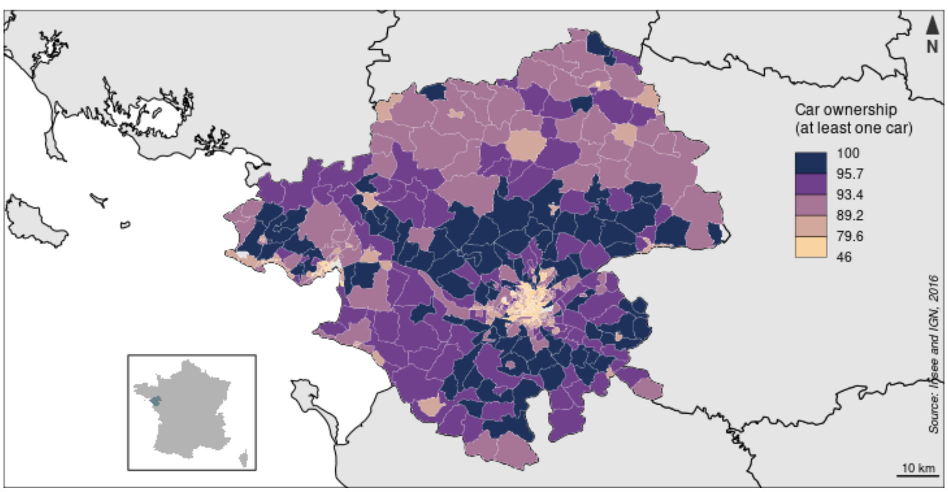


Figure 4: Car ownership, by IRIS

Based on the literature review (Section 2) that identified the main determinants of car ownership, we selected an initial set of explanatory variables from those available in the census and FiLoSoFi databases. This set of variables includes:

- family composition variables: shares of family categories (single member, couple without children, couple with children, single-parent and other composition);
- age variables: shares of different age groups (0-2, 3-5, 6-10, 11-14, 15-17, 18-24, 25-29, 30-39, 40-54, 55-64, 65-79 and 80/+);
- profession variables: shares of farmers, tradespeople, executives, intermediate occupations, clerical support workers, lower-skilled technical occupations, retiree and unemployed;
- income variable: median disposable annual income.

However, many of these variables are highly correlated (e.g. income and profession). In order to avoid co-linearity in our model, only a subset of these variables is retained. This subset is constituted of the following variables:

- share of couples with children variable to take into account the effect of larger households;
- share of individuals in the 25-54 age group variable as individuals in this age group tend to be more active;
- median disposable annual income per consumption unit (consumption units of a household are defined as follows (*Eurostat* 2021): 1 consumption unit (CU) for the first adult in the household, 0.5 CU for every other person in the household aged 14 years or older, 0.3 CU for each child under 14).

In addition to the three selected variables, we also computed a fourth variable, i.e. population density (in inhabitants/km^2), to integrate it into our model. This variable has been used in previous studies of car ownership as a proxy to accessibility variables (Clark 2007). It could be hypothesised that car ownership is likely to be higher in areas of lesser density because there is a poorer provision of public transport and people need to travel further to access shops and services. Therefore, population density of the IRIS is selected as one to help capture this accessibility effect. Table 2 lists the descriptive statistics of this set of explanatory variables.

Variable	Mean	Median	s.d.
car ownership (%)	87.41	91.79	10.92
couples with children (%)	28.41	29.21	11.17
25-54 yrs (%)	38.96	39.93	6.01
population density (inhabitants/km^2)	2362.89	591.80	3399.96
median income (€)	21,681	21,512	3,118

Table 2: Descriptive statistics (IRIS level)

4.2 Results

4.2.1 Spatial vectors

This analysis begins by defining the IRIS spatial neighborhood by means of a minimum spanning tree (see Section 3.1), which can be represented by an unweighted graph, as seen in Figure 5. For our study area, the longest edge of this tree has a length (t) equal to 7,604.5 meters. Beyond this truncation distance, IRIS pairs are deemed to be disconnected. The distance between disconnected IRIS pairs has been set at $4t$ in the distance matrix.

In order to compute spatial vectors, a principal coordinate analysis of the truncated distance matrix between the 499 IRIS has been carried out. This procedure resulted in 498 eigenvectors, 176 of which are associated with positive eigenvalues and the remaining 322 with negative eigenvalues. As explained in Section 3.2.2, only those eigenvectors associated with a positive eigenvalue have been taken into account in our analysis.

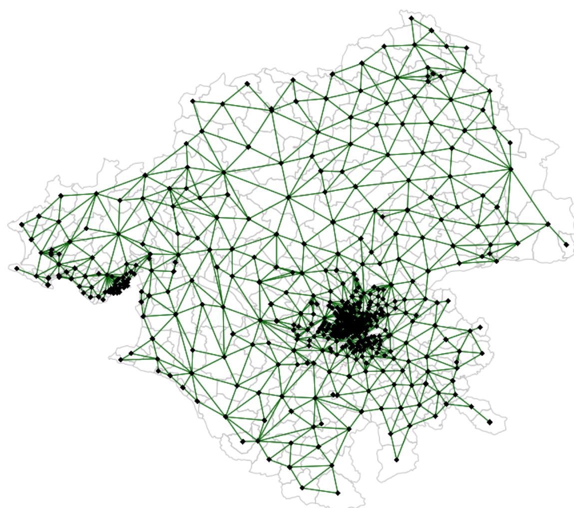


Figure 5: Spatial neighborhood

4.2.2 Significant spatial vectors

As the *car ownership* variable is expressed as a percentage, the standard techniques of statistical analysis (i.e. linear regression and ANOVA) are not appropriate. Thus, we use a generalized linear model (GLM) regression (with a quasi-binomial distribution to account for overdispersion and using a logit link function) in order to estimate the explanatory power of the 176 dbMEM vectors associated with a positive Moran's I value. Among the 176 vectors, only 23 significant ones are retained. The reduced model is presented in Table 3.

	Estimate	Std. Error	t-value
Intercept	2.168	0.029	72.676
MEM_1	-0.572	0.024	-23.575
MEM_4	-0.409	0.023	-17.077
MEM_6	0.131	0.028	4.606
MEM_8	-0.108	0.031	-3.406
MEM_9	0.077	0.026	2.897
MEM_{13}	-0.062	0.024	-2.615
MEM_{14}	-0.090	0.022	-3.959
MEM_{15}	-0.082	0.027	-3.035
MEM_{16}	0.074	0.027	2.678
MEM_{17}	0.087	0.031	2.797
MEM_{20}	0.069	0.022	3.027
MEM_{34}	-0.054	0.028	-1.902
MEM_{36}	-0.076	0.029	-2.577
MEM_{37}	-0.083	0.032	-2.597
MEM_{39}	0.094	0.031	2.964
MEM_{61}	-0.075	0.028	3.383
MEM_{69}	-0.075	0.025	-2.906
MEM_{73}	-0.062	0.029	-2.135
MEM_{80}	-0.050	0.024	-2.052
MEM_{96}	-0.057	0.027	-2.063
MEM_{129}	0.059	0.025	2.297
MEM_{137}	0.048	0.024	1.978
MEM_{167}	-0.060	0.027	-2.191
N			499
ϕ			0.033
Moran I stat. (p-value$<$0.001)			0.090

Table 3: GLM regression: significant eigenfunctions

4.2.3 Spatial sub-models

Significant dbMEM variables have been grouped into subsets corresponding to distinct spatial scales. They were partitioned into three distinct subsets corresponding to three scales by inspecting their geographical representation which are displayed in Appendix A:

- the first two significant dbMEM variables (MEM_1, MEM_4) have been grouped to represent the broad scale (BS);
- the next nine dbMEM variables (MEM 6, 8, 9, 13, 14, 15, 16, 17, 20) have been grouped to represent the medium scale (MS);
- and the last twelve dbMEM variables (MEM 34, 36, 37, 39, 61, 69, 73, 80, 96, 129, 137, 167) have been grouped to represent the fine scale (FS).

We define broad, medium and fine scale sub-models as mathematical operations for computing the variation of car ownership at the corresponding scales. In order to compute the variation of car ownership at each scale, the following steps are applied:

1. the sum of the chosen dbMEM vectors at each scale multiplied by their corresponding estimates (displayed in Table 3) added to the intercept is computed;
2. the variation of the car ownership at each scale is obtained by applying the inverse of the link function (used for GLM regression) to the result of the previous operation.

The values of car ownership variation at broad, medium and fine scale are shown respectively in Figures 6, 7 and 8. Spatial structure associated with broad scale sub-model matches the structure obtained distinguishing between main cities (Nantes and Saint-Nazaire) versus other areas (Figure 6). Spatial structure associated medium scale sub-model matches the structure obtained distinguishing between main city centers and close surrounding areas (Figure 7). Fine scale spatial structure is harder to interpret (Figure 8).

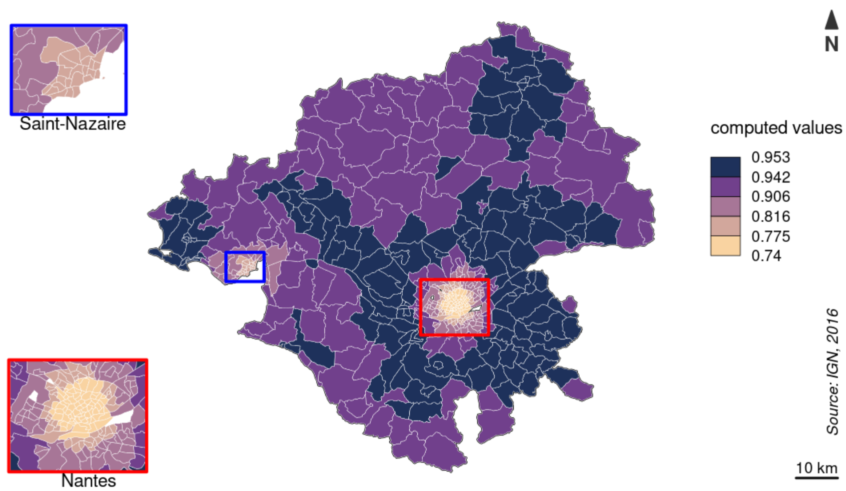


Figure 6: Variation in car ownership computed with the broad scale sub-model

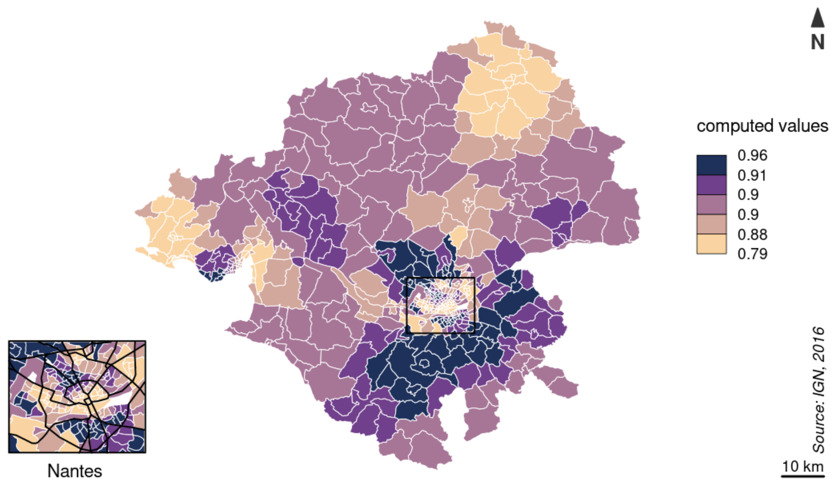


Figure 7: Variation in car ownership computed with the medium scale sub-model

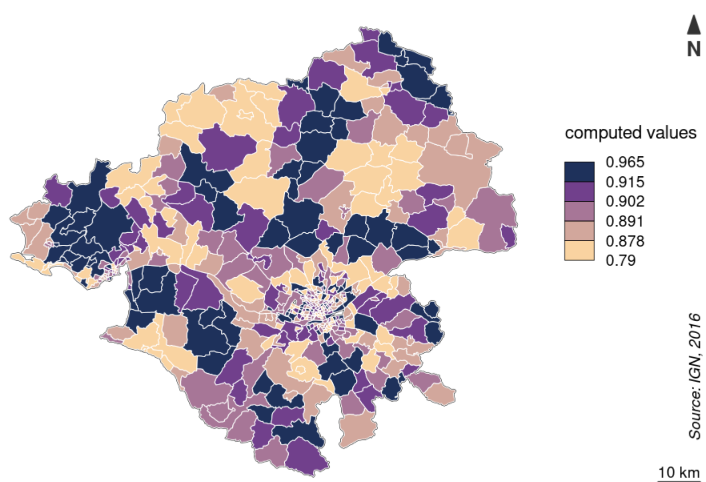


Figure 8: Variation in car ownership computed with the fine scale sub-model

4.2.4 Multi-scale analysis of explanatory variables and variation partitioning

Let us now turn to identifying the sociodemographic characteristics that contribute to the variation in car ownership rate at each identified scale. For this purpose, the following sociodemographic variables have been considered: share of couples with children, share of individuals in the 25-54 year age range, share of households living in single-family dwellings, population density, and income. These sociodemographic variables are used as explanatory variables in regression analysis of the output of each of the three sub-models. Outputs of broad scale, medium scale and fine scale sub-models are respectively regressed on these variables. The results of these regressions establish the contribution of sociodemographic variables at each scale.

The results of multi-scale analysis of explanatory variables are displayed in Table 4. The table reports the signs of the significant coefficients (with a p-value < 0.1) associated with each variable of the following GLM regressions: (i) *car ownership* on the sociodemographic variables (column 1 of the table); (ii) variations of car ownership at each scale (i.e. broad, medium, fine) as displayed by Figures 6, 7 and 8 on the sociodemographic variables (columns 2, 3 and 4).

	(1) *car ownership*	(2) Broad scale	(3) Medium scale	(4) Fine scale
Median income	+	-	+	+
Couples with children	+	+	+	+
25-54 yrs	-		+	
Population density	-	-		

Table 4: Signs of significant coefficients related to the spatial analysis of the *car ownership* variable

Some important elements should be highlighted as regards regression analysis displayed in Table 4.

Considering the regression of the car ownership on the sociodemographic variables (Table 4, column 1), the signs of the coefficients of the *median income* and the *population density* are as expected. In line with previous literature (Clark 2007, Clark & Finley 2010), the *car ownership* rate is positively correlated with the *median income*, in contrast, it is negatively correlated with the *population density*. Furthermore, the coefficient is positive for the share of *couples with children* and negative for the share of individuals between *25-54 years*.

Considering the three other regressions (Table 4, columns 2, 3 and 4), the multi-scale analysis provides a more nuanced view of the effect of the different sociodemographic explanatory variables. The variables *population density*, *median income* and *couples with children* are significant at the broad scale: *population density* and *median income* are negatively correlated with the response variable; while *couples with children* variable is positively correlated. It is interesting to note that *median income* is negatively correlated with car ownership variation at broad scale (Table 4, column 2) although it is positively correlated to car ownership in the purely sociodemographic model (Table 4, column 1). This derives from the fact that spatial structure at broad scale is closely related to the structure obtained distinguishing between main cities (Nantes and Saint-Nazaire) versus other areas as shown in Figure 6. Furthermore, centers of the two main cities in Loire-Atlantique Department tend to concentrate higher income households (i.e. *median income* in town centers is higher than other areas). The combination of these two factors yields a negative correlation between *median income* and car ownership variation at broad scale. At the medium and fine scales, *couples with children* and *median income* are significant and positively correlated with the response variable. The share of individuals in the 25-54 year age range is only significant at the medium scale (Table 4, column 3). As the medium scale distinguishes between main city centers and close surrounding areas (Figure 7), this implies a variation in the share of individuals in the 25-54 year age range between these two areas. Additional data are needed to further explore this result. Finally, one can note that the population density is significant at broad scale (Table 4, column 2) but not significant at medium and fine scales (Table 4, columns 3 and 4).

These results suggest that car ownership variation at each scale and sociodemographic variables are indeed correlated; hence, it can be inferred that at least a portion of the spatial variation in car ownership rate (i.e. structure) is induced by the relevant driving (i.e. sociodemographic) factors. In order to apportion car ownership variation between sub-models and sociodemographic variables, a variation partitioning analysis is undertaken. This analysis estimates both the amount of variation that can be attributed exclusively to sub-models (spatial variables) and the amount explained jointly by spatial and sociodemographic variables.

Figure 9 displays the results of this variation partitioning analysis following the method proposed by Randin et al. (2009). The explanatory matrices defined contain variables pertaining to sociodemographic determinants and various spatial scales. The first model (i.e. full) considers all significant dbMEM variables while the other three models (i.e. broad, medium and fine) consider only the subset of dbMEM variables associated with each scale. We partition the variation into four identifiable fractions of deviance: (i) pure sociodemographic (black area), (ii) pure spatial (grey area), (iii) shared sociodemographic and spatial (mustard area), and (iv) unexplained variation (yellow area).

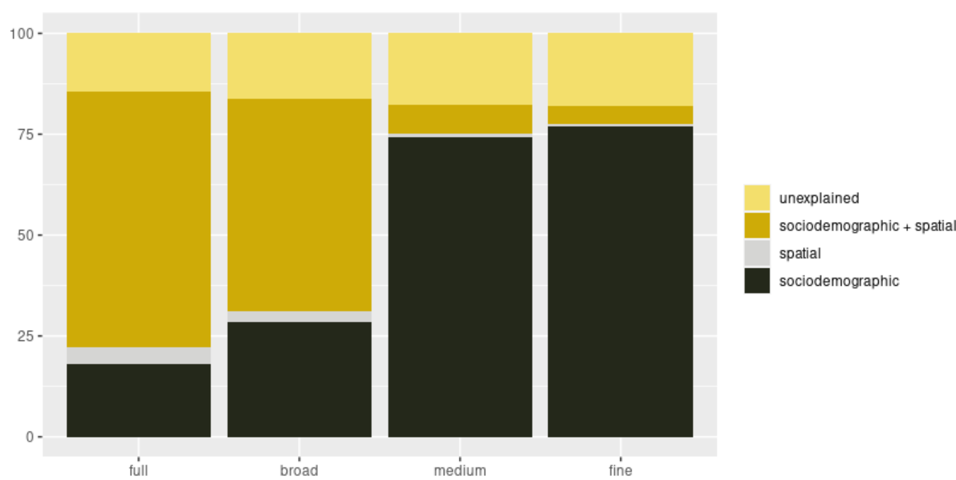


Figure 9: Variation partitioning

Some important elements should be highlighted as regards variation partitioning analysis. They are summarized as follows:

- The car ownership rate variation is well explained by the model: the proportion of variation explained by the model ranges from 85 % for the full model to 81 % for the model including the fine scale as spatial model.
- As expected, the sociodemographic model explains most of the variation: the sum of the deviance explained by the pure sociodemographic fraction and the shared sociodemographic and spatial fraction represents at least 81 % of the total variation.
- The car ownership rate variation explained by sociodemographic variables is highly spatialized: 63 % of the deviance in the full model is simultaneously explained by sociodemographic and spatial variables. The shared deviance between the socioeconomic and the spatial variables remains important for the broad scale model (i.e. 53 %) but diminishes dramatically for the medium and fine scale models (i.e. 7 % and 4 %, respectively).
- A small part of the car ownership rate variation is exclusively explained by the spatial structure: in the full model, this part is 4 %. This additional explanation by the spatial model is mainly due to the broad scale model: the independent contribution of the spatial variables is very small for the medium and fine scale models.

5 Discussion

The results presented in this paper underscore the importance of multi-scale spatial analysis as an explanatory tool applicable to transportation data. Since the data are spatially structured, these structures can be correlated with both explanatory factors and governance policies. Let us note that space *per se* is not considered as an explanation of transportation data variability; rather, spatial variables serve as proxies to quantify and dissect the spatial variation. A portion of this variation can then be attributed to some of the explanatory variables available for analysis, with the remainder being considered as spatial variation yet to be explained.

In this study, the car ownership rate has been shown to have a spatial structure at three distinct scales. At the broad scale, the spatial structure of car ownership rate variation is in line with the structure obtained when distinguishing between the two main cities (Nantes and Saint-Nazaire) versus other areas (Figs. 6 and A.1), which implies that living in one of the study area's two main cities is a relevant predictor of car ownership, as the rate is less in these cities. The broad-scale spatial structure of car ownership rate has been correlated with classical sociodemographic explanatory variables: negatively with population density and median income, and positively with share of couples with children. Population density is closely correlated with the availability of public transit as corresponding investments in denser areas are far greater. Part of the explanation of the negative correlation between median income and car ownership at broad scale model may lie in the fact that a part of the population with higher income lives in areas with better urban amenities which can offer a good substitute for car ownership (Mulalic & Rouwendal 2020). This assumption needs to be explored further by social scientists. In future research, it would be interesting to investigate if this finding is specific to the studied area or valid in different areas. This analysis illustrates the additional value to classical sociodemographic modeling brought by taking into Mems in analyzing the car ownership.

At the medium scale, the spatial structure of car ownership rate variation is in line with the structure obtained when distinguishing between large city centers and the nearby surrounding areas (Figs. 7 and A.2). In the case of Nantes city, three areas can be distinguished. The first is the inner core with a lower rate. The second area is constituted of IRIS located between the inner core and the ring road, where median income and individual house share are higher. These areas tend to have higher ownership rates. The third area, located outside the ring road and characterized by lower median income and a higher dwelling share, tends to have lower ownership. At the fine scale, spatial structures are harder to interpret (Figs. 8 and A.3).

A major advantage of distance-based Moran's eigenvector analysis is its explicit identification of scale patterns, which provides critical information in land use and transportation planning (Wang et al. 2013). Knowing these patterns (i.e. spatial structures) can be of major interest for transportation and urban planners and moreover allow decision-makers to act on targeted zones of the territory to influence the variable of interest. As shown by our analysis, issues are not the same throughout the territory considered. For example, since the car ownership rate is particularly high outside city centers, policymakers can increase the supply of public transit in these zones. Similarly, the analysis has identified a higher car ownership rate within the IRIS located in Nantes between the core and the ring road. A more targeted analysis of this area would serve to determine whether any levers could be used to reduce the rate. Another advantage of this analysis is its use as an explanatory tool to refine models for the dependent variable. When model residuals are not randomly distributed, exploring the spatial patterns of these residuals can in fact be very helpful in identifying missing explanatory factors. Supported by expert opinion, the resulting patterns will then guide the choice of omitted relevant explanatory variables (Paez 2019).

dbMEM vectors are obtained by performing a principal coordinate analysis of a truncated matrix of Euclidean distances between centroids of defined statistical areas. As our concern is transportation data analysis, a more sophisticated approach would use other distance measurements between statistical areas (instead of simple Euclidean distances) such as travel time.[4] In fact, travel time provides another relevant transportation-related measure of scale. In this paper, we have kept the classic Euclidean distance as a first approach and in order to facilitate the analysis. However, it will be the purpose of future works to analyze the data using others distances (e.g. travel time).

6 Conclusion

This study has contributed to the current transportation literature by introducing multi-scale spatial analysis techniques with space as an explicit predictor. Applying these techniques to study transportation variables (e.g. household car ownership) offers many important methodological and operational advantages. They provide transportation planners with a tool for identifying the spatial structures of response variables at various scales. These structures can be correlated with explanatory variables and used to target specific geographical areas in transportation planning policies.

In this paper, distance-based Moran's eigenvector maps have been implemented for car ownership spatial analysis. For this purpose, aggregated census data for small

[4] We thank an anonymous reviewer of this paper for suggesting travel time as another relevant transportation-related measure of scale.

statistical areas in France's Loire-Atlantique Department were used. Significant spatial vectors explaining car ownership variable were selected by performing generalized linear model regression. These spatial vectors were then divided into sub-models representing broad scale, medium scale and fine scale. Finally, variation partitioning analysis was performed with respect to variables pertaining to both sociodemographic determinants and the distinct spatial scales.

In conclusion, distance-based Moran's eigenvectors maps is an appealing method that can be used to study transport data. In this paper, it has been applied to study spatial structure of a single response variable. An avenue for future research is to apply this methodology in different urban areas to check whether the same findings still apply. For example, some counterintuitive results, such as the negative correlation between income and car ownership at broad scale, underline the interest of this approach, but this result needs to be verified in other urban areas. Furthermore, as this first analysis was based on Euclidean distance between statistical areas to explore spatial structures, it could be enriched by considering other distance measurements for transportation data analysis (e.g. travel time). Moreover, opportunities for continued investigation include applying this method to study multivariate response data. In this approach, spatial layout of multiple transport variables (e.g. single car, multiple cars and motorized two-wheelers ownership) can be explored simultaneously. Finally, another investigation opportunity is the study of both spatial and temporal layout of the response variable. In fact, dbMEM variables can represent effectively a spectral decomposition of both the spatial and temporal relationships.

Acknowledgments
Spatial analyses were performed using the packages 'adespatial' and 'spdep' developed in R language. For the variation partitioning analysis, the R package 'ecospat' was run.

References

Adjemian, M. K., Lin, C.-Y. C. & Williams, J. (2010), 'Estimating spatial interdependence in automobile type choice with survey data', *Transportation Research Part A: Policy and Practice* **44**(9), 661–675.

Anastasopoulos, P. C., Karlaftis, M. G., Haddock, J. E. & Mannering, F. L. (2012), 'Household automobile and motorcycle ownership analyzed with random parameters bivariate ordered probit model', *Transportation research record* **2279**(1), 12–20.

Anowar, S., Eluru, N. & Miranda-Moreno, L. F. (2014), 'Alternative modeling approaches used for examining automobile ownership: a comprehensive review', *Transport Reviews* **34**(4), 441–473.

AURAN (2006), 'Aire urbaine de nantes : Mémento des modes de déplacements'.

Bhat, C. R. & Pulugurta, V. (1998), 'A comparison of two alternative behavioral choice mechanisms for household auto ownership decisions', *Transportation Research Part B: Methodological* **32**(1), 61–75.

Boots, B. & Tiefelsdorf, M. (2000), 'Global and local spatial autocorrelation in bounded regular tessellations', *Journal of Geographical Systems* **2**(4), 319–348.

Borcard, D., Gillet, F. & Legendre, P. (2018), *Numerical ecology with R*, Springer.

Borcard, D. & Legendre, P. (2002), 'All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices', *Ecological modelling* **153**(1-2), 51–68.

Borcard, D., Legendre, P., Avois-Jacquet, C. & Tuomisto, H. (2004), 'Dissecting the spatial structure of ecological data at multiple scales', *Ecology* **85**(7), 1826–1832.

Borcard, D., Legendre, P. & Drapeau, P. (1992), 'Partialling out the spatial component of ecological variation', *Ecology* **73**(3), 1045–1055.

Brice, M.-H., Pellerin, S. & Poulin, M. (2016), 'Environmental filtering and spatial processes in urban riparian forests', *Journal of Vegetation Science* **27**(5), 1023–1035.

Brind'Amour, A., Boisclair, D., Legendre, P. & Borcard, D. (2005), 'Multiscale spatial distribution of a littoral fish community in relation to environmental variables', *Limnology and Oceanography* **50**(2), 465–479.

Brind'Amour, A., Mahévas, S., Legendre, P. & Bellanger, L. (2018), 'Application of moran eigenvector maps (mem) to irregular sampling designs', *Spatial Statistics* **26**, 56–68.

Brunsdon, C., Fotheringham, S. & Charlton, M. (1998), 'Geographically weighted regression', *Journal of the Royal Statistical Society: Series D (The Statistician)* **47**(3), 431–443.

Cao, J. & Cao, X. (2014), 'The impacts of lrt, neighbourhood characteristics, and self-selection on auto ownership: Evidence from minneapolis-st. paul', *Urban Studies* **51**(10), 2068–2087.

Cao, X. J., Næss, P. & Wolday, F. (2019), 'Examining the effects of the built environment on auto ownership in two norwegian urban regions', *Transportation Research Part D: Transport and Environment* **67**, 464–474.

Checon, H. H. & Amaral, A. C. (2017), 'Taxonomic sufficiency and the influence of rare species on variation partitioning analysis of a polychaete community', *Marine Ecology* **38**(1), e12384.

Chen, W., Jiao, S., Li, Q. & Du, N. (2020), 'Dispersal limitation relative to environmental filtering governs the vertical small-scale assembly of soil microbiomes during restoration', *Journal of Applied Ecology* **57**(2), 402–412.

Cilleros, K., Allard, L., Vigouroux, R. & Brosse, S. (2017), 'Disentangling spatial and environmental determinants of fish species richness and assemblage structure in neotropical rainforest streams', *Freshwater Biology* **62**(10), 1707–1720.

Clark, B., Chatterjee, K. & Melia, S. (2016), 'Changes in level of household car ownership: the role of life events and spatial context', *Transportation* **43**(4), 565–599.

Clark, S. D. (2007), 'Estimating local car ownership models', *Journal of transport Geography* **15**(3), 184–197.

Clark, S. D. & Rey, S. (2017), 'Temporal dynamics in local vehicle ownership for great britain', *Journal of Transport Geography* **62**, 30–37.

Clark, S. & Finley, A. O. (2010), 'Spatial modelling of car ownership data: a case study from the united kingdom', *Applied Spatial Analysis and Policy* **3**(1), 45–65.

Ding, C. & Cao, X. (2019), 'How does the built environment at residential and work locations affect car ownership? an application of cross-classified multilevel model', *Journal of Transport Geography* **75**, 37–45.

Ding, C., Wang, Y., Tang, T., Mishra, S. & Liu, C. (2018), 'Joint analysis of the spatial impacts of built environment on car ownership and travel mode choice', *Transportation research part D: transport and environment* **60**, 28–40.

Dray, S., Blanchet, G., Borcard, D., Guenard, G., Jombart, T., Larocque, G., Legendre, P., Madi, N., Wagner, H. H. & Dray, M. S. (2018), 'Package 'adespatial'', *R package version* pp. 3–8.

Dray, S., Legendre, P. & Peres-Neto, P. R. (2006), 'Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (pcnm)', *ecological modelling* **196**(3-4), 483–493.

Eurostat (2021), http://web.archive.org/web/20080207010024/http://www.808multimedia.com/winnt/kernel.htm. Accessed: 2021-04-14.

Gáspár, B., Bossdorf, O. & Durka, W. (2019), 'Structure, stability and ecological significance of natural epigenetic variation: a large-scale survey in plantago lanceolata', *New phytologist* **221**(3), 1585–1596.

Gebrehiwot, M., Kifle, D. & Triest, L. (2020), 'Partitioning the influence of hydrodynamics-induced physical variables and nutrients on phytoplankton assemblages in a shallow tropical reservoir (koka, ethiopia)', *Limnology* pp. 1–6.

Gower, J. C. (1966), 'Some distance properties of latent root and vector methods used in multivariate analysis', *Biometrika* **53**(3-4), 325–338.

Griffith, D. A. (1996), 'Spatial autocorrelation and eigenfunctions of the geographic weights matrix accompanying geo-referenced data', *Canadian Geographer/Le Géographe canadien* **40**(4), 351–367.

Griffith, D. A. (2000*a*), 'Eigenfunction properties and approximations of selected incidence matrices employed in spatial analyses', *Linear Algebra and its Applications* **321**(1-3), 95–112.

Griffith, D. A. (2000*b*), 'A linear regression solution to the spatial autocorrelation problem', *Journal of Geographical Systems* **2**(2), 141–156.

Griffith, D. A. (2009), 'Modeling spatial autocorrelation in spatial interaction data: empirical evidence from 2002 germany journey-to-work flows', *Journal of geographical systems* **11**(2), 117–140.

Griffith, D. A. (2011), 'Visualizing analytical spatial autocorrelation components latent in spatial interaction data: an eigenvector spatial filter approach', *Computers, Environment and Urban Systems* **35**(2), 140–149.

Griffith, D. A. & Peres-Neto, P. R. (2006), 'Spatial modeling in ecology: the flexibility of eigenfunction spatial analyses', *Ecology* **87**(10), 2603–2613.

Jiang, Y., Gu, P., Chen, Y., He, D. & Mao, Q. (2017), 'Influence of land use and street characteristics on car ownership and use: Evidence from jinan, china', *Transportation Research Part D: Transport and Environment* **52**, 518–534.

Jin, L. S., Yin, D., Fortin, M.-J. & Cadotte, M. W. (2020), 'The mechanisms generating community phylogenetic patterns change with spatial scale', *Oecologia* **193**, 655–664.

Laliberté, E., Paquette, A., Legendre, P. & Bouchard, A. (2009), 'Assessing the scale-specific importance of niches and other spatial processes on beta diversity: a case study from a temperate forest', *Oecologia* **159**(2), 377–388.

Lavery, T., Páez, A. & Kanaroglou, P. S. (2013), 'Driving out of choices: An investigation of transport modality in a university sample', *Transportation research part A: policy and practice* **57**, 37–46.

Legendre, P., Borcard, D. & Peres-Neto, P. R. (2005), 'Analyzing beta diversity: partitioning the spatial variation of community composition data', *Ecological monographs* **75**(4), 435–450.

Legendre, P. & Legendre, L. F. (2012), *Numerical ecology*, Elsevier.

Ma, J., Ye, X. & Shi, C. (2018), 'Development of multivariate ordered probit model to understand household vehicle ownership behavior in xiaoshan district of hangzhou, china', *Sustainability* **10**(10), 3660.

Moeckel, R. & Yang, D. (2016), 'Car ownership and the impact of built environment, demographics and transport systems', *Journal of Civil Engineering and Architecture* **10**(5), 587–595.

Moniruzzaman, M. & Páez, A. (2012), 'Accessibility to transit, by transit, and mode share: application of a logistic model with spatial filters', *Journal of Transport Geography* **24**, 198–205.

Mood, A. M. (1971), 'Partitioning variance in multiple regression analyses as a tool for developing learning models', *American Educational Research Journal* **8**(2), 191–202.

Moran, P. A. (1948), 'The interpretation of statistical maps', *Journal of the Royal Statistical Society. Series B (Methodological)* **10**(2), 243–251.

Morton, C., Lovelace, R., Philips, I. & Anable, J. (2018), 'Fuel price differentials and car ownership: A spatial analysis of diesel cars in northern ireland', *Transportation Research Part D: Transport and Environment* **63**, 755–768.

Mulalic, I. & Rouwendal, J. (2020), 'Does improving public transport decrease car ownership? evidence from a residential sorting model for the copenhagen metropolitan area', *Regional Science and Urban Economics* **83**, 103543.

Murakami, D. & Griffith, D. A. (2019), 'Eigenvector spatial filtering for large data sets: fixed and random effects approaches', *Geographical Analysis* **51**(1), 23–49.

Oakil, A. T. M., Manting, D. & Nijland, H. (2016), 'Determinants of car ownership among young households in the netherlands: The role of urbanisation and demographic and economic characteristics', *Journal of transport geography* **51**, 229–235.

Paez, A. (2019), 'Using spatial filters and exploratory data analysis to enhance regression models of spatial data', *Geographical Analysis* **51**(3), 314–338.

Páez, A., López, F. A., Ruiz, M. & Morency, C. (2013), 'Development of an indicator to assess the spatial fit of discrete choice models', *Transportation Research Part B: Methodological* **56**, 217–233.

Paleti, R., Bhat, C. R., Pendyala, R. M. & Goulias, K. G. (2013), 'Modeling of household vehicle type choice accommodating spatial dependence effects', *Transportation research record* **2343**(1), 86–94.

Peres-Neto, P. R. & Jackson, D. A. (2001), 'How well do multivariate data sets match? the advantages of a procrustean superimposition approach over the mantel test', *Oecologia* **129**(2), 169–178.

Pollice, A., Jona-Lasinio, G., Gaglio, M., Blanchet, F. & Fano, E. (2020), 'Modelling the effect of directional spatial ecological processes for a river network in northern italy', *Ecological Indicators* **112**, 106144.

Potoglou, D. & Kanaroglou, P. S. (2008), 'Modelling car ownership in urban areas: a case study of hamilton, canada', *Journal of Transport Geography* **16**(1), 42–54.

Randin, C. F., Jaccard, H., Vittoz, P., Yoccoz, N. G. & Guisan, A. (2009), 'Land use improves spatial predictions of mountain plant abundance but not presence-absence', *Journal of Vegetation Science* **20**(6), 996–1008.

Santos, R. C. L., Callil, C. T. & Landeiro, V. L. (2020), 'Unraveling the effects of water–sediment conditions and spatial patterns on unionida assemblages in seasonally connected floodplain lakes', *Hydrobiologia* **847**, 2909–2922.

Song, J. & Wang, R. (2017), 'Measuring the spatial dimension of automobile ownership and its associations with household characteristics and land use patterns: A case study in three counties, south florida (usa)', *Sustainability* **9**(4), 558.

Taddeo, S., Dronova, I. & Harris, K. (2021), 'Greenness, texture, and spatial relationships predict floristic diversity across wetlands of the conterminous united states', *ISPRS Journal of Photogrammetry and Remote Sensing* **175**, 236–246.

Van Acker, V. & Witlox, F. (2010), 'Car ownership as a mediating variable in car travel behaviour research using a structural equation modelling approach to identify its dual relationship', *Journal of Transport Geography* **18**(1), 65–74.

Wagner, H. H., Chávez-Pesqueira, M. & Forester, B. R. (2017), 'Spatial detection of outlier loci with moran eigenvector maps', *Molecular ecology resources* **17**(6), 1122–1135.

Wang, Y., Kockelman, K. M. & Wang, X. C. (2013), 'Understanding spatial filtering for analysis of land use-transport data', *Journal of Transport Geography* **31**, 123–131.

Whelan, G. (2007), 'Modelling car ownership in great britain', *Transportation Research Part A: Policy and Practice* **41**(3), 205–219.

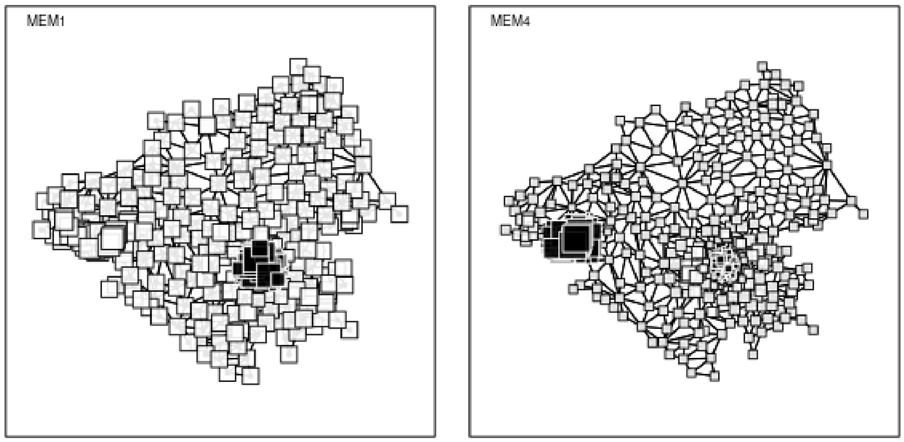
Yoshida, T. & Murakami, D. (2020), Chapter Seven - Implementation with R language, *in* Y. Yamagata & H. Seya, eds, 'Spatial Analysis Using Big Data', Academic Press, pp. 181–223.

Zbinden, Z. D. & Matthews, W. J. (2017), 'Beta diversity of stream fish assemblages: partitioning variation between spatial and environmental factors', *Freshwater biology* **62**(8), 1460–1471.

Zegras, C. (2010), 'The built environment and motor vehicle ownership and use: Evidence from santiago de chile', *Urban Studies* **47**(8), 1793–1817.

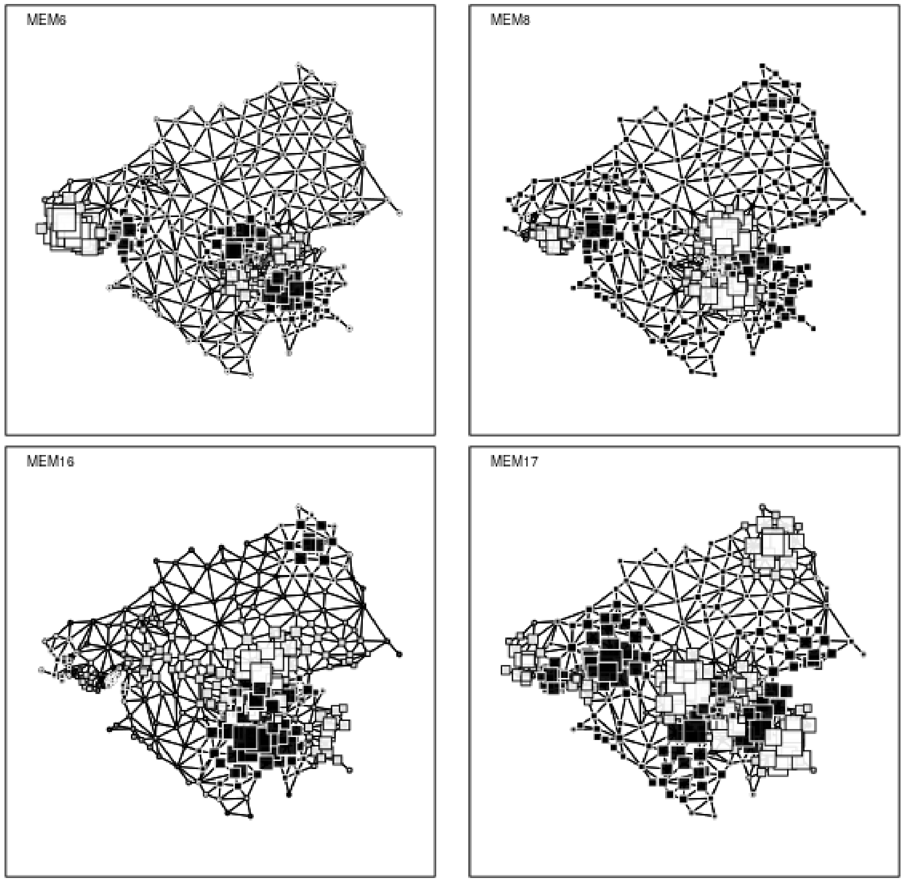
Appendices

A MEM spatial representation



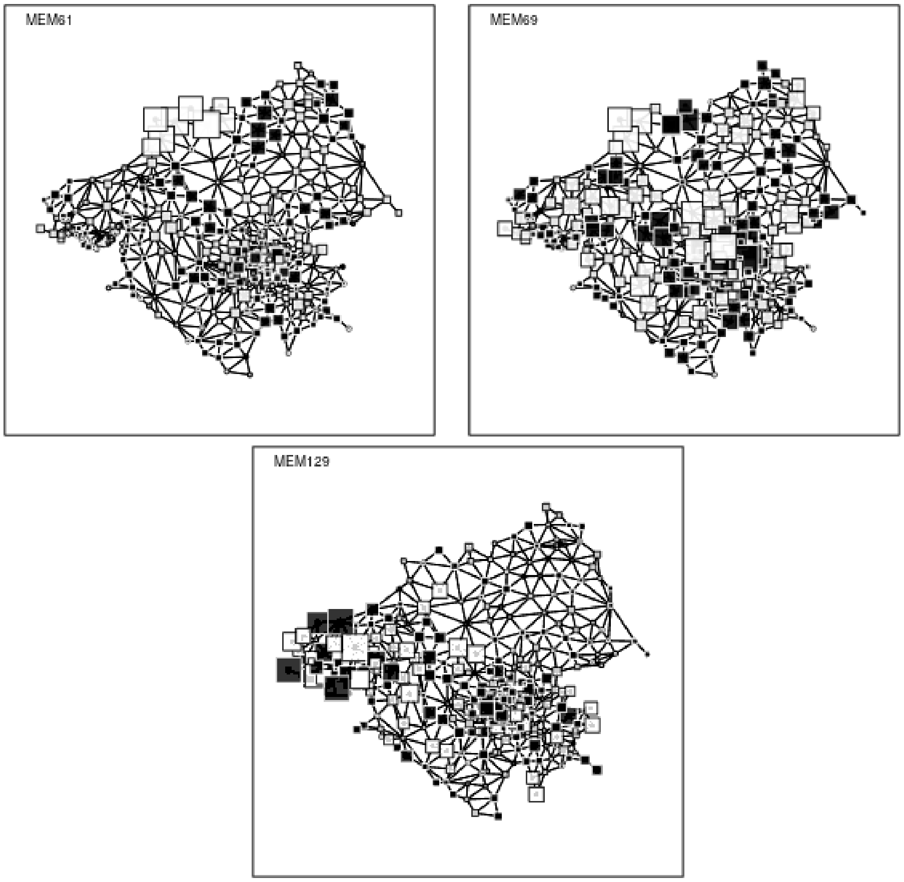
Note: Each dbMEM vector is plotted with the component of the vector associated to a given statistical area (IRIS) represented by a square. Black and White squares represent values of different signs (positive or negative). Square sizes are proportional to the absolute value of the component.

Figure A.1: MEM 1 and 4 form the broad scale sub-model. Spatial structure associated with these MEM matches the structure obtained distinguishing between main cities (Nantes and Saint-Nazaire) versus other areas.



Note: Each dbMEM vector is plotted with the component of the vector associated to a given statistical area (IRIS) represented by a square. Black and White squares represent values of different signs (positive or negative). Square sizes are proportional to the absolute value of the component.

Figure A.2: MEM 6, 8, 16 and 17 are part of the medium scale sub-model. Spatial structure associated with these MEM matches the structure obtained distinguishing between main city centers and close surrounding areas.



Note: Each dbMEM vector is plotted with the component of the vector associated to a given statistical area (IRIS) represented by a square. Black and White squares represent values of different signs (positive or negative). Square sizes are proportional to the absolute value of the component.

Figure A.3: MEM 61, 69 and 129 are part of the fine scale sub-model. Fine scale spatial structures are harder to interpret.

B Variables in the census database

Level	Variable	Definition [number of categories]	Categories
Household	Fam	Family composition [5]	Single member; The nuclear family is a couple without children; The nuclear family is a couple with children; The nuclear family is a single-parent family; Other composition
	ProfRP	Profession of the reference person [7]	Farmers, tradespeople; Executive; Intermediate occupations; Clerical support workers; Lower-skilled technical occupations; Retiree; Unemployed
	Size	Household size [2]	One person; Two persons or more
	Cars	Number of cars [3]	No car; One; Two or more
Individual	Age	Age [12]	0-2; 3-5; 6-10; 11-14; 15-17; 18-24; 25-29; 30-39; 40-54; 55-64; 65-79; 80/+
	Sex	Gender [2]	Female; Male
	Relate	Relationship to the household reference person [2]	Household reference person; Other household member
	Prof	Profession [7]	Farmers, tradespeople; Executive; Intermediate occupations; Clerical support workers; Lower-skilled technical occupations; Retiree; Unemployed
	Wstat	Work status [7]	In fixed-term employment; Permanent employment; Self-employed; Unpaid apprenticeships for those 15 or older; Unemployed; Under 15 years old; Other non-active persons
	Wtime	Working time [3]	Full-time worker; Part-time worker; Not applicable

Table B.1: Census sociodemographic variables