Applied Geography 94 (2018) 241–250

Contents lists available at ScienceDirect

Applied Geography

journal homepage: www.elsevier.com/locate/apgeog

The influence of spatial data allocation procedures on accessibility results: The case of high-speed rail networks

Emilio Ortega[a,b,*], Andrés Monzón[b], Elena López[b]

[a] Department of Forest and Environmental Engineering and Management, MONTES (School of Forest Engineering and Natural Environment), Universidad Politécnica de Madrid, 28040 Madrid, Spain
[b] Transport Research Centre (TRANSyT-UPM), Universidad Politécnica de Madrid, 28040 Madrid, Spain

ARTICLE INFO

Keywords:
Accessibility
High speed rail (HSR)
Spatial data allocation procedures

ABSTRACT

The computation of accessibility indicators requires the analyst to decide how spatial data are to be allocated in the defined areal units. This implies the selection of a geographical representation system and a spatial data distribution procedure to allocate data in each unit. However, there is no consensus on a "best choice" from the set of available spatial data allocation procedures, resulting in a significant heterogeneity across accessibility studies. In addition, so far little research has been done evaluating how accessibility results differ when computed under alternative procedures. Our approach addresses this gap and combines different geographical representation and population distribution procedures to produce different spatial data allocation methods for testing in case studies. In the case accessibility to high speed rail (HSR) networks is computed under five alternative population allocation procedures, in two rural areas in Spain served by a HSR line: Ciudad Real and Cuenca. Our results show that decisions related to spatial data allocation procedures must be carefully addressed when conducting accessibility analyses. Specifically, the population distribution method may in many cases produce differences of over 100% on accessibility values. We conclude that differences between dasymetric and vector procedures tend to be larger in highly populated and sprawling cities, and appear to be less relevant if the size of the cell is small. The results of the paper provide useful guidelines to interpret the influence on the results of different geographical representation and population disaggregation methods. Based on these findings we recommend that the selection of the most appropriate procedure should be explicitly taken into consideration in accessibility analysis methodologies.

1. Introduction

Transport-related accessibility studies place special emphasis on how the transportation network influences the spatial distribution of access to destinations. From this perspective, accessibility is understood as a feature of a given location, related to the ease with which "desired" destinations can be reached from it. These destinations widely vary across accessibility studies and include those as e.g. healthcare facilities (Coffee et al., 2012), supermarkets (Farber, Morang, & Widener, 2014), or HSR stations (Monzón, Ortega, & López, 2016), to cite some recent examples (for a review of existing accessibility formulations see (Geertman & Ritsema Van Eck, 1995; Geurs & van Wee, 2004; Martín, Reggiani, & Martin, 2007; Páez, Scott, & Morency, 2012; Reggiani, 2012)). Accessibility measurement is one of the key issues in the current research into transportation analysis (Farrington, 2007; Geurs & van Wee, 2004); as recently stated: "as long as the friction of distance continues to exist, accessibility will remain a relevant component of transportation studies" (Páez et al., 2012).

It has been suggested that further refinements are needed in the spatial analysis techniques used to compute accessibility measures (Kwan & Weber, 2008; Langford & Higgs, 2006; Monzón, Ortega, & López, 2013; Tsou, Hung, & Chang, 2005). One of such refinements refers to the sensitivity of accessibility results to the choice of the zoning system configuration, which is related to the well-known modifiable areal unit problem (MAUP) (Amrhein, 1995; Dark & Bram, 2007; Fotheringham & Wong, 1991; Openshaw, 1984). The MAUP deals with the implications of the selection of the boundaries of the study area, and their implications on accessibility analysis are currently under study –see the work on the influence of the *scale* of analysis or *planning level*– (Gutiérrez, Condeço-Melhorado, López, & Monzón, 2011; Ortega, López, & Monzón, 2012) and on how it is structured and modelled– i.e. the *configuration of the zoning system* (Langford & Higgs, 2006; Ortega,

* Corresponding author. Department of Forest and Environmental Engineering and Management, MONTES (School of Forest Engineering and Natural Environment), Universidad Politécnica de Madrid, 28040 Madrid, Spain.
E-mail address: e.ortega@upm.es (E. Ortega).

https://doi.org/10.1016/j.apgeog.2018.03.020
Received 9 December 2015; Received in revised form 28 February 2018; Accepted 27 March 2018

López, & Monzón, 2014).

Our research addresses another issue in accessibility analysis interrelated but different than the MAUP: the importance of the spatial data allocation procedure. The computation of most accessibility indicators requires the modelisation of variables of both the transport and land use systems (Geurs & Ritsema van Eck, 2001; Geurs & van Wee, 2004) and their implementation in a GIS. The combined use of data from these two systems introduces some methodological issues, as the corresponding data are computed and stored in different ways. On the one hand, regarding the transport system, the variables stored are the travel times between a set of transport network nodes, which act as trip origins (O) and destinations (D). These travel times are computed for each node and stored in an O/D matrix. On the other hand, land use system data, such as population data, are not available at these nodes; it is frequent to find population values for certain administrative units (such as e.g. municipalities or census tracks), or distributed in grids -see e.g. (Freire et al., 2016). Although resulting accessibility values will inevitably be affected by the choice of the allocation procedure, there is little research quantifying this effect.

In addition, the analyst needs to define the areal units in which spatial data are to be allocated; these are mostly using a grid cell, or existing administrative divisions. Surprisingly, there is currently no consistent theoretical basis guiding the definition of areal units (Jacobs-Crisioni, Rietveld, & Koomen, 2014; Wu, 2004); indeed it seems to rely mostly on "*arbitrary, subjective reasons or scientific tradition*" (Verburg, Schot, Dijst, & Veldkamp, 2004), or on a confusing mix of factors related to data availability, accuracy requirements and computing effort restrictions (Huby, Cinderby, White, & de Bruin, 2009; Kwan & Weber, 2008). Continuous advances in GIS and computing capacities have made it easier to reduce the size of spatial units, at the sake of the use of additional time-consuming resources (Guo & Bhat, 2004; Jacobs-Crisioni et al., 2014; Sparks, Bania, & Leete, 2010). Although it is widely acknowledged that the recommendations derived from these analyses may potentially be inaccurate and based on erroneous findings (Kwan & Weber, 2008; Langford & Higgs, 2006; Páez et al., 2012; Tsou et al., 2005), there is little systematic research on the implications of alternative configurations (Huby et al., 2009; Langford & Higgs, 2006; Ortega et al., 2012, 2014).

A few notable exceptions are briefly reviewed. First, the research by Hewko, Smoyer-Tomic, and Hodgson (2002), which demonstrated that assuming evenly distributed populations within midsized areas can lead to aggregation error in calculating accessibility indicators. Another key reference is that of Apparicio, Abdelmajid, Riva, and Shearmur (2008), who also found differences in accessibility values calculated using two alternative aggregation methods: (a) census tract centroids; and (b) a more accurate method: a population-weighted mean of the accessibility values for census blocks within census tracts. The same rationale informs the work of Langford and Higgs (2006), who used accessibility measures to assess spatial inequalities in healthcare deliveries. They investigated the implications for equity results of adopting three alternative spatial representations of population, and concluded that the choice of the spatial representation model plays a key role in equity outcomes. Another example is a recent research (Jacobs-Crisioni et al., 2016), which highlights the importance of incorporating future population levels when assessing accessibility changes of transport infrastructure investments.

In this context, our research work investigates the influence of alternative population allocation procedures in resulting accessibility measures. We have selected as a case study the analysis of the accessibility *to* high speed rail (HSR) stations, particularly on low-density/rural areas. When a new HSR network is built, these are the areas that predominantly suffer the closure of existing conventional regional rail stations and services (Martínez Sanchez-Mateos & Givoni, 2012; Monzón et al., 2013, 2016). The research computes accessibility to HSR stations to determine the influence of: (1) how spatial units are designed, using different geographical representation procedures; and (2) how the population is assigned to these spatial units. Our approach combines different spatial unit designs and population distribution procedures to produce five different spatial data allocation procedures methods for testing in a case study. The results of the paper provide useful guidelines to interpret the influence on the results of different geographical representation and population disaggregation methods.

The structure of the paper is as follows. Section 2 describes the analysis of spatial data allocation procedures and includes the case study of two different rural areas in Spain served by a HSR line. Accessibility analysis and spatial data allocation implications are discussed in Section 3; and finally Section 4 contains some concluding remarks.

2. Spatial data allocation procedures analysis

The influence of spatial data allocation procedures on accessibility values is analysed by measuring accessibility to high-speed rail stations under alternative procedures. First, an analysis of the spatial data allocation is made, considering a case study, to define the spatial units and the origins for the accessibility calculations.

2.1. Case study description

The research topics are investigated by applying the methodology to two study areas: the two Spanish provinces (NUTS-3[1]) of Ciudad Real and Cuenca. They are about the same size and are located in the middle of two different HSR lines connecting Madrid (the capital city) with Seville and Valencia respectively (see Fig. 1). They were selected for their different population densities.

Ciudad Real has 530,250 inhabitants distributed in 102 municipalities (Fig. 2), of which 13 have more than 10,000 and are well connected by road. Ciudad Real is on the HSR corridor connecting Madrid with Seville and has two HSR stations: the city of Ciudad Real (75,000 inhabitants) and Puertollano (52,000 inhabitants).

Cuenca province has only 219,138 inhabitants scattered in 237 municipalities (Fig. 2). Unlike Ciudad Real, its towns are less populated and more dispersed, and are poorly connected by road. The north of the province is a mountainous area with low-quality roads with a low average speed. Cuenca is on the HSR corridor that connects Madrid with Valencia and has only one HSR station: Cuenca (57,000 inhabitants). Fig. 2 shows the municipality population distribution in the two provinces.

2.2. Design of spatial data allocation procedures

The definition of each type of spatial data allocation procedure for the definition of each spatial data unit *z* depends on how they represent the territory. This representation will depend on the data representation method used (vector or raster), and the level of detail (map accuracy or cell size).

If a vector model is selected, the population is assigned within the administrative boundaries of the polygons that configure the territory, while in a raster format the territory is divided into cells, each of which must have a value. Population values are then "extended" throughout the territory using GIS tools. The cell size has important implications for the precision of the data needed in the planning process. It must be small enough to capture the required detail but large enough to perform efficiently. A smaller cell achieves greater detail in the size of the features represented in a raster layer, and in a raster format there are also several methods for calculating the population distribution in the

[1] NUTS-3: The NUTS classification is a hierarchical system for dividing up the economic territory of the EU for the purpose of the collection, development and harmonization of EU regional statistics. NUTS-3 corresponds to the third level (provinces in the case of Spain). NUTS-5 corresponds to the fifth level (municipalities in the case of Spain).