Journal of Transport Geography 72 (2018) 1–12

Contents lists available at ScienceDirect

Journal of Transport Geography

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

Effect of distance from home to school and spatial dependence between homes on mode of commuting to school

Jorge Chica-Olmo[a,*], Carlos Rodríguez-López[b], Palma Chillón[b]

[a] Department of Quantitative Methods for Economics and Business, Faculty of Economics and Business Sciences, University of Granada, Granada, Spain.
[b] PROFITH "PROmoting FITness and Health through physical activity" research group, Department of Physical Education and Sport, Faculty of Sport Sciences, University of Granada, Granada, Spain.

ARTICLE INFO

Keywords:
Distance
Active commuting
School
Areas of influence
Spatial autocorrelation
Regression-kriging

ABSTRACT

Active commuting to school (i.e., walking and cycling) has health implications for young people. Therefore, it of interest to determine how the distance students walk to school varies depending on where they live and how their decision to walk is affected by contextual/environmental variables. This study aimed to examine which of the distances (Euclidean, Manhattan, walking-network and driving-network) is the best predictor of the decision to walk to school and determine the areas of influence of active commuting to school for four high schools in Granada, Spain. To achieve these aims, the regression-kriging method was used. The results indicated that the Euclidean and the walking distances were the best predictors of the decision to walk to school. Spatial dependence produced by some locational variables and spatial contagion among students was found to be moderate to strong. In addition, the spatial range of this spatial dependence is approximately 1000 m to 1600 m. Regression-kriging could be implemented in a geographic information system to determine the areas of influence of schools and aid urban designers and planners in developing neighborhoods that support active modes of commuting. Identifying the areas of influence is important for promoting active modes of transport by local governments.

1. Introduction

Most urban models, such as "new urbanism", aim to minimize car use and its negative impact on the environment while promoting more pedestrian-friendly features (Crane and Crepeau, 1998). These models focus on reducing the distance between locations and increasing the feasibility of alternative modes of transport, such as walking, cycling or public transportation. To this end, policies have been designed to reduce road congestion, as well as urban pollution and dispersion. These factors are also associated with the location of schools in urban areas. Studies that have investigated the effects of school location have been based on aggregated (zonal) or detailed (household) data. The latter type are more appropriate when studying the effects of distance and accessibility on the choice of commuting mode (Handy, 1996).

The commuting mode children use to go to school has been widely discussed in recent years and has economic, social, health and environmental effects (Faulkner et al., 2013; Li and Zhao, 2015; Mandic et al., 2015; Wilson et al., 2010). The effect of the distance from household to school on transport mode choice is well known. Several studies have shown that the distance from home to school is a stronger predictor of active commuting to school (ACS) in school-age students (Chillón et al., 2015; Davison et al., 2008; McDonald, 2008) and that shorter distances are associated with higher rates of active travel (Mandic et al., 2015; Pont et al., 2009; Rodríguez-López et al., 2017). Previous evidence is available on the current distance that young people are willing to walk to school. In the United States, 31% of trips from home to school made by walking are under one mile (1.6 km) (US Department of Health Human Services, 2008). The criterion distance for walking to school has been reported to be 1.2 miles (2.0 km) among Belgian adolescents (Van Dyck et al., 2010) and 1.5 miles (2.4 km) among Irish adolescents (Nelson et al., 2008). In the UK, the statutory walking distance for English children has been set at 3 miles (4.8 km) (Government/DfES, 2005), while for Japanese students it is approximately 2.5–3.7 miles (4.0–6.0 km), although the actual distance depends on the children's age (Mori et al., 2012). Although there is no universal criteria, some studies have shown that a distance of about 2.5 miles (4.0 km) is considered reasonable for adolescent walkers (Nelson et al., 2008). Therefore, one question of interest is to determine the distance that children and adolescents are expected to walk from home to school. In this study, walking distance is defined as the radius

* Corresponding author at: Faculty of Economics and Business Sciences (Department of Quantitative Methods for Economics and Business), University of Granada, Campus de la Cartuja s/n, 18071 Granada, Spain.
E-mail addresses: jchica@ugr.es (J. Chica-Olmo), crdguez@ugr.es (C. Rodríguez-López), pchillon@ugr.es (P. Chillón).

https://doi.org/10.1016/j.jtrangeo.2018.07.013
Received 1 December 2016; Received in revised form 24 May 2018; Accepted 31 July 2018
Available online 10 August 2018
0966-6923/

of influence around the school, which may vary depending on the particular locational factors of each area of the city, such as crime, traffic, the environment, urban design, socioeconomic characteristics, and others.

As a result of these factors, the area of influence of schools need not be necessarily circular but may form irregular polygons within which the school is located. To the best of our knowledge, few studies have investigated the effects of spatial dependence and distance between households on travel mode choice and the presence of spatial distribution patterns (Mitra et al., 2010). In our study it has been assumed that households located close to each other in a given geographic area are affected by similar locational factors. These locational factors may cause young people living in the same area to encourage others located nearby to make similar transport mode choices, thus leading to the presence of spatial autocorrelation. In addition, these variables are not always easy to observe, measure or specify in a specific model. Moreover, since the decision to walk to school may depend on locational factors, it could cause a contagion effect, which can have a major impact on urban traffic, environmental pollution and climate change, but also provide substantial health benefits (de Hartog et al., 2010).

ACS provides an opportunity for increasing daily physical activity level (Chillon et al., 2011; Panter et al., 2016; Sun et al., 2015; van Loon and Frank, 2011). Indeed, it has been shown that youth who walk to school are more physically active than those who are driven (Faulkner et al., 2009). Because ACS is a complex behavior influenced by multiple factors at different levels of the person-environment system and has several public health benefits, it is of interest to conduct research that can aid policymakers in implementing effective strategies targeted at promoting ACS.

Several methods have been used in the literature to measure distance, among them self-reported questionnaires to determine the distance of the route from home to school (Burke and Brown, 2007; Panter et al., 2010), geographic information systems (GIS) (Panter et al., 2010; Timperio et al., 2006), sketch maps of the route (Schantz and Stigell, 2009), Google Maps™ (Mendoza et al., 2011; Voss and Sandercock, 2010) or global positioning systems (GPS) (Duncan and Mummery, 2007; Duncan et al., 2007). Discrete choice models, such as the classic logistic regression model or logit model which are equivalent to log-linear models (Wasserman and Pattison, 1996), have been used to determine the probability that an individual will choose a particular means of transport based on a set of explanatory variables. The classic logistic regression model has been widely used in the literature on the mobility of children to schools (Pont et al., 2009). Different types of geometric distances can be measured in different ways. Although the most commonly used measure of distance in the literature on child mobility is the Euclidean distance (McDonald, 2007a), other measures of distance, such as the Manhattan distance, the walking distance and the driving distance, are also used in this study (a graphical description of each is shown in Fig. 1). In addition, the classic logistic regression model is based on the assumption of independence of perturbations, which are considered unobserved random variables. This hypothesis is not satisfied when the presence of spatial autocorrelation between perturbations is suspected due to the presence of locational factors and contagion effects. If spatial autocorrelation is present in the model perturbations, a spatial regression model is more suitable (Overmars et al., 2003). Therefore, to account for the presence of spatial autocorrelation in perturbations, the regression-kriging method is used (Cressie, 1991). In addition to spatial dependence, spatial heterogeneity is another aspect to be considered in the regression model (Anselin, 1988). Geographically weighted regression is a method that has been shown to be suitable for modeling spatial heterogeneity in the field of transportation (Feuillet et al., 2018).

Although the Euclidean distance is the most commonly used metric in research on child mobility, other types of geographical distances have also been analyzed, such as the network and Manhattan distances (Arafat and Abed Al Musa, 2017). These authors examined the network

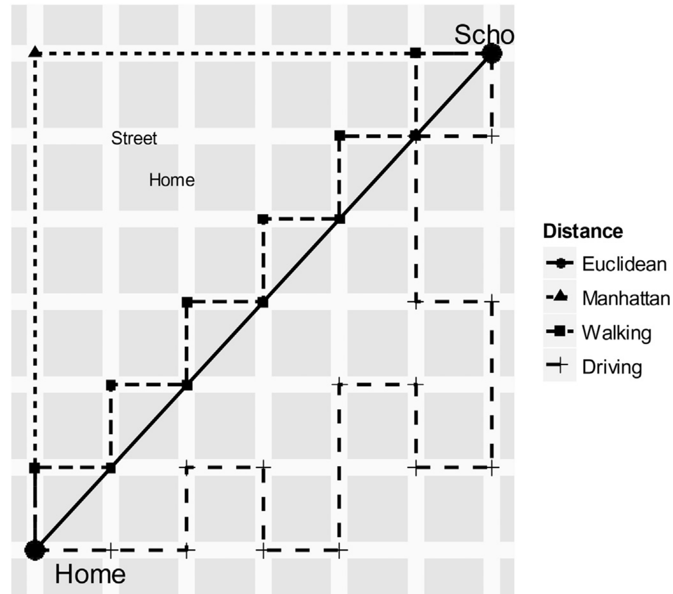

Fig. 1. Graphical representation of the four types of distances.

and Manhattan distances and the presence of spatial autocorrelation in the model perturbations. Model perturbations are due to the contagion effect of household, family and unobserved locational factors (crime, traffic, the environment, urban design, socioeconomic characteristics, etc.).

Using a survey conducted in 2012 of 527 students aged 12–18 years old attending four high schools located in the city of Granada, Spain, the areas of influence of active commuting for each school have been determined for four types of distances. The study has two aims: (a) to examine which of the four types of distances (i.e., Euclidean, Manhattan, walking-network and driving-network) could be the best predictor of the decision to walk to school and (b) to determine the areas of influence of ACS for four high schools in the city of Granada, Spain, considering the distance from home to school and the spatial dependence between households.

The rest of the paper is structured as follows. In the next section, a review of the literature on the factors that influence children's decision to walk to school is presented, focusing on the presence of spatial dependence. The data and methodology are then described in the following section. Finally, the results are discussed and analyzed and conclusions are drawn, highlighting the most important findings of the research.

2. Literature review

ACS behavior is associated with different demographic, personal, school, family, environmental and social factors (Chillón et al., 2014; Hume et al., 2009; Sirard and Slater, 2008). The main predictive factor of ACS is distance from home to school, which is examined here in relation to spatial dependence.

2.1. Accessibility, distance and proximity

Geographic information systems (GIS) use different methods to measure accessibility. The measure of accessibility between two locations includes the concept of time, transport mode and physical distance (Jang and Kang, 2015). Accessibility can be measured as the distance, cost or time between two locations. In a study by Apparicio et al. (2003), the authors examined whether physical distance and time are correlated, and concluded that at the metropolitan level using distance to estimate the shortest network time does not introduce major error. Geographical proximity can be measured by some type of physical distance (i.e., Euclidean, Manhattan, walking-network and driving-network). Distance is a measure of the proximity between the household and the school, but is also considered the main barrier to walking to school (McDonald, 2007b) in addition to other environmental variables such as crime and traffic (Pont et al., 2009). The closer the household and school are to each other, the greater the probability that children will walk to school (Chillón et al., 2014). The Euclidean distance and the Manhattan distance are two common and easy to calculate measures for determining geographical proximity. The Euclidean distance is the straight-line or "as the crow flies" distance, while the Manhattan distance is the sum of the horizontal and vertical components of a right-angled triangle. Another measure of physical distance is the shortest network distance on foot or by car (Apparicio et al., 2008). Of the four types of distances addressed here to measure how users perceive the concept of proximity, the Euclidean distance is the shortest, whereas the driving distance is the longest due to the presence of one-way roads and pedestrian streets (Fig. 1). Golledge and Hubert discussed how the use of a Euclidean metric can be supported for the construction of mental maps for interurban analysis (Golledge and Hubert, 1982). The Euclidean, Manhattan and network distances and time have been used to measure accessibility to health services (Apparicio et al., 2017) and green space (Higgs et al., 2012), as well as walking to green spaces, cycling to public high schools, driving to health care services, and walking, cycling and driving to supermarkets (Logan et al., 2017). In this paper, the Euclidean, Manhattan and shortest network distance on foot and by car are used.

2.2. Spatial dependence in mode choice

Spatial dependence can be defined as the existence of a relationship between what happens at a point of the plane and what happens elsewhere (Anselin, 1988). Spatial dependence is also referred to as "spatial autocorrelation" (Anselin and Rey, 1991), although the notion of spatial autocorrelation is more limited than that of spatial dependence (Anselin and Rey, 1991). In practice, both concepts are usually considered the same (Anselin, 1988; Anselin, 1999). However, in the geostatistics and spatial statistics fields, Cressie (1991) primarily used the term "spatial dependence." To clarify, in the current study, the term "spatial autocorrelation" refers to a variable or model perturbations (in the sense of Anselin and Bera, 1998, "as a moment of the joint distribution") and the term "spatial dependence" is used when referring to the phenomenon under study. The idea of spatial dependence is expressed in the First Law of Geography, which posits that "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970).

In the case of transport mode choice, spatial dependence is expected to occur among individuals who are neighbors, since the locational variables, such as crime rates or traffic, will have a similar effect on decisions regarding the mode of ACS. For example, a low crime rate in an area will positively influence the decision to walk in that area. In a context of this type, a positive face-to-face spatial contagion is also expected, since neighboring individuals will tend to make similar decisions. Thus, for example, if an individual decides to walk to school, this can influence their neighbors directly (positive contagion), although it could also have an additional indirect effect (negative contagion). However, because both locational variables and contagion are often unobservable, they are omitted in the models.

The concept of spatial dependence has been addressed in different fields such as ecology, forestry, geology, medical, urban demography and land use, among others (Anselin and Bera, 1998). These authors also highlight that spatial dependence is due to certain misspecifications which result in spatial measurement errors and spatial autocorrelation between the perturbances. In this line, Dubin (1988) found that omitted variables will be included in the error term, causing this term to be spatially autocorrelated. On the other hand, Galster et al. (2001) explored a wide variety of concepts related to the land use patterns (i.e., density, centrality, contiguity, continuity, etc.) that explain the level of sprawl, which might be related to the concept of spatial dependence.

However, few studies have addressed the spatial interaction effects that influence children's transport mode choice. These effects may occur because households share common unobserved factors (Sidharthan et al., 2011) that depend on the location and are difficult to observe, measure and specify in a model. These factors could be spatially autocorrelated due to spatial interaction in two ways: across spatial zones, neighborhoods or blocks and across individual behaviors (Rainham et al., 2012). Another cause of spatial dependence is spatial spread, since individuals who are close in space are influenced by their neighbors' decisions (contagion), thus prompting other children to choose the same method of transport (Sidharthan et al., 2011).

The most common barriers to ACS are distance and traffic (Beck and Greenspan, 2008; Chillón et al., 2014). Moreover, Beck and Greenspan (2008) showed that families' behavior is not homogeneous throughout a geographical area and that their choices may differ depending on the area in which they live. Therefore, besides distance, there are other locational factors that cause school travel choices to be dependent on location. These locational factors together with the contagion effect may influence students' decisions, resulting in a spatial distribution of that does not form concentric circles around high schools. Indeed, families' decisions are likely related to the characteristics affecting a given area, leading to the presence of spatial dependence. Pont et al. (2009) conducted a review of some of the main factors related to ACS. The authors investigated the physical, economic, social, cultural and political factors of active transportation among young people. Of these factors, we are particularly interested in the environmental characteristics of the neighborhood as they have a spatial influence on the mode choice for school transportation (Mitra, 2013). This author (Mitra and Buliung, 2012) analyzed the presence of scale effects on the relationship between ACS and the built environment, taking into account the modifiable areal unit problem (MAUP).

In addition to distance, we address the effect of both local, face-to-face contagion and locational variables that have similar effects on households located in the same area and lead to the presence of spatial dependence.

On the other hand, the classic logistic regression model assumes that the decisions of individuals living in nearby houses are independent. However, locational factors such as crime, traffic, environmental factors, among others, may influence these decisions, leading to spatial dependence. Sidharthan et al. (2011) estimated a spatial autoregressive model (SAR) to test the effect of spatial interdependence on the choice of transport mode. They found that the decision of an individual related to children's school trips can affect the decisions of other individuals in their environment. This spread may be due to spatial diffusion, social spillover and effects related to unobserved variables that are spatially autocorrelated (Anselin, 1988; Ripley, 1981). For example, if a person decides to use active transport, this could persuade their neighbors to do the same. Thus, a spatial contagion effect occurs by which some individuals encourage other individuals in their neighborhood to make similar decisions (Boix and Trullén, 2007). This spread effect, in addition to the effect of locational variables, which will be similar for all families living in the same area, can lead to the presence of spatial dependence. In a similar line, Mitra et al. (2010) detected the prevalence of walking to and from school for children aged 11–13 years old in Toronto and the global presence of spatial autocorrelation and the local clusters of walking trips.

In this paper, the regression-kriging method (RK) is used to achieve the two main objectives of this work: to determine which of the four types of distance best explains ACS and the area of influence of each school. To achieve the first objective, 16 models have been estimated (four for each high school) and the results have been compared. To achieve the second aim, ACS has been predicted taking into account not only the distance from home to school, but also the presence of spatial autocorrelation among the perturbations. We assume that the presence of spatial correlation is due to the contagion effect of neighbors and the locational variables (crime, traffic, the environment, urban design, socioeconomic characteristics, etc.). These predictions have been used to obtain isoprobability maps that represent the area of influence of each school.

3. Data description

3.1. Participants

A cross-sectional study was carried out in November 2012. Students aged 12–18 years old attending four high schools (Alhambra, Ganivet, Soto de Rojas and Zaidín high schools) located in the city of Granada were invited to participate in the study. All of the high schools were public with a student body of a low to medium socio-economic level and were located in residential neighborhoods close to the city center of Granada. The schools were recruited using convenience sampling. High schools in Spain are divided into areas of influence to give preferential access to families whose residence or place of work is located within each school zone. Although there is a limit on the number of students per high school, it is compulsory for all students to be admitted to a high school. A total of 527 adolescents accepted to participate and the inclusion criteria was to have provided data on the mode of commuting to school and an accurate household address to calculate the distance from home to school. The sample represents 43% of all students attending these schools.

The research ethics committee of the University of Granada approved the research and the informed consent procedure (case no. 817). The schools that participated in the study were informed about the purpose of the research and were responsible for informing the students and parents, as well as obtaining their approval and written consent.

3.2. Mode of commuting to/from school

The students completed a self-reported questionnaire which included questions that have been proposed as the most appropriate measure for assessing mode of commuting to school based on a systematic review of 158 studies (Herrador-Colmenero et al., 2014). The questions have been recently validated in young Spanish people (Chillón et al., 2017). Specifically, the students were asked how they usually commuted to and from school. The response options were: walk, bike, motorbike, car, bus. A dichotomous variable was obtained: walking and cycling were categorized as "active commuting," and the use of car, motorbike or bus as "passive or motorized commuting." Respondents who actively commuted on at least one leg of the trip (to or from school) were categorized as active or walking, while those who passively commuted both to and from school were categorized as passive or motorized. Given that most active commuters were walkers (less than 0.3% of participants biked to school), a dichotomous variable walkers vs. motorized commuters was created using the same process and omitting those who biked in order to study the criterion distance.

Four types of distances were considered to capture proximity effects: the Euclidean distance, the Manhattan distance, the shortest walking network distance and the shortest driving distance. The first two distances (Euclidean and Manhattan) were obtained automatically once the 527 households had been georeferenced (Fig. 2) according to the following expression (Arafat and Abed Al Musa, 2017):

$$D_{ij} = ((u_j - u_i)^p + (v_j - v_i)^p)^{1/p} \quad (1)$$

where u and v are geographical coordinates and $p = 1$ for the Manhattan distance and $p = 2$ for the Euclidean distance.

The two network distances (walking and driving) were calculated using Google Maps™ software, which is considered an accurate method (Logan et al., 2017). This software allows selecting the origin (the house), the destination (the school) and the transport mode. Each of the households (origin) of the students attending the four schools (destinations) was located, and the shortest walking network path and shortest driving network path from home to school was chosen. All the distances were obtained and expressed in meters. The four distances were used as the explanatory variable in the 16 models that were estimated using logistic regression as discussed in the results section.

3.3. Distance from home to school

The objective measure of the commuting distance from home to school for each participant was estimated selecting the shortest walking network path between the home address and the school using Google Maps™ software. Only in case where it failed to recognize the address, Via Michelin software was used to locate the address and it was then manually extrapolated to Google Maps software. The walking distance from home to school (i.e., distance for those participants who only walked) was calculated. The distance was collected and expressed in meters.

4. Methods

Kriging is a method used for spatial prediction in the field of geostatistics. Although there are many variants of kriging, we chose regression-kriging (RK) because it is the most suitable method when there are auxiliary variables and it can be more easily combined with other methods. RK is a spatial interpolation method that combines a multiple-linear regression model (or a variant of generalized linear models, generalized additive models, etc.) with kriging of the regression residuals (Hengl et al., 2007a; McBratney et al., 2000). This method has been used in several fields. Odeh et al. (1995), for instance, compared different methods and concluded that RK generally performed well. In the housing market, an iterative RK method has been used for the spatial prediction of housing prices and locational rents (Chica-Olmo, 1995). A network RK method was used to estimate the ridership of the new Second Avenue Subway in New York City (Zhang and Wang, 2014), while a mapping methodology was used to analyze urban structure characteristics and children's active transport (Broberg et al., 2013).

In our specific case, the aim is to obtain isoprobability maps in order to estimate the likelihood that a child living in any location of the plane will walk to school. These probabilities are represented on the plane and permit "drawing" the polygon that determines each school's area of influence.

When the outcome is a binary variable, RK uses a logistic regression. Other authors (Hengl et al., 2007a) have used RK with logistic regression to map the presence/absence of a binary variable and the result was the probability of occurrence of this variable. The predictions can be made using the following expression (Hengl et al., 2007b):

$$\hat{z}(s_0) = \hat{m}(s_0) + \hat{e}(s_0) = \frac{1}{1 + \exp(-\hat{\beta}' x(s_0))} + \hat{e}(s_0) \quad (2)$$

where $0 \le \hat{z}(s_0) \le 1$ is the estimated response of the binary variable (in our case this binary variable is the mode of ACS, as we are interested in predicting z in any location s_0); the binary variable is observed in locations $s_1, s_2, \ldots s_n$ (the households) where $s_i = (u_i, v_i)$ and u_i, v_i are the spatial coordinates; $\hat{m}(s_0)$ is the trend (this is the classic logistic regression and allows us to obtain the probability that a student will travel on foot considering the distance from home to school); $\hat{\beta}$ is a vector kx1 of the estimated regression coefficients; $x(s_0)$ is a vector kx1 of the explanatory variables (the distance to school) and $\hat{e}(s_0)$ is the estimated kriging residual. The residuals are estimations of the perturbations of the classic logistic regression model.

As indicated in the literature review, ACS behavior is associated

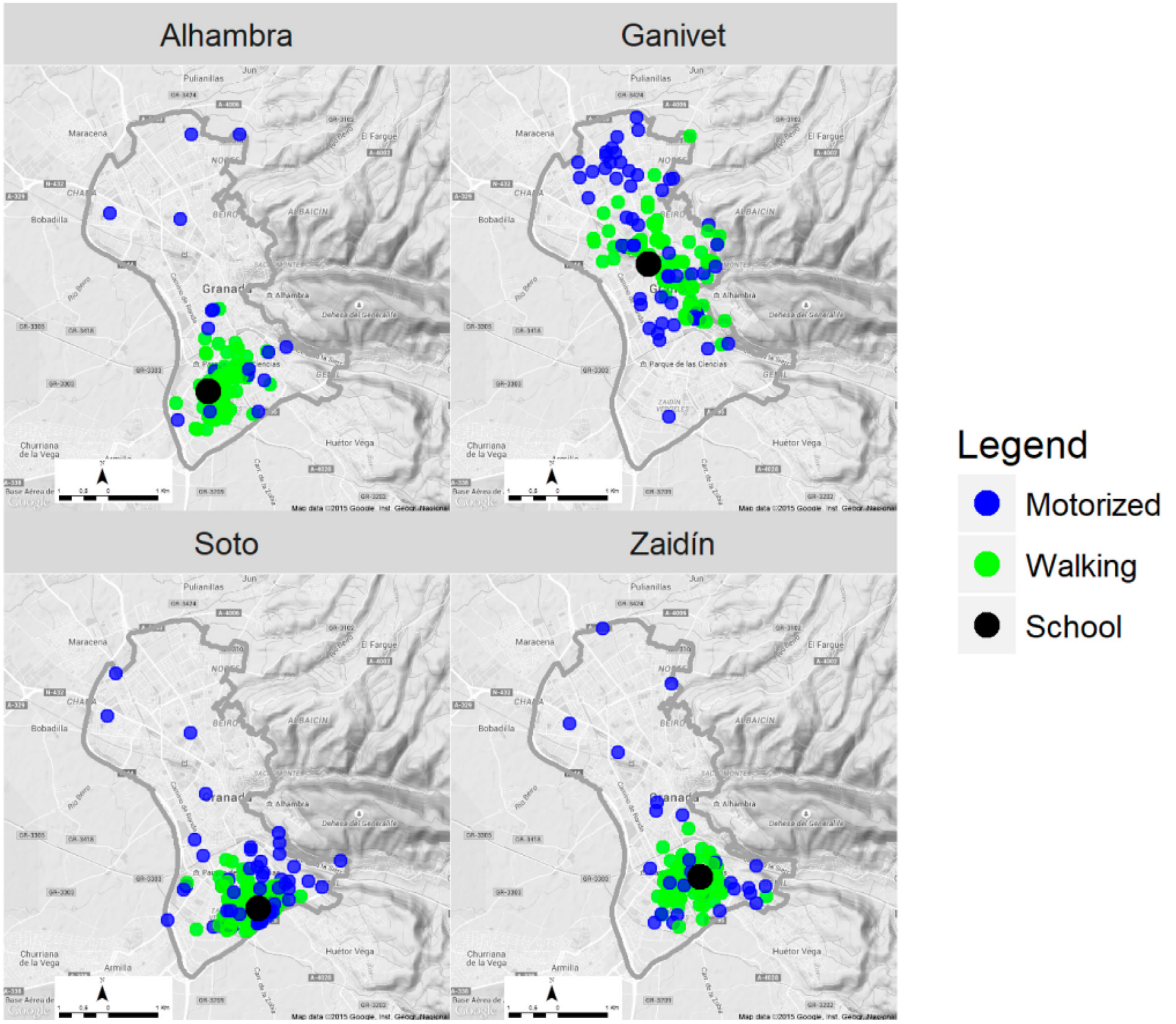

Fig. 2. Location of households (blue and green points) and schools (black points). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

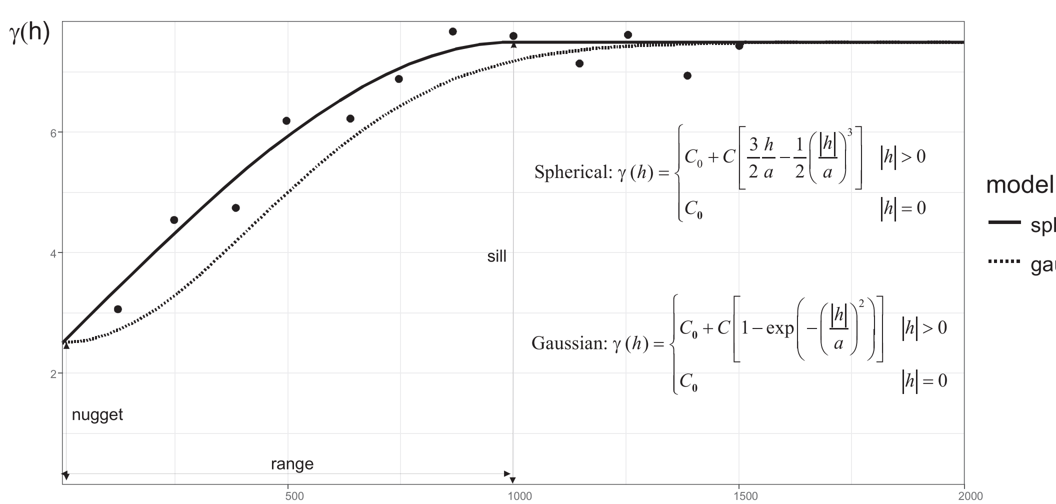

Fig. 3. Empirical variogram (dots) and spherical (sph) and Gaussian (gau) variogram models.

Table 1
Descriptive statistics of transport mode choice, distance to school and distance between neighbors, in meters.

School (n)	Alhambra (103)		Ganivet (127)		Soto de Rojas (187)		Zaidín (110)	
Mode choice	Walking	Motorized	Walking	Motorized	Walking	Motorized	Walking	Motorized
Frequency (ni)	87	16	72	55	132	55	84	26
Percent (%)	84	16	57	43	71	29	76	24

Distance to school	Eucl.	Walk.	Manh.	Driv.	Eucl.	Walk.	Manh.	Driv.	Eucl.	Walk.	Manh.	Driv.	Eucl.	Walk.	Manh.	Driv.
Min.	62.94	120	79	120	4.89	30	5	30	30.32	30	56	30	48.26	66	87	66
Max.	5636.69	7100	6314	9500	3354.89	5100	3752	9200	5970.9	6600	8194	9200	5804.58	6600	7512	7700
Mean	815.1	1054.85	1017.77	1460.19	1255.4	1521.91	1595.19	2709.7	660.24	839.63	851.17	1196.47	753.44	989.15	968.25	1274.6
SD	929.55	1212.37	1078.12	1427.59	739.84	877.75	936.34	1680.27	717.15	822.12	973.9	1073.99	828.88	992.75	1086.03	1194.8
t-test of difference between means and correlation coef.	Eucl.	Walk.	Manh.	Driv.	Eucl.	Walking	Manh.	Driv.	Eucl.	Walk.	Manh.	Driv.	Eucl.	Walk.	Manh.	Driv.
Eucl.	–	4.739 (0.000) [0.92]	10.005 (0.000) [0.99]	8.733 (0.000) [0.88]	–	8.916 (0.000) [0.69]	15.055 (0.000) [0.98]	11.367 (0.000) [0.51]	–	4.549 (0.000) [0.76]	9.800 (0.000) [0.99]	9.837 (0.000) [0.73]	–	8.446 (0.000) [0.96]	8.115 (0.000) [0.99]	11.400 (0.000) [0.95]
Walk.		–	0.794 (0.429) [0.92]	8.451 (0.000) [0.95]		–	1.991 (0.048) [0.67]	10.040 (0.000) [0.72]		–	0.238 (0.811) [0.75]	11.659 (0.000) [0.93]		–	0.693 (0.489) [0.96]	9.637 (0.000) [0.98]
Manh.			–	6.305 (0.000) [0.88]			–	8.803 (0.000) [0.53]			–	5.958 (0.000) [0.70]			–	7.717 (0.000) [0.94]
Driv.				–				–				–				–

Distance between neighbors	Alhambra	Ganivet	Soto	Zaidín
Min.	3.16	6.4	3.1	6.4
Max.	6491.97	6549.64	6461.11	6849.99
Mean	1136.97	1722.4	953.33	1105.82
SD	1191.52	1021.31	903.01	1087.97

Note: t-test of difference between means, *p*-values shown in parenthesis. Correlation coefficients shown in brackets.

with a range of personal, school, family and social explanatory variables (age, income, etc.), which can have an influence on the fact that two students who are at the same distance make different transport mode choices. However, these personal factors have not been included in the models presented in this paper because one of the objectives has been to obtain isoprobability maps, and to obtain such maps the explanatory variables must be known at any point of the plane. For example, the age of the surveyed individuals is known in the households where they live, but is not known at any point on the map (s_0). Therefore, we consider an individual with average personal characteristics, and assume that the decision made by that individual depends only on factors related to the location. The isoprobability map shows how likely that particular individual will be to vary his or her transport mode choice depending on the location.

Once this individual is selected, the individual's mode of ACS will vary spatially as a function of the relevant explanatory variables, such as distance to school, the locational variables and contagion. For various reasons, the locational variables and contagion are sometimes not measured, so they are omitted in the models. We assume that the omitted variables are present in the perturbations and that the presence of spatial autocorrelation in them is caused by the omission of these relevant explanatory variables (Dubin, 1988).

In geostatistics, the key instrument to detect the presence of spatial dependence is the variogram (Matheron, 1965; Cressie, 1991; Hengl et al., 2007a; Hengl et al., 2007b; Chica-Olmo, 1995). An unbiased estimator of the variogram of residuals is (Matheron, 1965):

$$\hat{\gamma}_e(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [e(s_i + h) - e(s_i)]^2 \tag{3}$$

where $(s_i + h)$ and (s_i) are the locations of two households (one located in s_i the other located at an 'h' distance from the household located in s_i) and $N(h)$ is the number of h distant household-pairs, that is, the number of households that are separated by an 'h' distance. The value of the empirical variogram ($\hat{\gamma}_e(h)$) is shown in dots for households that are separated by an approximate distance of 125 m, 250 m, 375 m and so on (Fig. 3). It can also be observed how the spatial variability between the residuals increases as the distance between the households increases.

To perform estimations with the kriging method, it is necessary to fit a model to the empirical variogram. To do so, we have used the spherical and Gaussian variogram models (see Fig. 3).

The model fit depends on three parameters: the nugget effect (C_0), the range (a) and the sill ($C_0 + C$), where C is the partial sill. The nugget effect is a measure of a random component, the range is defined by the value of h in which the value of the variogram stabilizes and the sill is the value that the variogram model attains at the range. The range in Fig. 3 is 1000 m, meaning that the residuals are spatially autocorrelated up to a distance of approximately 1000 m. The Gaussian and spherical models are two of the most commonly used models in geostatistics and represent continuous phenomena (Armstrong, 1998). Both models show how the distance between houses increases, the variability between the residuals also increases. This may be due to the fact that students who live in houses proximate in space are affected by similar locational variables and face-to-face contagion, and logically make similar decisions. In contrast, when the distance between the houses increases, these variables are less similar and the contagion is smaller, which leads the students to make less similar decisions. Moreover, the distances to everything else also increase and the areas may be perceived by the students as not very walkable.

A measure of the intensity of the spatial dependence can be investigated using the relative nugget effect, that is, the ratio of the nugget effect and sill. A ratio of less than 25% indicates strong spatial dependence, a ratio between 25% and 75% indicates moderate spatial

Table 2
Logistic regression and variogram model estimated by regression-kriging.

Dependent variable: mode of ACS 1 = walking 0 = motorized (passive)		School (sample size)			
Distance		Alhambra (103)	Ganivet (127)	Soto (187)	Zaidín (110)
Eucl.	*Logistic regression*				
	Const	4.884	2.029	2.766	3.898
	Eucl. distance	−0.0034	−0.0014	−0.0029	−0.0035
	AIC	**55.31**	150.7	**181.03**	84.56
	Δ_i	**0**	0.8	**0**	1.78
	R2	0.530	0.249	0.331	0.456
	Variogram model	Gau	Sph	Sph	Sph
	range	1427	1025	1630	1490
	p-sill	0.09	0.05	0.08	0.11
	nugget	0.03	0.16	0.11	0.05
	r-nug	25	76	58	31
Manh.	*Logistic regression*				
	const	4.795	1.930	2.566	3.714
	Manh. distance	−0.0026	−0.0010	−0.0020	−0.0025
	AIC	57.20	152.7	185.92	85.67
	Δ_i	1.89	2.8	4.89	2.89
	R2	0.504	0.231	0.302	0.445
	Variogram model	Gau	Sph	Sph	Sph
	range	1481	1123	1560	1590
	p-sill	0.10	0.05	0.1	0.12
	nugget	0.03	0.16	0.1	0.05
	r-nug	23	76	50	29
Walking	*Logistic regression*				
	const	4.109	2.123	3.819	3.939
	Walking distance	−0.0019	−0.0012	−0.0036	−0.0023
	AIC	59.83	**149.9**	186.26	**82.78**
	Δ_i	4.52	**0**	5.25	**0**
	R2	0.475	0.256	0.300	0.473
	Variogram model	Gau	Sph	Sph	Sph
	range	1550	1065	1140	1310
	p-sill	0.11	0.03	0.08	0.14
	nugget	0.03	0.17	0.11	0.02
	r-nug	21	85	58	13
Driving	*Logistic regression*				
	const	4.578	1.070	3.879	4.077
	Driving distance	−0.0017	−0.0003	−0.0024	−0.0021
	AIC	59.07	171.7	192.73	84.29
	Δ_i	3.76	21.8	11.7	1.51
	R2	0.485	0.051	0.261	0.458
	Variogram model	Gau	Sph	Sph	Sph
	range	1593	1112	1120	1390
	p-sill	0.10	0.05	0.08	0.09
	nugget	0.03	0.17	0.11	0.06
	r-nug	23	77	58	40

Note: p-value is 0.000 for all the parameters, const and distances. $\Delta_i = AIC_i - AIC_{\min}$.
Variogram model: Gau = Gaussian; Sph = spherical; r-nug = ratio of the nugget effect in %; p-sill = partial sill; R2 = Nagelkerke adjusted R-squared.
The bold means the best model (smaller AIC) for each school.

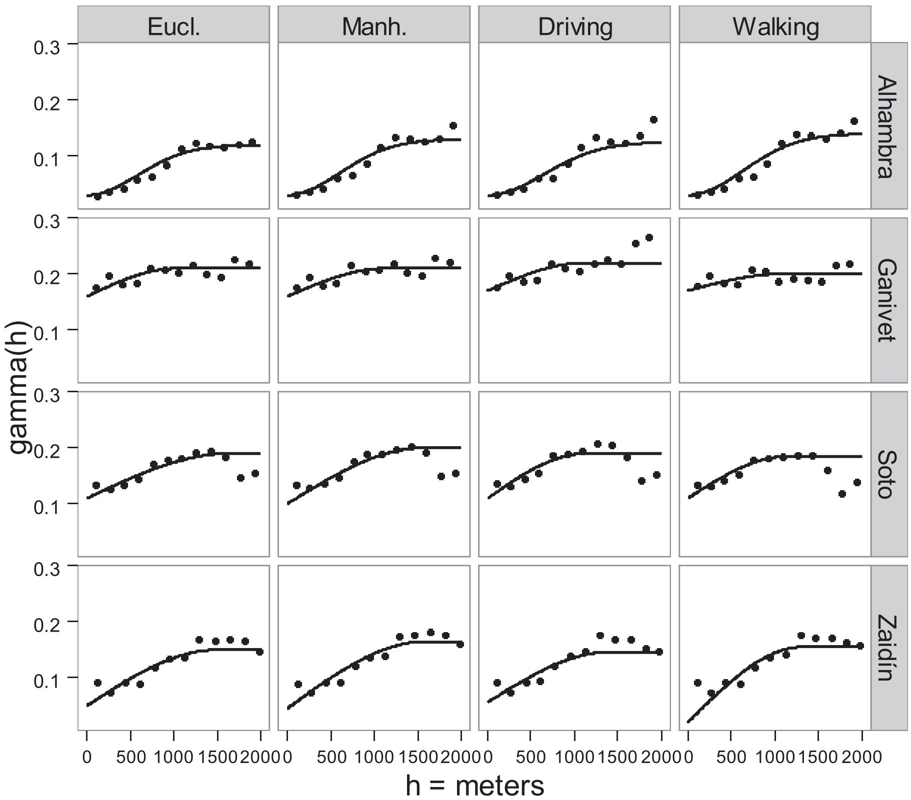

Fig. 4. Experimental variograms and fitted models for the residuals of the estimated spatial logistic regression models.

dependence, and a ratio greater than 75% indicates weak spatial dependence (Rouhani, 1996).

The kriging estimator of residuals is:

$$\widehat{e}\,(s_o) = \sum_{i=1}^{n} \lambda_i e(s_i) \tag{4}$$

This estimator is a weighted average of the known values (e) in the locations s_i, and where λ_i are the kriging weights. These weights are obtained by minimizing the variance of the estimation error and considering the structure of spatial autocorrelation detected with the variogram (Matheron, 1965). The advantage to this method is that it generally provides weights that are greater than the residuals located closest to the location s_0 and smaller weights for those furthest away from the location.

In our study, expression (2) allows estimating the response of the binary variable ACS by the sum of $\widehat{m}\,(s_0)$, which depends on the distance between the house and the school, and $\widehat{e}\,(s_0)$, which considers the spatial autocorrelation structure of the residuals. Hence, we assume that unobserved factors are utilized to improve the prediction by considering the spatial autocorrelation of residuals, instead of assuming (as in classic logistic regression) that they are white noise (Zhang and Wang, 2014). Expression (2) is an estimator which has the advantage of allowing estimations at any point of the plane.

5. Results and discussion

The frequency and percentage of students who travel to school on foot or by motorized modes of transportation are presented in Table 1. Alhambra High School was found to have the highest percentage of students who walk to school (84%), while Ganivet High School had the lowest (57%). The descriptive characteristics of the four types of distances used to measure the proximity between home and school are also shown in Table 1. As can be seen, the mean Euclidean distance is always the shortest distance for all schools, while the mean driving distance is always the longest. These results are in line with other studies that show that the Euclidean distance is shorter than the Manhattan distance and the network distance (Arafat and Abed Al Musa, 2017). However, an interesting result of our work is that the mean walking distance and the mean Manhattan distance are similar. There is not enough evidence to reject our null hypothesis that the means of the walking distance and the Manhattan distance are equal, since the p-value was greater than 0.05 for three schools (Alhambra, Soto de Rojas and Zaidín) and greater than 0.01 for one (Ganivet). By contrast, there is evidence to reject the null hypothesis that the differences between the means of the other types of distances are equal, since the p-value was always less than 0.01. Therefore, there are no significant differences between the mean Manhattan distance and the mean walking distance, but there are significant differences between the other distances. This has practical implications because it is much more cumbersome to obtain walking distances than Manhattan distances.

The location of the students' homes and their mode of commuting to school are shown in Fig. 2. There is a large concentration of households around the Alhambra, Soto de Rojas and Zaidín high schools, while the location of the households of students attending Ganivet High School was more dispersed. Actually, the coefficient of variation is similar to 1 for the three first high schools and 0.59 for Ganivet High School, thus showing an almost double dispersion in this school (see Table 1). In addition, this school has the longest average distance between households (1722.4 m), which also affects the rates of spatial dependence. Hence, the students at Ganivet High School have lower spatial dependence rates as indicated by the high value of the nugget effect (r-nug) for this school (see Table 2).

As can be seen in Fig. 2, the Alhambra, Soto de Rojas and Zaidín high schools are located in the southern part of the city. This is a transition zone, with median housing prices and whose inhabitants are of middle socioeconomic status. However, Ganivel High School is located in the downtown area of the city, where housing prices are high and the socioeconomic status is generally higher, which could be

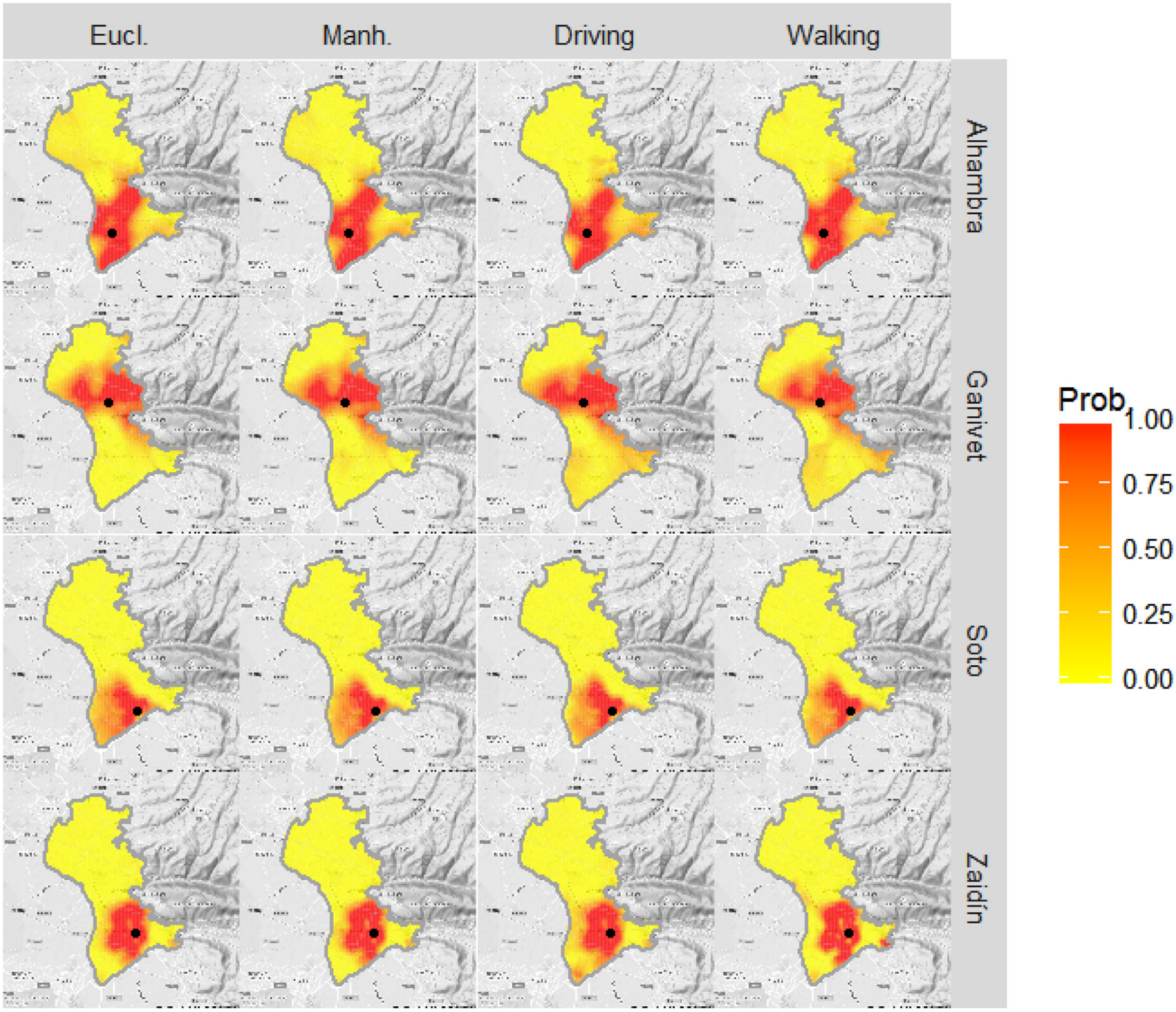

Fig. 5. Isoprobability maps of walking to school for the four types of distances.

related to the more frequent use of motorized vehicles (see Table 1).

In this study, 16 logistic regression models were obtained for each of the four high schools, and four types of distances were performed (Table 2**)**. In all cases, the coefficients of the models are significant and have a negative sign, thus indicating that as the distance increases, the likelihood of ACS to school decreases. This result coincides with most of the cases in the study of Pont et al. (2009). In order to compare the different kinds of distances (for the same school) we have used the Nagelkerke adjusted R-squared and the Akaike Information Criterion (AIC), which describe a trade-off between the goodness-of-fit and its complicity (Burnham and Anderson, 2004). According to these statistics, the type of distance that properly explains the transport mode choice at the Alhambra and Soto de Rojas schools is the Euclidean distance as the statistical value of AIC is lower. However, the walking distance shows the best results for the Ganivet and Zaidín schools. Therefore, the Euclidean and walking distances best express the concept of proximity between home and school for schools in these types of environment.

It is also possible to calculate the AIC difference = $\Delta_i = AIC_i - AIC_{\min}$ (where $AIC_{\min}$ is the minimum value). A rule of thumb is when $\Delta_i > 10$, the model fails to explain some substantial variation in the data and the best model is when $\Delta_i = 0$ (Burnham and Anderson, 2003). In this study, none of the distances considered in the Alhambra and Zaidín high schools meet the stated rule, so all distances substantially explain the variation in the data. On the other hand, the distance by car for the Ganivet and Soto high-schools fulfills this rule of thumb. Therefore, the level of empirical support of this distance to explain ACS in these schools is basically low or none (Burnham and Anderson, 2003). The low values of Δ_i for the Alhambra and Zaidín high schools are due to the strong correlation between the different types of distances (see Table 1).

To detect the presence of the contagion effect of neighbors and the locational variables, we have used the experimental variograms of residual $e(s_i)$ and its fitted models, which is expressed in Fig. 4. A spherical model was adjusted to all the experimental variograms with the exception of Alhambra High School, for which a Gaussian model was fitted. The spherical model is linear for small distances, while the Gaussian model represents an extremely continuous phenomenon (Armstrong, 1998). Therefore, the residuals of Alhambra High School present greater spatial continuity, that is, small changes in the distance between houses (h) cause small changes in the variance of the residuals. For the other schools, however, small changes in the distance between houses lead to major changes in variability. In other words, the contagion effect of the neighbors and the locational variables is less likely to vary in space for the students attending Alhambra High School compared to the other schools. However, this contagion effect and the locational variables vary substantially in the space for the students attending Ganivet High School.

The range of the variogram represents the distance at which the residuals are correlated, that is, the range of spatial dependence. This

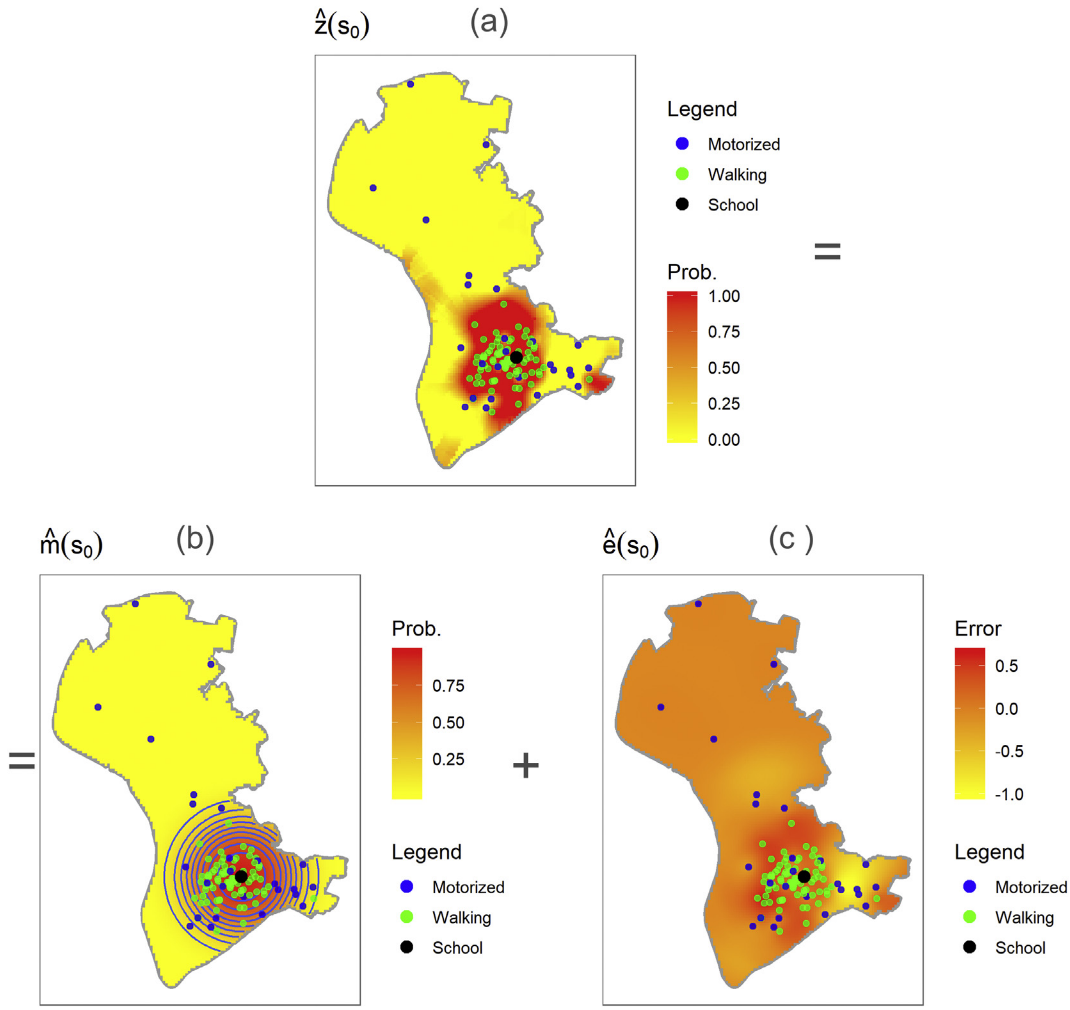

Fig. 6. (a) Isoprobability map of walking to Zaidín High School, (b) isoprobability map considering only the distance to the school and (c) isovalue map of residuals.

means that students living within that range are affected by similar locational variables and face-to-face contagion and make similar decisions. However, the residuals associated to the students living at a further distance from this range are not correlated: when the variogram reaches the sill, the covariance is canceled (Cressie, 1991). In the present study the range is between 1025 (Ganivet High School) and 1630 m (Soto de Rojas High School). Thus, the effect of spatial dependence between households has a radius of influence of about 1000–1600 m. This radius is quite similar to that obtained in other studies (Broberg et al., 2013). This spatial dependence may be due to both local or face-to-face contagion effects or caused by unobserved locational variables (crime, traffic, the environment, urban design, socioeconomic characteristics, etc.) that have similar effects on households located in the same zone.

The ratio of the nugget effect in percentages (%) was also obtained to determine the degree of spatial dependence (r-nug, Table 2). Spatial dependence was found to be weak for Ganivet High School, moderate for Soto de Rojas High School and strong for the Alhambra and Zaidín schools.

It is important to note that a lower percentage of students at Ganivet High School walk to school. The school also shows higher average distances to school, as well as a lower degree of spatial dependence (higher r-nug). In contrast, Alhambra is the high school with the highest percentage of students who walk. Moreover, although the mean distance from home to school is not the lowest (due to its high statistical dispersion), Alhambra is one of the schools with a greater degree of spatial dependence because the relationship between the nugget and sill is lower (lower r-nug). Moreover, the variograms of Alhambra High School (Fig. 4) showed that students have a low elasticity to distance between households that are close to each other (0–500 m), since there is little variability between the residuals of the houses.

In line with the second aim of the study, isoprobability maps were created to determine the likelihood that a child living anywhere in the city will actively commute to school (Fig. 5). These maps represent the areas of influence of ACS for four high schools in the city of Granada taking into account the structure of variability detected by the variograms. The estimated probabilities reflect the areas of influence of each school for each type of distance and have been obtained with RK (expression (2)). The red areas indicate the areas with the highest probability that a child will walk to school. These areas have been calculated taking into account the distance from home to school and the spatial dependence detected with variograms in the unobservable locational variables ($e(s_i)$). The distances that students in these schools will walk delimit the areas of influence. These areas are not circular in shape because the students' travel mode choices are not affected only by the distance to school, but by other unobservable locational factors that can be obtained for any point level $s_0$by the kriging method ($\hat{e}(s_0)$).

To clearly illustrate the results of the RK method, Zaidín High School has been taken as an example. Fig. 6 shows: (a) the map of ACS isoprobabilities considering the Euclidean distance and the presence of autocorrelation between the errors ($\hat{z}(s_0)$); (b) the isoprobabilities map obtained by classic logistic regression considering only the distance from home to school ($\widehat{m}(s_0)$) and (c) the map of isovalues of residuals ($\hat{e}(s_0)$).

As can be observed, the final result for ACS ($\widehat{z}(s_0)$) is affected not only by the distance from home to school ($\widehat{m}(s_0)$), but also by the contagion effect and other locational variables that have not been included in the model ($\widehat{e}(s_0)$). It is clearly observed that if only distance is considered, the probability of walking to school would have a homogeneous spatial distribution in the form of concentric circles around the school. However, the presence of autocorrelation among the residuals caused by other locational factors deforms these concentric circles, resulting in an irregular isoprobabilities map.

5.1. Study limitations

One of the main limitations of this work is that only one locational variable (distance) has been used to explain the ACS model. The inclusion of other locational variables (crime, traffic, the environment, urban design, socioeconomic characteristics, etc.) would have improved the overall understanding of the ACS model, but were not measured because it was not on the aims of the study. Moreover, these variables are often unavailable or difficult or costly to measure. However, we consider that the omission of these variables does not invalidate the results obtained in this work since the global effect of these factors on ACS has been considered indirectly by taking into account the presence of spatial autocorrelation in the model perturbations caused by omitting these locational variables. Other limitations include the convenience sampling of the participants: the sample is limited to Spanish youth from southeastern Spain. Nonetheless, the sample size is quite large. The main strengths in the current study was the use of a valid questionnaire to measure children's commuting to school. An additional strength of the work was the objective measurement of the area of influence of each school.

6. Conclusions and policy implications

This study has analyzed the spatial distribution of households with students who travel to four high schools located in Granada, Spain. Four different distances were examined to explain the effect of proximity between the homes and the schools and measure the effect of each distance and the presence of spatial dependence on the decision to walk to school. The analysis has permitted us to determine the area of influence of each school.

The Euclidean distance has traditionally been considered to explain the mode of commuting (Pont et al., 2009). The results we have obtained indicate that the two types of distances that best explain the decision to walk to school are the Euclidean distance and the walking distance. Moreover, the presence of spatial dependence effects arising from interactions among geographically close households has also been observed. In general, the spatial dependence produced by locational variables and spatial contagion by the students is moderate to strong, except for Ganivet High School, for which it is weak. In addition, the spatial range of the locational variables and the contagion effect is approximately 1000 m to 1600 m. These findings have policy implications since the transport mode choice of the students is influenced by locational variables and neighbors' decisions. They also highlight the need for public institutions to improve the surrounding environment of the schools in order to increase their walkable area of influence to promote active commuting to school reiterating its social benefits and promote a “domino effect” regarding individuals' decisions. These interventions would decrease negative behavoir patterns (i.e., passive commuting to school). In addiction, identifying these walkable areas may be of interest for local services such as restaurants, supermarkets, shops, etc. It is our hope that this study will provide some interesting insights.

Recent developments in GIS have opened up new opportunities for the use of spatial methodologies (Broberg and Sarjala, 2015). RK could be implemented in a GIS to allow state agencies and municipalities to determine the areas of influence of schools and aid urban designers and planners in developing neighborhoods that support ACS in order to feed into policy decisions.

Acknowledgements

We would like to thank the school officials, students and their teachers and parents for their unconditional participation in this study and without whom this project would not have been possible. We also acknowledge all those who helped with the data collection and all the members involved in the field work for their effort, enthusiasm and support.

Funding

This research did not receive any specific grants from funding agencies in the public, commercial or not-for-profit sectors.

References

Anselin, L., 1988. Spatial Econometrics: Methods and Models. Kluwer Academic Publishers, Dordrecht.
Anselin, L., 1999. Spatial Econometrics. CSISS. Spatially Integrated Social Science. pp. 1–31.
Anselin, L., Bera, A.K., 1998. Spatial dependence in linear regression models with an introduction to spatial econometrics. In: Ullah, A., Giles, D. (Eds.), Handbook of Applied Economic Statistics. vol. 155. Marcel Dekker, pp. 237–290.
Anselin, L., Rey, S., 1991. Properties of tests for spatial dependence in linear regression models. Geogr. Anal. 23 (2), 112–131.
Apparicio, P., Shearmur, R., Brochu, M., Dussault, G., 2003. The measure of distance in a social science policy context: advantages and costs of using network distances in eight Canadian metropolitan areas. J. Geogr. Inf. Decis. Anal. 7, 105–131.
Apparicio, P., Abdelmajid, M., Riva, M., Shearmur, R., 2008. Comparing alternative approaches to measuring the geographical accessibility of urban health services: distance types and aggregation-error issues. Int. J. Health Geogr. 7, 1.
Apparicio, P., Gelb, J., Dubé, A.-S., Kingham, S., Gauvin, L., Robitaille, É., 2017. The approaches to measuring the potential spatial access to urban health services revisited: distance types and aggregation-error issues. Int. J. Health Geogr. 16 (1), 32.
Arafat, A.N., Abed Al Musa, M., 2017. Evaluating sustainable housing locations: comparing the use of network to euclidean proximity in the scoring process. Int. J. Plan. Urban Sustain. Dev. 5 (2), 52–62.
Armstrong, M., 1998. Basic Linear Geostatistics. Springer Science & Business Media.
Beck, L.F., Greenspan, A.I., 2008. Why don't more children walk to school? J. Saf. Res. 39, 449–452.
Boix, R., Trullén, J., 2007. Knowledge, networks of cities and growth in regional urban systems*. Pap. Reg. Sci. 86, 551–574.
Broberg, A., Sarjala, S., 2015. School travel mode choice and the characteristics of the urban built environment: the case of Helsinki, Finland. Transp. Policy 37, 1–10.
Broberg, A., Salminen, S., Kyttä, M., 2013. Physical environmental characteristics promoting independent and active transport to children's meaningful places. Appl. Geogr. 38, 43–52.
Burke, M., Brown, A.L., 2007. Distances people walk for transport. Road Transp. Res. 16, 16–29.
Burnham, K.P., Anderson, D.R., 2003. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. Springer Science & Business Media.
Burnham, K.P., Anderson, D.R., 2004. Multimodel inference understanding AIC and BIC in model selection. Sociol. Methods Res. 33, 261–304.
Chica-Olmo, J., 1995. Spatial estimation of housing prices and locational rents. Urban Stud. 32, 1331–1344.
Chillon, P., Ortega, F.B., Ruiz, J.R., De Bourdeaudhuij, I., Martinez-Gomez, D., Vicente-Rodriguez, G., Widhalm, K., Molnar, D., Gottrand, F., Gonzalez-Gross, M., Ward, D.S., Moreno, L.A., Castillo, M.J., Sjostrom, M., 2011. Active commuting and physical activity in adolescents from Europe: results from the HELENA study. Pediatr. Exerc. Sci. 23, 207–217.
Chillón, P., Hales, D., Vaughn, A., Gizlice, Z., Ni, A., Ward, D.S., 2014. A cross-sectional study of demographic, environmental and parental barriers to active school travel among children in the United States. Int. J. Behav. Nutr. Phys. Act. 11, 1.
Chillón, P., Panter, J., Corder, K., Jones, A., Van Sluijs, E.M., 2015. A longitudinal study of the distance that young people walk to school. Health Place 31, 133–137.
Chillón, P., Herrador-Colmenero, P., Hidalgo-Migueles, J., Cabanas-Sánchez, V., Fernández-Santos Jorge, R., Veiga, O.L., Castro-Piñero, J., 2017. Scandinavian. J. Public Health 45 (6), 612–620.
Crane, R., Crepeau, R., 1998. Does neighborhood design influence travel?: a behavioral analysis of travel diary and GIS data. Transp. Res. Part D: Transp. Environ. 3, 225–238.
Cressie, N., 1991. Statistics for Spatial Data. John Wiley & Sons, USA.
Davison, K.K., Werder, J.L., Lawson, C.T., 2008. Children's active commuting to school: current knowledge and future directions. Prev. Chronic Dis. 5, A100.
de Hartog, J.J., Boogaard, H., Nijland, H., Hoek, G., 2010. Do the health benefits of cycling outweigh the risks? Environ. Health Perspect. 118, 1109–1116.
Dubin, R.A., 1988. Estimation of regression coefficients in the presence of spatially