CrossMark

# Is accessibility relevant in trip generation? Modelling the interaction between trip generation and accessibility taking into account spatial effects

Ruben Cordera[1] · Pierluigi Coppola[2] · Luigi dell'Olio[1] · Ángel Ibeas[1]

**Abstract** The influence of accessibility to opportunities in trip generation continues to be debated in the specialised literature given its relevance to simulate phenomena such as induced demand. This article estimates multiple linear regression models (MLR), spatial autoregressive models (SAR), spatial autoregressive models in the error term (SEM) and spatially filtered Poisson regression models (SPO) to discover whether or not accessibility is a significant factor in trip generation using data from the urban area of Santander (Spain). The results obtained provide evidence which shows that, on an intraurban scale, more accessibility to opportunities decreases trip production in private vehicle for work purpose, whereas it increases trip production in other transport modes for non—mandatory purposes. For the correct interpretation of the estimated parameters it was important to consider the direct and indirect effects of the independent variables in the SAR production models. Finally, the validation of the models showed that the SAR and SEM models had a mean squared error slightly lower than the MLR models in predicting overall trip production. This was because the spatial models reduced the correlation of the residuals present in the MLR models. Furthermore, the SPO models performed better in validation mode than all the continuous models.

**Keywords** Accessibility · Trip generation · Induced demand · Spatial—econometric models

✉ Ruben Cordera
corderar@unican.es

[1] Department of Transportation, University of Cantabria, Av. de Los Castros s/n, 39005 Santander, Cantabria, Spain

[2] Department of Civil Engineering, "Tor Vergata" University of Rome, Via del Politecnico 1, 00133 Rome, Italy

## Introduction and objectives

Trip generation models are the first and most important phase in traditional transport modelling given that they provide the basic input for the remaining simulation stages (McNally 2008). The classic specification of four stage models starts from the assumption that changes in the distribution of activities and changes in the transport network do not affect the number of trips being generated, a hypothesis which is difficult to justify over the medium to long term (Martínez 2000; Ortúzar and Willumsen 2001).

This article proposes the inclusion of accessibility indicators in the trip generation models to try to minimize this problem by making them sensitive to changes which may occur within the system of activities and the transport network. The theoretical relationship between accessibility and trip generation is supported by two basic theories (Thill and Kim 2005). Firstly, the act of making a journey is not generally an end in itself but more of a means to perform other activities. Secondly, the fact that trip generation depends on the services provided by the transport infrastructure, in other words, on the existence of the phenomena known as induced demand (Cervero 2002; Goodwin 1996; Hymel et al. 2010). The first supposition suggests that a greater presence of opportunities should lead to a greater disposition to make journeys. The latter implies that lower journey costs will create greater demand for transport both in terms of the number of journeys being made as well as their length. Furthermore, greater accessibility to opportunities could also suppose a reduction in journeys made by private vehicle (PV), an effect detected in previous research (Cervero and Kockelman 1997). Therefore, in a determined zone, changes made both to the location of activities and to the network should have an effect on trip generation if they coincide with any change in opportunities or travelling costs. If the trip generation models are not sensitive to these changes, then the transport models could provide biased predictions for example, about the resulting demand from the construction of a new transport infrastructure.

Nevertheless, in spite of its theoretical importance, the inclusion of accessibility indicators in trip generation models has proven to be problematic. In some estimations the indicators that were used have turned out not to be significant or have produced an unexpected sign which is why some authors have recommended against using it in practice (Ewing et al. 1996; Hanson and Schwab 1987).

This article proposes a new evaluation of the explanatory capacity of the accessibility variable in trip generation models controlled by the spatial effects which may occur. Accessibility will be measured by three indicators: a gravity accessibility indicator from each of the zones to employment, a gravity accessibility indicator from each of the zones to education places and an indicator of the journey time to the city centre. The possible presence of spatial autocorrelation in the error term has been scarcely addressed by traditional trip generation linear regression models in spite of the possible bias and inefficiencies in the estimated parameters. Spatial autocorrelation could occur due to social effects resulting from interaction between households creating different patterns of trip generation in the study area. This is a hypothesis which links in with the growing interest around the influence of social factors in transport choices (Carrasco and Farber 2014; Scott et al. 2013). Furthermore, the lack of any relevant variable in the model which affects the population in a spatially differentiated way could also cause spatial autocorrelation in the residuals of the models. It is therefore important to control the presence of these spatial effects for the correct estimation of the models' parameters.

The proposed models have been estimated using disaggregated household data. Multiple linear regression models (MLR), spatial autoregressive models (SAR), spatial autoregressive models in the error term (SEM) and spatially filtered Poisson regression models (SPO) have been estimated. The four types of models will be compared as a function of the theoretical coherence and statistical significance of the estimated parameters as well as their goodness of fit. Finally, the models will be validated using a K—Fold type technique, which is a recommended method for testing the predictive capacity of models, previously not applied in this line of research (James et al. 2014).

The obtained results show that the household trip production for work and other purposes was sensitive to the expressed accessibility conditions. However, whereas greater accessibility supposed a reduction in PV trips for work purpose, it also supposed an increase in the number of trips made using other transport modes (NPV) for non—mandatory purposes. Trips made for study purposes showed that they could decrease in zones providing greater accessibility using PV, although the effect was only significant using the journey time to the city centre as an indicator. These results provide evidence supporting the idea that accessibility is a relevant factor in modelling discretionary trip production using modes other than the car (greater accessibility coincides with greater household trip production, ceteris paribus) and that it could be a factor in the reduction of PV trip generation, especially for work purposes.

Furthermore, the models which consider the presence of spatial dependence between observations, SAR, SEM and SPO showed that there was a degree of significant spatial dependence in trip production. It was important to consider the presence of spatial dependence between observations, especially for the direct and indirect effects of the variables in the SAR production models in order to correctly interpret the estimated parameters.

Finally, the validation of the models showed that the SAR and SEM had a mean squared error which was slightly lower than the MLR models in predicting overall trip production. This was because the MLR models showed a degree of significant spatial correlation with positive residuals concentrated in the periphery of the study area for trips made using PV and concentrated in the centre for trips made for other purposes using NPV. This effect was minimized by most of the spatial models until it lost its significance. This could be due to the interaction between households in different zones in the study area or alternatively, to the omission of a spatial variable with a different effect between the centre and the periphery. Furthermore, the Poisson regression models behaved better for both, the work journeys and those made for other purposes, underlining the importance of considering the truncated and discrete nature of trip generation.

The article will be structured in the following way. "State of the Art" section will provide a review of the state of the art in work relating to the inclusion of accessibility as a variable and the consideration of spatial effects in trip generation models. The methodology and the study area used for estimating the models will be presented in "Methodology and Application" section. This will be followed by the presentation of the results in "Results and discussion" section and finally "Conclusions" section will cover the main conclusions that have been drawn.

## State of the art

The relationship between accessibility and trip generation has still not yet been satisfactorily clarified in the specialised literature. Some authors have found that this relationship is positive and significant whereas others have found that there is no relationship whatsoever between accessibility and trip generation.

One of the first studies was by Vickerman (1974), who tested the relationship between trip generation and accessibility using data from Oxford (United Kingdom). The author found evidence that variations in accessibility were positively linked to trip generation. Vickerman, however, also underlined the difficulty of estimating the model given the high level of correlation between the independent variables, a factor which the author believed was the main difficulty when trying to differentiate between the effects accessibility and other factors have on trip generation.

Later, Koenig (1980) studied the relationship between trip generation and accessibility using data from five French cities. The author used a gravity type indicator of accessibility and correlated it with trip generation rates controlled by car possession, age group and journey purpose. The relationship was positive in all the groups leading the author to conclude that accessibility was a strong determining factor in trip generation.

Hanson and Schwab (1987) examined the relationship between accessibility and different journey choices, one of which was frequency, using data from the town of Uppsala (Sweden). The work was controlled by various sociodemographic characteristics and the authors concluded that, differently from Koenig (1980), the relationship between accessibility and trip generation was weak and lower than expected by theoretical hypothesis. This led the authors to state that on an intraurban scale the level of accessibility does not need to be incorporated into trip generation models. A similar result was obtained by Ewing et al. (1996) who tested the impact of accessibility and other land use variables on trip rates using data from Florida (USA). The results showed that density, land use diversity and accessibility did not have significant effects on trip generation so the authors claimed that the conventional trip generation models without regard to location were good enough. Negative results were also obtained by Kitamura et al. (2001) who examined data from the Kyoto-Osaka-Kobe metropolitan area (Japan) and southern California (USA) in order to test the effect of accessibility on trip behaviour. The authors concluded that accessibility did not affect automobile use in study areas characterized by high industrial and urban development.

Thill and Kim (2005) estimated multiple regressions to predict the generation of private vehicle journeys with data from Minneapolis—St. Paul (USA). The authors used two types of indicators of accessibility: gravity and accumulated opportunities, as well as different manually calibrated impedance parameters. The researchers concluded that both, the trip production and attraction models, are significantly affected by accessibility even though the variable sign was positive or negative depending on the model.

Another line of research has focused on the role of other land use characteristics in trip generation. Cervero and Kockelman (1997) studied the importance of the build environment in PV and NPV trip generation considering three dimensions: density, diversity and design, using data from the San Francisco Bay Area (USA). The authors considered the accessibility to jobs and services as part of the density concept. This research found that the three land use factors (including accessibility) can reduce PV trip rates encouraging the use of more sustainable transport modes. Cervero et al. (2009) examined how different facilities and built environment characteristics: density, diversity, accessibility and

proximity to transit influenced walking and cycling behaviour in Bogotá (Colombia). The results showed that factors such as connectivity and proximity to cycle lanes were associated with more physical activity whereas other factors such as density and diversity were not. According to the authors, these results showed that given the high accessibility of most neighbourhoods in Bogotá only the design factors were relevant. Purvis et al. (1996) estimated non-work trip generation models incorporating the effect of work trip duration as a measure of accessibility using data from the San Francisco Bay Area (USA). The authors found that decreases in the work trip duration were correlated with increases in trip generation for shopping and other purposes. In the same study area, Wu et al. (2012) compared the auto trip rates predictions of an activity—based model with and without considering accessibility. The authors applied the model to the San Francisco urban area (USA) and the results allowed the authors to conclude that if trip generation had been calculated without considering accessibility, the model would have significantly overestimated the number of trips made by car in denser neighbourhoods.

Other authors like Walters et al. (2013), Clifton et al. (2015) and Millard-Ball (2015) have focused on methods of adjustment of vehicle trip generation estimates from the Institute of Transportation Engineer's (ITE) Trip Generation Manual (Institute of Transportation Engineers 2012). In general, the authors have showed that the ITE trip rates overestimate private vehicle trips in significant percentages such as 55 % in the case of Millard-Ball (2015) and 49 % for AM peak traffic in Walters et al. (2013). This is because they were often based on surveys conducted before the 1960s in areas with low accessibility to opportunities, without mixed-use and where there was less congestion and fewer transport alternatives. This line of research has highlighted the importance of taking into account the built environment and the accessibility of every place in order to estimate more accurate trip generation rates.

MLR is among the more commonly used methods for estimating the future evolution of trip generation. However, MLR presents certain problems associated with modelling trip generation as a continuous variable and not considering the possible presence of spatial autocorrelation in the residuals of the model. Given that the data used in trip generation models contain a strong spatial component, independently of their aggregated or disaggregated nature, specific econometric techniques need to be applied to minimize the problem. If not, then the estimated parameters could show bias and may provide erroneous inferences about the influence of variables such as accessibility.

Spatial dependence could be defined simply as the impact of the dependent variable or the independent variables of neighbouring areas, on the dependent variable of a determined area (Anselin 1988). In order to define the relationship between areas the analyst needs to specify a neighbourhood matrix. Gamas et al. (2006) estimated trip generation models for Mexico City (Mexico) considering the trips generated by unit area and the presence of spatial effects. After finding that several of their variables were autocorrelated, the authors applied a SAR model and thereby avoided the estimation of biased parameters resulting from conventional MLR models (LeSage and Pace 2009).

Kwigizile and Teng (2009) estimated trip production and attraction models using data from Las Vegas Valley (USA). Initially, the authors specified and estimated trip production and attraction models using MLR. After performing a Moran I test on the models' variables, the authors detected the presence residual autocorrelation of the trip attraction model but not in the trip production model. This led the authors to estimate a trip attraction model considering the presence of spatial dependence between observations. Comparing the fit between the spatial and non-spatial models, the former showed an average deviation from the data of 17 % faced with 31 % for the latter.

Finally, Kim et al. (2009) detected the presence of spatial effects in the non-household based (NHB) trip generation models in the metropolitan area of Daegu (South Korea). The NHB trip production and attraction models showed spatial correlation in the error term leading the authors to estimate SEM models. The NHB attraction model also showed spatial correlation. The authors concluded that the spatial econometric models were a useful tool for making estimations about trip generation in study areas in which there are agglomeration economics i.e. positive externalities derived from proximity among firms and people, that induce more travel demand (Glaeser 2010).

Although the authors in the three studies mentioned above showed that the models considering the presence of spatial correlation in the residuals had a better fit to the data, they did not validate the models using different data from that used in the estimation phase. However, Lopes et al. (2014) did validate trip generation models estimated using data from 1974 with an origin–destination survey in 2003 for the city of Porto Alegre (Brazil). The authors found that the spatial models made better predictions for the 2003 journeys than the traditional linear regression models, although in both cases the predictions were not good enough due to the huge changes experienced by the city over 30 years. In any case, the predictive capacity of these models compared to the traditional MLR models is yet to be completely clarified.

Furthermore, additional evidence is required to check if the theoretical hypothesis that accessibility affects trip generation is correct addressing also for spatial effects between observations usually present in cross-sectional data. These spatial effects in trip production may be caused by various factors. Firstly, by the spatial dependence between nearby households due to mutual social influence. Recent research has placed more emphasis on the importance of considering social factors together with traditional factors when explaining transport choices. Social norms (Bamberg et al. 2007), transport habits (Minnen et al. 2015), social networks (Sherwin et al. 2014) or social cohesion (Clark and Scott 2013) may implicate spatial dependencies in the behaviour of nearby households which generate a greater or lesser trip production than households in other areas. Another factor to consider is the inability to measure, ore precisely measure, variables that have a differential effect in space. This specification error may also imply autocorrelation in the residuals of the MLR models with the estimation of biased or inefficient parameters. Both effects can be captured with the support of models which explicitly consider situations of spatial dependence, such as those presented in this study.

## Methodology and application

This section presents the theoretical formulation of the trip generation models incorporating the accessibility factor and the presence of spatial autocorrelation in the residuals of the models. The available data for estimating the trip generation models are also presented here.

### Modelling trip generation considering accessibility to opportunities

In trip generation modelling, a distinction must be made between trip production and trip attraction models. Trip production models aim to predict the home-based journeys (HB), where the origin or destination is the home, and the origin of non-home-based trips (NHB).

Trip attraction models instead includes the trips which end is different from the home in HB journeys, and the destination of NHB journeys (Ortúzar and Willumsen 2001).

Trip generation models can be classified by purpose. The three most commonly used categories include: journeys for work, journeys for education and journeys for other purposes. This latter class can be further divided into multiple sub-classes depending on the requirements of the model (shopping, leisure, etc.). The estimation can be carried out using disaggregated household data in the case of trip production or using aggregated zonal data in modelling trip attraction. The conventional form of trip generation models corresponds to a multiple linear regression:

$$y = X\beta + \varepsilon \tag{1}$$

where y is a vector with information about the dependent variable (produced or attracted trips), $\beta$ is a vector of estimated parameters and $\varepsilon$ is a vector of independent and identically distributed (IID) errors. The matrix X contains information about the independent variables.

Some authors have proposed the use of models which consider the truncated and discrete nature of trip production (Chang et al. 2014; Lim and Srinivasan 2011). Among the available alternatives is the non-linear Poisson regression model with probability given by (Greene 2003):

$$P(y_i) = \frac{\lambda_i^{y_i} e^{-\lambda_i}}{y_i!} \tag{2}$$

The Poisson regression assumes that each dependent variable $y_i$ is extracted from a Poisson type discrete distribution with parameter $\lambda_i$, which commonly is logarithmically linked to a linear combination of independent variables:

$$\ln(\lambda_i) = x_i' \beta \tag{3}$$

However, this model has the limitation of considering that the mean and variance conditioned by the Poisson distribution are equal, which could be problematic if overdispersion in $y_i$ is detected. This hypothesis can be relaxed by estimating an additional dispersion parameter, leading to a quasi-Poisson regression model (Zeileis et al. 2008).

Among the most commonly used independent variables in trip generation models are those such as number of vehicles, household income, household size, land use and accessibility. Various indicators have been proposed in the literature to capture the effect of accessibility. Handy and Niemeier (1997) proposed a classification with three large types: accumulated opportunities, gravity type and based on utility. The most frequently used indicators are of the gravity type and generally take the following form:

$$A_i = \sum_j f(E_j, C_{ij}) \tag{4}$$

where $E_j$ is a measure of the attraction of zone j and $C_{ij}$ is a measure of the journey cost between zones i and j. The indicators of accumulated opportunities can be interpreted in a specific way of (4) where $C_{ij}$ is equal to 1 if the opportunities are found within a cost cut-off point defined by the analyst and 0 if not (Koenig 1980). The gravity indicators have an advantage over the above in that they can differentiate the opportunities according to journey cost without setting a binary cut-off point. The gravity indicators are also zonal,

meaning that they provide a better understanding of the aggregate relationship between transport and land use from those based on utility at the individual level (Horner 2004).

Three accessibility indicators were chosen for this work. The simplest consisted of measuring access times from the centroids to the network plus travel times along the arcs of the network to the city centre, from each of the zones in the study area. This type of indicator can be useful in monocentric urban areas where there is a high centralised concentration of opportunities. Access and travel times are not the only costs that could be considered, but they provide a good proxy to the generalized travel cost of the trip. A gravity indicator proposed by Cascetta (2009) and Coppola and Nuzzolo (2011) has also been used for urban areas with a polycentric nature. These authors differentiate between the active accessibility of a zone which can be defined as the capacity to reach the opportunities present in other zones, and passive accessibility which is defined as the ability of a zone to be reached from other zones. Active accessibility can be seen as a valid indicator to predict the number of journeys produced by a zone and is formulated in the following way:

$$ACC\_ACT_i = \sum_j \left[ \exp(\alpha_2 \cdot C_{ij}) \cdot E_j^{\alpha_1} \right] \tag{5}$$

where $E_j$ and $C_{ij}$ are the same as those present in Eq. (4) and $\alpha_1$ and $\alpha_2$ are parameters to be estimated. This indicator can be easily calibrated using ordinary least squares (OLS) transforming both sides of the expression (5) logarithmically.

## Models considering the presence of spatial dependence between observations

LeSage and Pace (2009) provide an introduction to the spatial econometric models developed in the literature. The best known spatial model is the SAR which assumes the existence of a spillover effect in the dependent variable. The model is specified as:

$$y = \rho W y + X\beta + \varepsilon \tag{6}$$

where $\rho$ is the parameter of spatial autocorrelation, W is a weighted matrix N x N where N is the number of observations and the remaining variables are equal to those present in (1).

If the only requirement is to specify the presence of spatial dependence in the error term then a SEM model can be used, as follows:

$$y = X\beta + u \tag{7}$$

$$u = \lambda W u + \varepsilon \tag{8}$$

where $\lambda$ is a parameter of autocorrelation of errors μ, and $\varepsilon$ is a vector of IID errors. So in this model the dependent variable of a location is a function of not only the independent variables but also of the μ errors of the neighbouring locations. The matrix W can be defined in different ways depending on whether zonal or point data are available. The four most common types of neighbourhood are: queen, rook, predetermined number of closest neighbours and the specification of a maximum neighbourhood distance. The queen type contiguity considers as neighbours all the adjacent locations sharing a border or a vertex with the given location, while the rook type contiguity considers as neighbours those locations that share a border with the reference location (Anselin 1988). Lesage and Pace (2010a) proposed different measurements of the correlation between neighbourhood matrices and showed how the influence of specifying W on the estimations of the

parameters is minimal if they are correctly interpreted from the true partial derivatives (direct impacts + indirect impacts) and if the model is well specified.

Griffith (2013) has proposed another method for addressing spatial autocorrelation which could be present in the residuals of the linear and non-linear models. Spatial filtering, the method based on spatial filters defined as linear combinations of the eigenvectors chosen from a neighbourhood matrix, allow this correlation to be removed. The eigenvectors are chosen in such a way that they show a spatial autocorrelation index which is higher than a critical value according to the specified neighbourhood matrix. The first eigenvector chosen $E_1$ will be the group of real numbers with the greatest index of spatial correlation according to the Moran I indicator. The vector $E_2$, will present the highest index of spatial correlation between the groups of real numbers not correlated with the $E_1$ vector, and so on. In this way these eigenvectors capture the spatial correlation present in the data acting as control variables for the model.

## Application to the urban area of santander

The MLR, SAR, SEM and SPO trip generation models were estimated using data from the city of Santander (Spain). Santander is a small—medium sized city with a population of 178,400 in its urban nucleus rising to more than 280,000 in its sphere of influence. Most of the trips in the city are made by foot (45 %) or by car (45 %), with the rest of trips made by bus (8 %) and using other modes.

The data for the models' estimation came from two sources. Firstly, a household survey asked in 2010 by the University of Cantabria to 1,000 households and, secondly, the national population and housing census (INE 2011) which provided data about the socio-demographic characteristics and housing in the city. The households were chosen randomly from a list of 73,395 households on the municipal register. The sample size was determined by supposing a 90 % confidence level and a 5 % error level. In order to avoid bias, the interviewer presented and collected the survey in person giving instructions about how to fill in the forms. If there were problems when returning to collect the survey, the interviewee was instructed to continue returning to the households up to three times in order to minimise the number of non-response. Given that the postal address was available for each of the surveyed households, their positions were located geographically onto an information system (GIS) to know the area where they were located in, among a total of 86 zones in which the city was divided. This allowed the authors to relate the household database with the census zonal data in preparation for estimating the trip production models.

A total of 28 variables were considered for using with the household and zonal data (see Tables 1, 2). Note that of the 1,000 households surveyed only 817 provided complete answers. These were the ones used to estimate the trip production models.

The variables HBW-PV, HBS-PV and HBO-PV represent the trips produced by the households: home-based work trips, home-based study trips and home-based other purposes trips, using private vehicle (car or motorcycle). The variables HBW-NPV, HBS-NPV, HBO-NPV represent the same trip purposes but in which the trips are made using non—private vehicles (mainly on foot but also using public transport). These were the dependent variables used in the trip production models.

The AGE variable represents the mean age of the household members. Given that the relationship between age and trip production is probably quadratic, the variable has also been specified squared in the models ($AGE^2$). The variable CHILDREN represents the number of children in the household under 6 years old. The variables WOMEN,

**Table 1** Descriptive statistics of the variables contained in the household survey database (N = 817 households)

| Variable | Minimum | Maximum | Mean | SD | Description |
|---|---|---|---|---|---|
| TTRIPS | 0 | 22 | 5.53 | 3.55 | Total Trips |
| HBW-PV | 0 | 12 | 0.64 | 1.40 | Home-based Work Trips by private vehicle |
| HBS-PV | 0 | 8 | 0.21 | 0.86 | Home-based Study Trips by private vehicle |
| HBO-PV | 0 | 8 | 0.55 | 1.29 | Home-based Other Trips by private vehicle |
| HBW-NPV | 0 | 6 | 0.82 | 1.46 | Home-based Work Trips by non-private vehicle |
| HBS-NPV | 0 | 12 | 0.72 | 1.63 | Home-based Study Trips by non-private vehicle |
| HBO-NPV | 0 | 12 | 2.60 | 2.56 | Home-based Other Trips by non-private vehicle |
| AGE | 15 | 95 | 50.19 | 17.60 | Households members mean age |
| AGE$^2$ | 225 | 9025 | 2828.23 | 1847.27 | AGE variable squared |
| CHILDREN | 0 | 2 | 0.02 | 0.18 | N° household children <6 years |
| WOMEN | 0 | 4 | 1.22 | 0.70 | N° Household women |
| WORKERS | 0 | 4 | 0.93 | 0.86 | N° Household workers |
| SIZE | 1 | 6 | 2.24 | 0.95 | Household size |
| D_INCOME2 | 0 | 1 | 0.29 | 0.45 | Dummy income = 2 |
| D_INCOME3 | 0 | 1 | 0.54 | 0.50 | Dummy income = 3 |
| D_INCOME4 | 0 | 1 | 0.13 | 0.34 | Dummy income = 4 |
| VEH | 0 | 4 | 0.96 | 0.77 | N° household vehicles |

**Table 2** Descriptive statistics of the variables contained in the zonal database (N = 86 zones)

| Variable | Minimum | Maximum | Mean | Std. deviation | Description |
|---|---|---|---|---|---|
| ACCW-PV | 490.56 | 535.56 | 524.56 | 7.80 | Active Accessibility to work places using PV |
| ACCS-PV | 105.04 | 118.70 | 115.05 | 2.43 | Active Accessibility to education places using PV |
| ACCW-NPV | 181.22 | 1493.72 | 1042.67 | 304.89 | Active Accessibility to work places using NPV |
| ACCS-NPV | 33.68 | 213.74 | 140.77 | 45.95 | Active Accessibility to education places using NPV |
| CBD-PV | 0 | 18.57 | 13.24 | 1.79 | Travel time to CBD using PV (min) |
| CBD-NPV | 0 | 90.55 | 29.51 | 17.44 | Travel time to CBD using NPV (min) |
| HOUSEHOLDS | 124 | 2104 | 727.28 | 430.25 | N° Households |
| EMP | 37 | 4604 | 778.38 | 958.27 | N° work places |
| EDU | 0 | 12857 | 503.45 | 1458.19 | N° Education places |
| POP | 458 | 5904 | 2074.38 | 1217.57 | Population |
| POP_DEN | 1.54 | 891.81 | 210.37 | 208.58 | Population density (POP/ha) |

WORKERS and SIZE represent, respectively, the number of women, workers and the total number of people in the household.

The income variable referring to the overall monthly household income can be within the following ranges: (1) income less than 600 euros (€), (2) income from 600 to 1200 €, (3) income from 1200 to 2500 € and (4) income over 2500 €. To avoid non—linearity problems, the income variable has been transform in three dummy variables: D_INCOME2, D_INCOME3 and D_INCOME4. As each household can only belong to one of the intervals, the corresponding dummy variable takes a value of 1 in that class and 0 in the others. If the household belongs to the interval (1), the three dummy variables take the value 0.

The accessibility indicators: ACCW-PV, ACCS-PV, ACCW-NPV and ACCS-NPV were calculated from expression (5). These four accessibility indicators represent, respectively: accessibility to work by PV, accessibility to places of education by PV, accessibility to work by NPV and accessibility to places of education by NPV. The dependent variables used in their calibration were: the trips by PV for work and other purposes, the PV trips for study purpose, the NPV trips for work and others purposes and the NPV trips for study purpose. Given that the accessibility indicators try to evaluate the interactive potential of an area, the real trip distribution provides the most suitable data for their calibration. The number of jobs and the education places in zone j (see variables EMP and EDU in Table 2) were introduced into ACCW and ACCS, respectively, as variables of opportunities. The places of education have covered all levels: primary, secondary and university. Accessibility to employment is considered to be a good indicator of both work related and other purposes trips, while places of education are a better accessibility indicator of trips made for study purpose. The cost matrix used corresponded to the sum of access times, from each centroid to the transport network, and travel times at free flow by car for the PV accessibility indicators. The composite cost (Ortúzar and Willumsen 2001) of access plus travel times by foot and bus modes was used for the NPV accessibility indicators.

As can be seen in Table 3 the estimations of the parameters gave fits, according to $R^2$ of around 0.6 except for the case of accessibility to education places by PV where the fit was poorer. The parameters had the expected signs (positive for opportunities and negative for travel times) where the impedance parameter is somewhat greater in the case of accessibility to education places and using NPV which is what was to be expected.

The CBD-PV and CBD-NPV variables can also be interpreted as measures of accessibility but in this case only for the opportunities present in the city's urban centre. They

**Table 3** Estimation of the parameters of the accessibility indicators

| Variable | Accessibility Employment (PV) | | Accessibility Education places (PV) | | Accessibility Employment (NPV) | | Accessibility Education places (NPV) | |
|---|---|---|---|---|---|---|---|---|
| | α | p value | α | p value– | α | p value | α | p value |
| Ln_E | 0.330 | 0.000 | 0.185 | 0.000 | 0.596 | 0.000 | 0.472 | 0.000 |
| Cost | −0.019 | 0.000 | −0.026 | 0.000 | −0.041 | 0.000 | −0.061 | 0.000 |
| $R^2$ | 0.66 | | 0.32 | | 0.73 | | 0.59 | |

represent the access and travel time, in min, taken by a private car at free flow and the access and travel composite cost, in min, of foot and bus modes respectively. The city of Santander has a marked monocentric character, meaning the variables of accessibility and CBD show an important negative correlation with $R^2$ below $-0.7$ in all the indicators (see the times to CBD in Fig. 1). The employment density (see Fig. 2) also highlights this monocentric nature of the city with a great many jobs concentrated along this central axis.

The meanings of the remaining variables: HOUSEHOLDS, EMP, EDU, POP and POP_DEN are described in Table 2.

## Results and discussion

### Trip production for work purpose without and considering spatial dependence between observations

Table 4 present the parameters for HBW trip production models without and considering spatial dependence between observations. The models were estimated using OLS (MLR), maximum likelihood (SAR and SEM) and quasi-maximum likelihood (SPO). The specifications of the MLR models include all the variables present in the data base except for HOUSEHOLDS and POP_DEN as they showed a high correlation with the variable POP. In addition, all the models are shown only with the ACC variable in the specification given that this and CBD could not be used simultaneously because of their strong correlation. Only the variables which were clearly significant in the MLR models were considered in the case of the spatial models. In brackets, below the estimated parameters, are the p—value of the t test considering that the parameter is significantly different from zero if the
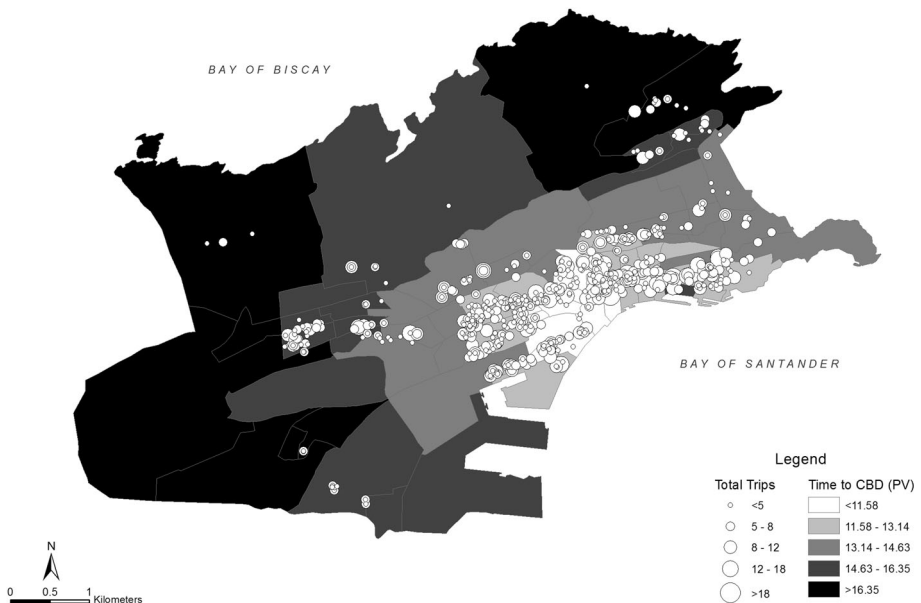


**Fig. 1** Distribution of households, total trips by household and travel time to CBD by Private Vehicle
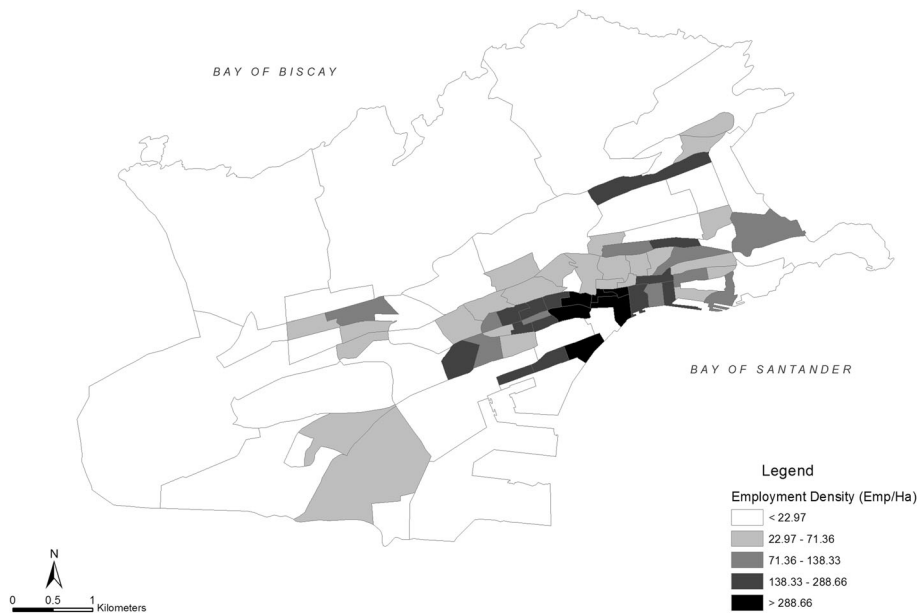
**Fig. 2** Employment density in the study area

value is equal to or lower than 0.1 (a confidence level of at least 90 % to avoid a type 1 error).

Eight models were estimated corresponding to the home-based trips for work purpose. The spatial models were specified considering the possible presence of spatial dependence in the dependent variable (SAR), in the residuals of the model (SEM) and using spatial filtering (SPO). The programming language R and the software package spdep (Bivand et al. 2015) were used to estimate the models.

The MLR models had a fit of 20 and 30 % of the explained variance. The variable WORKERS was clearly significant and positive with an effect of 0.5 trips in the PV model and 1 trip per additional worker in the NPV model. The variable AGE was significant to a 95 % confidence level and had a positive sign only in the NPV model. In this case $AGE^2$ was negative suggesting that after reaching a certain age, the journeys for work purpose start to decrease. The variable WOMEN was not significant for any case, while CHILDREN was negative for HBW-NPV journeys, suggesting that households with children tend to use less alternative transport modes when the workers travel to their jobs. The vehicles owned by the households (VEH variable) were clearly significant in the two models although with opposite signs (car ownership reduces the journeys made using alternative modes to the private vehicle). Household income was not a significant factor in explaining trip production in any of the cases and POP was only significant and positive in some of the PV models.

The variable ACC is the most relevant to the aims of this study. The variable was significant in the PV trips and clearly not significant in the NPV trips. In the PV trips, ACC showed a negative sign, meaning these are reduced, ceteris paribus, as accessibility to work opportunities increased. The introduction of the variable in a logarithmic form was also tested to capture if the accessibility had decreasing returns, but the results did not increase

**Table 4** Estimated parameters of the HBW trip production models

| Variables | MLR-HBW-PV | SAR-HBW-PV | SEM-HBW-PV | SPO-HBW-PV | MLR-HBW-NPV | SAR-HBW-NPV | SEM-HBW-NPV | SPO-HBW-NPV |
|---|---|---|---|---|---|---|---|---|
| (Constant) | 6.705(.025) | 4.178(.155) | 5.692(.104) | 6.241(.172) | −0.842(.083) | −0.377(.347) | −0.344(.383) | −3.976(.000) |
| AGE | −0.011(.535) | – | – | – | 0.034(.047) | 0.033(.049) | 0.032(.057) | 0.156(.000) |
| $AGE^2$ | 0.000(.558) | – | – | – | −0.000(.037) | −0.000(.038) | −0.000(.044) | −0.002(.000) |
| CHILDREN | 0.406(.124) | – | – | – | −0.701(.007) | −0.688(.007) | −0.686(.007) | −0.562(.121) |
| WOMEN | 0.044(.603) | – | – | – | 0.069(.404) | – | – | – |
| WORKERS | 0.491(.000) | 0.479(.000) | 0.481(.000) | 0.711(.000) | 0.964(.000) | 1.003(.000) | 1.007(.000) | 0.924(.000) |
| SIZE | −0.091(.208) | – | – | – | 0.029(.680) | – | – | – |
| D_INCOME2 | −0.073(.752) | – | – | – | 0.137(.542) | – | – | – |
| D_INCOME3 | 0.111(.639) | – | – | – | 0.143(.534) | – | – | – |
| D_INCOME4 | −0.0001(1.00) | – | – | – | 0.267(.327) | – | – | – |
| VEH | 0.416(.000) | 0.388(.000) | 0.382(.000) | 0.554(.000) | −0.469(.000) | −0.438(.000) | −0.445(.000) | −0.448(.000) |
| ACC | −0.013(.023) | −0.009(.111) | −0.012(.085) | −0.016(.061) | 0.000(.173) | – | – | – |
| POP | 0.000(.008) | 0.026(.045) | 0.000(.049) | 0.000(.132) | 0.000(.870) | – | – | – |
| $\rho/\lambda$ | – | 0.267(.000) | 0.217(.018) | – | – | 0.008(.917) | −0.041(.631) | – |
| $R^2$ | 0.21 | – | – | – | 0.31 | – | – | – |
| $R^2_{adj}$ | 0.20 | – | – | – | 0.30 | – | – | – |
| F | 17.96(.000) | – | – | – | 30.28(.000) | – | – | – |
| Moran's I | 0.067(.000) | 0.013(.308) | 0.027(.051) | 0.023(.100) | 0.017(.132) | 0.018(.184) | 0.026(.059) | 0.020(.140) |
| Robust LM—lag | 17.93(.000) | – | – | – | 1.36(.243) | – | – | – |
| Robust LM—error | 4.20(.040) | – | – | – | 0.24(.626) | – | – | – |
| AIC | 2695 | 2675 | 2683 | 1649 | 2653 | 2649 | 2649 | 1776 |
| Log-Likelihood | −1335 | −1331 | −1334 | −814 | −1314 | −1317 | −1316 | −879 |
| LR (MLR) | – | 15.87(.000) | 8.01(.005) | – | – | 0.01(.916) | 0.22(.640) | – |
| N | 817 | 817 | 817 | 817 | 817 | 817 | 817 | 817 |

the goodness of fit of the models or the significance of the variable. If the models are estimated using the CBD variable, the results for other parameters were very similar, and the CBD parameter showed a positive sign and was more clearly significant than the accessibility indicator (parameter value: 0.125 and $p$ value $= 0.000$ in the MLR-HBW-PV model). The effect is therefore the same, as the distance from the centre increased so did the number of journeys made by private vehicle.

For the two MLR models, the Moran I index was calculated for the residuals of the regression. The index was significant in the case of the PV trips, leading to the acceptance of the hypothesis of spatial autocorrelation in the residuals of the model. The Robust Lagrange Multiplier Test (Robust LM Test, see Table 4) was also performed on the MLR models in order to confirm the presence of spatial dependence in the dependent variable (LM—lag) or in the error term (LM-Error). The test was clearly significant for the two cases in the HBW-PV trips although the LM—lag test showed a higher value.

The trip production models considering spatial dependence were specified with the same variables which had shown to be significant and not correlated in the MLR models. Different specifications were also made to check if the variables that had previously not been significant became significant using the spatial models, however, in all cases the results were negative. Spatial autoregressive models with autoregressive disturbances (SARAR models) were also tested but the results were very similar to the SAR and SEM models considering the parameters estimated and the goodness of fit.

Given the nature of the household data (point data structure), a neighbourhood matrix with ten nearby neighbours was used. It was preferred to use this type of matrix rather than a distance matrix because the latter may generate a very irregular distribution of neighbours between observations. This phenomena is to be avoided because it may violate the regularity hypothesis which is required to obtain the asymptotic properties of the estimators and the statistical tests (Anselin 2002).

The parameters obtained in the MLR models are similar to those from SAR and SEM, although the MLR models do have a certain tendency to slightly overvalue some values compared to the spatial models (for example the parameters of the variable ACC equal to $-0.013$ vs. $-0.009$ in the SAR model). The SPO model was estimated as a quasi—Poisson model, given that overdispersion was significant using a regression-based test (Cameron and Trivedi 1990).

The accessibility indicator ACC was nearly significant in the SAR model and significant, at a 90 % confidence level, in the SEM and SPO models. If the CBD variable is introduced, the parameter was clearly significant (e.g. parameter value: 0.094 and p-value $= 0.000$ in the SAR-HBW-PV model). By calculating the elasticity of the ACC variable in the HBW trips using the expression $\beta(\overline{X}/\overline{Y})$, it can be seen how the PV generated trips are elastic with an absolute value between 7.4 and 10.7 %, depending on the model per 1 % increase in the accessibility indicator. In the case of the Poisson model, the estimated parameter gives a reduction of 1.6 % in the trip production by PV, per one additional unit increase in the ACC indicator.

The fit of the SAR and SEM PV models was very similar but slightly better in the case of the SAR model. In both cases the $\rho$ or $\lambda$ parameters were clearly significant. The LR test, which allows the fit of the SAR and SEM models to be compared with the fit of the MLR models (Ortúzar and Willumsen 2001), presented significant values (i.e. the spatial models had a significantly better fit to the data). The Moran I test performed on the residuals, also showed that the spatial correlation decreased and was not statistically significant at a 95 % confidence level in all the models.

Therefore, the results seem to indicate the presence of a moderate degree of spatial autocorrelation in the error term, in the case of the MLR-HBW-PV model. Figure 3 shows the Getis—Ord Gi* statistic (Ord and Getis 1995) estimated with the residuals of this model. The point observations from the households have been transformed into polygons using the Thiessen polygons technique. Note how the positive residuals, where the model makes predictions below the observed values, are significantly concentrated in the city outskirts, whereas the negative residuals are concentrated significantly closer to the city centre. This would suggest that the accessibility variables ACC and CBD partially capture the effect of increased HBW and PV trip generation producing an effect of spatial correlation in the residuals. This effect is minimized by the spatial models meaning spatial correlation ceases to be significant.

## Trip production for study purpose without and considering spatial dependence between observations

The models of trip production for study purpose can be seen in Table 5. The fit of the MLR models was similar to that of the models for work purpose although slightly higher in the case of the NPV trips and slightly lower for the PV trips. The variable accessibility to places of education was clearly not significant for the PV and NPV trips in the MLR models. However, if the models are specified using the CBD variable, this was significant for the PV trips with an estimated parameter of 0.04 (p—value: 0.02) and an elasticity of 2.5 % for an increase of 1 % in journey time. The LR tests were significantly better in the cases of SAR-HBS-PV and SEM-HBS-NPV although with moderate values. The Poisson models also showed a significant overdispersion because of which they were estimated as quasi—Poisson models. In the case of SPO-HBS-PV, if the model was specified with the
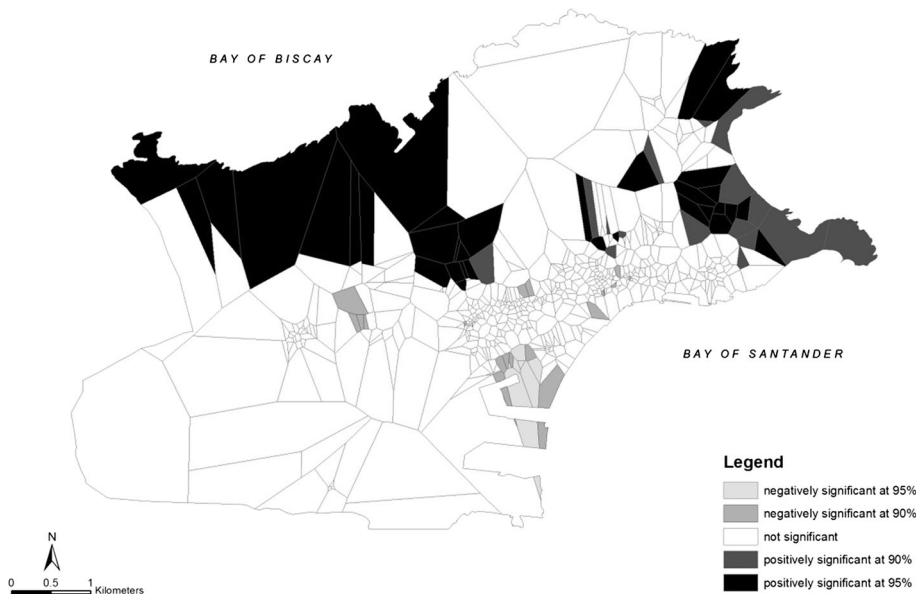


**Fig. 3** Getis—Ord Gi* statistic in the residuals of the MLR-HBW-PV models

**Table 5** Estimated parameters of the HBS trip production models

| Variables | MLR-HBS-PV | SAR-HBS-PV | SEM-HBS-PV | SPO-HBS-PV | MLR-HBS-NPV | SAR-HBS-NPV | SEM-HBS-NPV | SPO-HBS-NPV |
|---|---|---|---|---|---|---|---|---|
| (Constant) | 1.881(.164) | 0.997(.000) | 1.029(.000) | −0.763(.500) | 3.204(.000) | 3.330(.000) | 3.427(.000) | −1.038(.142) |
| AGE | −0.036(.001) | −0.040(.000) | −0.040(.000) | −0.044(.456) | −0.140(.000) | −0.137(.000) | −0.139(.000) | 0.036(.354) |
| $AGE^2$ | 0.000(.010) | 0.000(.002) | 0.000(.002) | 0.000(.749) | 0.001(.000) | 0.001(.000) | 0.001(.000) | −0.001(.006) |
| CHILDREN | 0.044(.795) | – | – | – | −1.269(.001) | −1.281(.000) | −1.308(.000) | −0.561(.004) |
| WOMEN | 0.067(.213) | – | – | – | 0.166(.045) | 0.158(.054) | 0.148(.070) | 0.102(.201) |
| WORKERS | −0.165(.001) | −0.124(.004) | −0.124(.004) | −0.576(.000) | −0.383(.000) | −0.368(.000) | −0.368(.000) | −0.369(.000) |
| SIZE | 0.039(.393) | – | – | – | 0.813(.000) | 0.829(.000) | 0.836(.000) | 0.748(.000) |
| D_INCOME2 | −0.140(.340) | – | – | – | 0.174(.444) | – | – | – |
| D_INCOME3 | −0.148(.324) | – | – | – | 0.177(.447) | – | – | – |
| D_INCOME4 | -0.013(.942) | – | – | – | 0.301(.274) | – | – | – |
| VEH | 0.402(.000) | 0.417(.000) | 0.418(.000) | 1.418(.000) | −0.271(.000) | −0.251(.000) | −0.255(.000) | −0.202(.011) |
| ACC | −0.008(.487) | – | – | – | 0.001(.182) | – | – | – |
| POP | 0.000(.668) | – | – | – | −0.000(.605) | – | – | – |
| $\rho/\lambda$ | – | 0.156(.038) | 0.118(.013) | – | – | 0.080(.202) | 0.135(.075) | – |
| $R^2$ | 0.16 | – | – | – | 0.44 | – | – | – |
| $R^2_{adj}$ | 0.14 | – | – | – | 0.43 | – | – | – |
| F | 12.25(.000) | – | – | – | 52.55(.000) | – | – | – |
| Moran's I | 0.049(.000) | 0.019(.166) | 0.023(.100) | 0.009(.468) | 0.050(.000) | 0.036(.009) | 0.025(.061) | 0.020(.147) |
| Robust LM—lag | 4.90(.027) | – | – | – | 0.19(.663) | – | – | – |
| Robust LM—error | 0.84(.360) | – | – | – | 5.89(.015) | – | – | – |
| AIC | 1953 | 1946 | 1949 | 736 | 2670 | 2666 | 2665 | 1354 |
| Log-Likelihood | −963 | −966 | −967 | −362 | −1322 | −1323 | −1322 | −667 |
| LR (MLR) | – | 4.34(.037) | 1.99(.158) | – | – | 1.52(.218) | 2.91(.009) | – |
| N | 817 | 817 | 817 | 817 | 817 | 817 | 817 | 817 |

**Table 6** Estimated parameters of the HBO trip production models

| Variables | MLR-HBO-PV | SAR-HBO-PV | SEM-HBO-PV | SPO-HBO-PV | MLR-HBO-NPV | SAR-HBO-NPV | SEM-HBO-NPV | SPO-HBO-NPV |
|---|---|---|---|---|---|---|---|---|
| (Constant) | −0.447(.880) | −0.180(.126) | −0.074(.524) | −1.803(.000) | −4.293(.000) | −3.844(.000) | −3.719(.000) | −1.493(.000) |
| AGE | −0.026(.132) | — | — | — | 0.124(.000) | 0.109(.000) | 0.108(.000) | 0.040(.000) |
| AGE² | 0.000(.135) | — | — | — | −0.001(.003) | −0.001(.004) | −0.001(.005) | −0.000(.011) |
| CHILDREN | −0.288 (.274) | — | — | — | 0.447(.324) | — | — | — |
| WOMEN | 0.016(.851) | — | — | — | −0.081(.575) | — | — | — |
| WORKERS | −0.192(.010) | −0.207(.001) | −0.207(.001) | −0.316(.001) | −0.893(.000) | −0.917(.000) | −0.963(.000) | −0.399(.000) |
| SIZE | 0.186(.010) | 0.193(.000) | 0.192(.000) | 0.272(.001) | 1.246(.000) | 1.154(.000) | 1.168(.000) | 0.450(.000) |
| D_INCOME2 | 0.037(.872) | — | — | — | 0.076(.849) | — | — | — |
| D_INCOME3 | 0.007(.977) | — | — | — | 0.239(.556) | — | — | — |
| D_INCOME4 | −0.028(.921) | — | — | — | −0.344(.475) | — | — | — |
| VEH | 0.447(.000) | 0.401(.000) | 0.400(.000) | 0.615(.000) | −0.523(.000) | −0.526(.000) | −0.510(.000) | −0.181(.000) |
| ACC | 0.002(.733) | — | — | — | 0.002(.000) | 0.001(.000) | 0.002(.000) | 0.001(.000) |
| POP | 0.000(.238) | — | — | — | −0.000(.185) | — | — | — |
| ρ/λ | — | 0.189(.007) | 0.180(.014) | — | — | 0.287(.000) | 0.319(.000) | — |
| R² | 0.08 | — | — | — | 0.31 | — | — | — |
| R²adj | 0.07 | — | — | — | 0.30 | — | — | — |
| F | 5.93(.000) | — | — | — | 29.58(.000) | — | — | — |
| Moran's I | 0.058(.000) | 0.024(.087) | 0.027(.050) | 0.020(.152) | 0.103(.000) | 0.028(.043) | 0.023(.090) | 0.019(.158) |
| Robust LM—lag | 6.04(.014) | — | — | — | 3.56(.059) | — | — | — |
| Robust LM—error | 1.91(.167) | — | — | — | 2.62(.106) | — | — | — |
| AIC | 2689 | 2673 | 2674 | 1773 | 3580 | 3560 | 3560 | 3227 |
| Log-Likelihood | −1332 | −1330 | −1331 | −876 | −1777 | −1771 | −1771 | −1593 |
| LR (MLR) | — | 7.02(.007) | 5.63(.018) | — | — | 22.04(.000) | 21.51(.000) | — |
| N | 817 | 817 | 817 | 817 | 817 | 817 | 817 | 817 |

CBD variable, the estimated parameter was also significantly different from 0, with a value for the parameter of 0.163 (p-value: 0.018).

## Trip production models for other purposes without and considering spatial dependence between observations

Table 6 shows the parameters for the trip production for other purposes in the MLR, SAR, SEM and SPO models.

The MLR models had a poor fit in the case of the PV trips and a better performance in the NPV with 30 % of the explained variance. The most noteworthy change with respect to HBW and HBS models was the positive sign of the ACC parameter in the case of NPV trips. This means that in this case, greater accessibility to opportunities significantly increases the number of trips that are produced. In the case of the journey time to the city centre, the longer the journey time, the fewer household trips were produced by NPV with a parameter value equals to $-0.032$ (p—value: 0.000). The Moran I index presented clearly significant values whereas the LM—test was only significant for the lag type model.

In the spatial models, the accessibility variables continue to be clearly significant for the HBO-NPV. If the elasticities of the ACC variable are calculated for the HBO-NPV trips, it can be seen how the trips being produced are inelastic, with a value between 0.4 and 0.8 % according to the model. The quasi—Poisson SPO-HBP-NPV model estimated an increase of 0.1 % for one additional unit in the indicator of accessibility to opportunities. This shows that although accessibility increases so does the production of HBO trips using alternative modes to the private vehicle, although the growth is not very high.

Figure 4 shows a representation of the Getis—Ord $G^*$ statistic for the residuals of the MLR-HBO-NPV model, once again representing the households as Thiessen polygons. In this case, contrary to the details shown in Fig. 3, the positive residuals are mainly
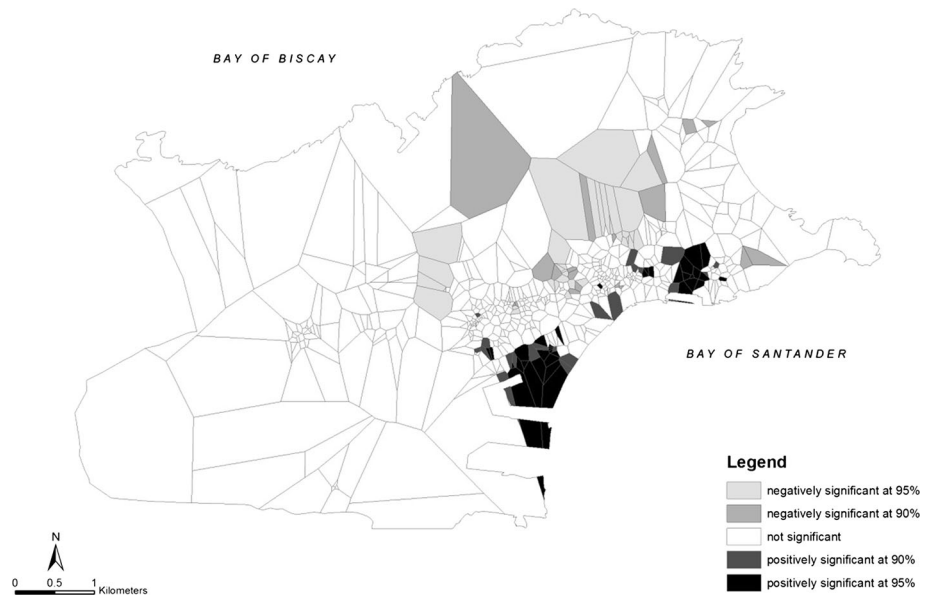


**Fig. 4** Getis—Ord Gi$^*$ statistic in the residuals of the MLR-HBO-NPV model

Table 7 Total impacts of the SAR models

| Variables | SAR-HBW-PV | SAR-HBW-NPV | SAR-HBS-PV | SAR-HBS-NPV | SAR-HBO-PV | SAR-HBO-NPV |
|---|---|---|---|---|---|---|
| AGE | – | 0.033(.033) | –0.047(.000) | –0.149(.000) | – | 0.153(.000) |
| AGE$^2$ | – | –0.000(.030) | 0.000(.004) | 0.001(.000) | – | –0.001(.006) |
| CHILDREN | – | –0.693(.003) | – | –1.392(.000) | – | – |
| WOMEN | – | – | – | 0.171(.045) | – | – |
| WORKERS | 0.653(.000) | 1.010(.000) | –0.146(.002) | –0.400(.004) | –0.256(.000) | –1.286(.000) |
| SIZE | – | – | – | 0.901(.000) | 0.238(.000) | 1.619(.000) |
| D_INCOME2 | – | – | – | – | – | – |
| D_INCOME3 | – | – | – | – | – | – |
| D_INCOME4 | – | – | – | – | – | – |
| VEH | 0.528(.000) | –0.441(.000) | 0.495(.000) | –0.273(.000) | 0.494(.000) | –0.738(.000) |
| ACC | –0.012(.108) | – | – | – | – | 0.002(.000) |
| CBD | – | – | – | – | – | – |
| POP | 0.000(.014) | – | – | – | – | – |

**Table 8** Eigenvectors coefficients for constructing the spatial filters

| Eigenvector | SPO-HBW-PV | SPO-HBW-NPV | SPO-HBS-PV | SPO-HBS-NPV | SPO-HBO-PV | SPO-HBO-NPV |
|---|---|---|---|---|---|---|
| EV1 | −5.927(.000) | 2.172(.089) | 11.368(.002) | 5.533(.000) | 6.653(.006) | 2.095(.025) |
| EV2 | −3.282(.003) | 3.557(.010) | – | −4.689(.004) | −9.320(.000) | 3.135(.008) |
| EV3 | 4.268(.002) | −6.312(.000) | – | – | 5.464(.001) | 1.774(.017) |
| EV4 | 6.532(.000) | – | – | – | −2.927(.162) | −4.367(.000) |
| EV5 | −7.688(.000) | – | – | – | 2.916(.072) | −2.406(.002) |
| EV6 | 2.799(.027) | – | – | – | −8.055(.014) | 1.827(.004) |
| EV7 | – | – | – | – | −12.537(.000) | −3.783(.002) |
| EV8 | – | – | – | – | – | −1.688(.014) |
| EV9 | – | – | – | – | – | 1.961(.010) |
| EV10 | – | – | – | – | – | −1.456(.066) |
| EV11 | – | – | – | – | – | 1.644(.049) |
| EV12 | – | – | – | – | – | −1.792(.022) |
| EV13 | – | – | – | – | – | 4.722(.004) |
| EV14 | – | – | – | – | – | −2.751(.005) |

concentrated in the city's central zone, whereas the negative residuals are located in the peripheral areas. This effect explains the spatial correlation in the residuals of the MLR models, an effect which is minimized by the spatial models in which the Moran I index ceased to be significantly different from zero with the exception of the SAR-HBO-NPV model.

## Total impacts of the SAR models, eigenvector coefficients and comparison of the results

The total impacts on the dependent variable were also calculated in the SAR models. These effects take into account the direct effect of each independent variable on the dependent variable, as well as the indirect effects of spatial spillover from the dependent variable towards neighbouring observations and from neighbouring observations towards the local dependent variable (see Table 7) (LeSage and Pace 2010b). Taking these direct and indirect effects into account, the SAR models infer that the total effect of changing certain variables, particularly SIZE, WORKERS and VEH, is greater than that estimated using the MLR models. This data provides evidence of the existence of spatial effects on the trip production rates. The parameters estimated in the Poisson models for the eigenvectors of the spatial filter can also be seen in Table 8. The SPO-HBO-NPV model was the one that presented a greater number of significant eigenvectors when constructing the spatial filter.

Comparing these results with other studies, these agree with those of Vickerman (1974) who showed that accessibility was a positive factor in trip production, mainly for optional journeys for shopping or leisure purposes. Furthermore, the fits of the models were similar to those obtained by Vickerman after they removed the collinearity problems associated with the models. Using a similar study area to the one used in this research (monocentric small—medium sized city), Hanson and Schwab (1987) also found a significant, although weak, effect of accessibility on optional journeys, although the only indicator of accessibility used by these authors was the accumulated opportunities type, they did not use gravity type indicators or journey time to city centre in spite of even the authors themselves recognising "If distance of the residence from the central business district (CBD) can be taken as a valid surrogate for the density of opportunities (…) then there is some evidence that low accessibility levels are related to lower trip frequencies". Thill and Kim (2005) found that trip production was affected positively or negatively by accessibility, similarly to the results of the models presented in this research considering the difference between using PV and NPV. Cervero and Kockelman (1997) also found that accessibility could be a negative factor in trip production for the PV mode as has been found in the models presented.

If these results are compared to those obtained in other research including the possible presence of spatial dependence between observations, Gamas et al. (2006) found that the MLR models they estimated showed biased parameters that overvalued the influence of specific variables compared to spatial models. This phenomenon is similar to that found in the present study when addressing only the direct effects of each independent variable. However, if the indirect effects are also considered in the case of the SAR model (untested by Gamas et al. (2006)) the effect of certain variables may be greater than that found by the MLR models. The results of Kwigizile and Teng (2009) also found that the ordinary regression models generally overvalued the estimated parameters when compared with the spatial models, although the authors did not address the influence of indirect effects. Finally, Kim et al. (2009) found the presence of spatial autocorrelation in both non-home-based (NHB) trip production and attraction. There seems to be evidence that both trip

production and attraction may present the effects of spatial correlation and that this is dependent on the characteristics of the study area.

## Validation of the models

The models in the predictive phase were validated using a K—Fold type cross—validation (James et al. 2014). The models chosen for the validation were MLR-HBW-PV, SAR-HBW-PV, SEM-HBW-PV and SPO-HBW-PV among the PV models and MLR-HBO-NPV, SAR-HBO-NPV, SEM-HBO-NPV and SPO-HBO-NPV among the models in other modes of transport. These models were selected because they were the ones that showed a greater degree of spatial correlation and where the LR test demonstrated a better improvement in the fit compared with the MLR models.

The SAR and SEM models can work in predictive phase using the technique proposed by Bivand (2002). Bivand proposes separating the model terms in the trend, the signal and the noise. The trend in both the SAR and SEM models is given by the matrix $X\beta$, in other words, the matrix of independent variables and the vector of parameters to be estimated. The signal in the SAR model can be found using:

$$\hat{\rho} W \hat{y} = \hat{\rho} W (I - \hat{\rho} W)^{-1} X \hat{\beta} \tag{9}$$

The above has the inconvenience of losing the part of the signal belonging to the error term. In the SEM model the signal is equal to 0 meaning that the prediction is given only by the trend.

The validation was done by randomly dividing the sample into 10 groups, successively estimating the models with 9 of the groups and performing the prediction with the remaining group. This process was repeated with all 10 groups until all the observations had a prediction value in validation mode. Later, the mean squared error (MSE) was calculated for the residuals of the predictions compared to the real value. In various empirical applications the cross—validation with 5 or 10 groups has shown no excessive bias or variance in the estimation of the MSE (James et al. 2014; Kohavi 1995). The cross-validation estimate is computed for k groups as:

$$CV_k = \frac{1}{k} \sum_{i=1}^{k} MSE_i \tag{10}$$

The results of the validation were similar for the MLR, SAR and SEM models (see Table 9). However, the SAR and SEM models have a slight advantage over the MLR based

| Table 9 Cross—Validation of the models | Model | CV |
|---|---|---|
| | MLR-HBW-PV | 1.578 |
| | SAR-HBW-PV | 1.572 |
| | SEM-HBW-PV | 1.576 |
| | SPO-HBW-PV | 1.431 |
| | MLR-HBO-NPV | 4.685 |
| | SAR-HBO-NPV | 4.682 |
| | SEM-HBO-NPV | 4.646 |
| | SPO-HBO-NPV | 4.396 |

model in terms of their ability to predict trip production in both the PV and NPV modes. This improvement in prediction is almost certainly due to the spatial models helping to reduce spatial correlation in the residuals generated in the periphery and the urban centre with the HBW and HBO trips, respectively. The Poisson regression models SPO-HBW-PV and SPO-HBO-NPV presented the lowest MSE. This was probably due to the better adaptation of the Poisson distribution to the patterns of household trip generation.

## Conclusions

This article has described work to estimate MLR, SAR, SEM and quasi-Poisson models considering spatial effects between observations, to predict the number of trips made by a sample of households in the urban area of Santander (Spain). The models were specified with indicators of accessibility to opportunities in order to determine if this is a significant factor when explaining trip generation. This aspect is important when considering that journey demand should be sensitive to the transport costs implied in accessing the opportunities, an expected fact according to the available theory about induced demand and microeconomic logic.

The estimated MLR models showed how the production of journeys is sensitive to the ACC variable, a gravity type accessibility indicator, and the CBD variable, the time it takes to get to the city centre. This sensitivity was the opposite in the trips produced using private vehicle and other modes of transport. While in the case of PV trip production for work purpose, accessibility elastically reduced the number of trips, for other modes of transport (mainly on foot), accessibility could increase trip production, especially for optional trips, although in a non-elastic way. These results provide evidence that, on an intraurban scale and for an urban area like the one being studied, accessibility could be a significant factor in the production of trips and that trips made using a private vehicle could be reduced by increasing accessibility to opportunities. Therefore, it would be recommendable to use some type of accessibility indicator for simulating the possible effects of changes made to the transport supply or the pattern of activities. If this is not done then the predictions would surely misestimate the transport demand of optional trips by NPV or mandatory trips by PV in the network being studied. In the case of the urban area of Santander, a city with a monocentric distribution of opportunities, the journey time from each zone to the city centre was an indicator for estimating trip production as good as the gravity type indicator in terms of the goodness of fit of the models. In the case of other urban areas with more decentralised distributions of opportunities, gravity accessibility indicators could be more suitable for predicting trip production.

The models which considered the presence of spatial dependence between observations, SAR, SEM and SPO were able to support these results and deal with the existence of a significant degree of spatial dependence in the production of HBW-PV and HBO trips and, to a lesser degree, HBS trips. The consideration of the presence of spatial dependence and the direct and indirect effects of the variables in the SAR production models was important in the estimation and correct interpretation of the parameters which in the cases of the variables SIZE, WORKERS and VEH showed greater effects than those estimated by the MLR models. This phenomenon points to the effects of spatial correlation in the residuals derived from the fact that in the urban periphery the MLR models underestimate the production of HBW trips using private vehicle, while the production of HBO trips by alternative modes of transport was underestimated in the urban centre. Although this effect

is not of any great magnitude in the study area (a small-medium sized city without significant congestion problems), it could be higher in larger urban areas with worse traffic conditions. This effect could be due to mutual social influence between households resulting from different social characteristics between city centre households and those in the suburbs. In addition, the lack of a variable with spatially differentiated effects could also cause spatial autocorrelation in the residuals of the linear models although the fit of the SAR models and the LM-lag tests were higher than the fit of the SEM models and the LM-error tests for both, the MLR-HBW-PV and the MLR-HBO-NPV specifications.

Finally, the validation of the models using cross—validation with 10 groups, showed that the spatial models had a mean squared error slightly lower than the MLR models in the prediction of overall trip production, derived from the explicit consideration of the effect of spatial correlation in the residuals in the centre and periphery of the study area. In addition, the SPO models performed better than the continuous models by considering the truncated and discrete nature of trip generation.

Future lines of research stemming from the present work could involve the application of similar models in other study areas to determine the importance of differences in accessibility and spatial correlation between observations in larger cities with greater problems of traffic congestion. Furthermore, it would also be of interest to test the existence of spatial heterogeneity in the parameters of trip generation models, a path already opened by other researchers (Roorda et al. 2010) which could also be applied in evaluating the accessibility effects.

# References

Anselin, L.: Spatial econometrics: methods and models. In: Studies in operational regional science 4 Kluwer. Academic Publishers, Dordrech (1988)

Anselin, L.: Under the hood: issues in the specification and interpretation of spatial regression models. Agric. Econ. **27**, 247–267 (2002)

Bamberg, S., Hunecke, M., Blöbaum, A.: Social context, personal norms and the use of public transportation: two field studies. J. Env. Psychol. **27**, 190–203 (2007). doi:10.1016/j.jenvp.2007.04.001

Bivand, R.: Spatial econometrics functions in R, Classes and methods. J. Geograph. Syst. **4**, 405–421 (2002). doi:10.1007/s101090300096

Bivand, R., et al.: Package 'spdep'. Weighting Schemes, Statistics and Models, Spatial Dependence (2015)

Cameron, A.C., Trivedi, P.K.: Regression-based tests for overdispersion in the Poisson model. J. Econom. **46**, 347–364 (1990). doi:10.1016/0304-4076(90)90014-K

Carrasco, J.A., Farber, S.: Selected papers on the study of the social context of travel behaviour. Trans. Res. Part. A. **68**, 1–2 (2014). doi:10.1016/j.tra.2014.09.011

Cascetta, E.: Transportation systems analysis: Models and applications. In: Springer optimization and its applications. 2nd edn, vol. 29. Springer, New York (2009)

Cervero, R.: Induced Travel Demand: research Design. Empir. Evid. Norm. Pol. J. Plan. Lit. **17**, 3–20 (2002). doi:10.1177/088122017001001

Cervero, R., Kockelman, K.: Travel demand and the 3Ds: density, diversity, and design. Transp. Res. Part. D **2**, 199–219 (1997). doi:10.1016/S1361-9209(97)00009-6

Cervero, R., Sarmiento, O.L., Jacoby, E., Gomez, L.F., Neiman, A.: Influences of built environments on walking and cycling: lessons from Bogotá. Inter. J. Sustain. Transp. **3**, 203–226 (2009). doi:10.1080/15568310802178314

Clark, A.F., Scott, D.M.: Does the social environment influence active travel? An. Investig. walk. Hamilt. Can. J. Transp. Geogr. **31**, 278–285 (2013). doi:10.1016/j.jtrangeo.2013.06.005

Clifton, K.J., Currans, K.M., Muhs, C.D.: Adjusting ITE's Trip Generation Handbook for urban context. J. Trans. Land. Use. **8**, 5–29 (2015)

Coppola, P., Nuzzolo, A.: Changing accessibility, dwelling price and the spatial distribution of socio-economic activities. Res. Transp. Econ. **31**, 63–71 (2011). doi:10.1016/j.retrec.2010.11.009

Chang, J.S., Jung, D., Kim, J., Kang, T.: Comparative analysis of trip generation models: results using home-based work trips in the Seoul metropolitan area. Transp. Lett. **6**, 78–88 (2014). doi:10.1179/1942787514Y.0000000011

Ewing, R., DeAnna, M., Li, S.-C.: Land use impacts on trip generation rates. Transp. Res. Rec. **1518**, 1–6 (1996)

Gamas, J., Anderson, W., Pastor, C.: Estimation of trip generation in Mexico City,Mexico, with Spatial Effects and Urban Densities Transportation Research Record. J. Transp. Res. Board. **1985**, 49–60 (2006). doi:10.3141/1985-06

Glaeser, E.L.: Agglomeration economics National Bureau of Economic Research conference report. The University of Chicago Press, Chicago (2010)

Goodwin, P.: Empirical evidence on induced traffic Transportation **23**, 35–54 (1996). doi:10.1007/BF00166218

Greene, W.H.: Econometric analysis. Prentice-Hall International, Upper Saddle River, New Jersey (2003)

Griffith, D.A.: Spatial autocorrelation and spatial filtering: gaining understanding through theory and scientific visualization. Springer Science & Business Media, New York (2013)

Handy, S.L., Niemeier, D.A.: Measuring accessibility: an exploration of issues and alternatives. Env. Plan. A. **29**, 1175–1194 (1997)

Hanson, S., Schwab, M.: Accessibility and intraurban travel. Env. Plan. A. **19**, 735–748 (1987). doi:10.1068/a190735

Horner, M.W.: Spatial dimensions of urban commuting: a review of major issues and their implications for future geographic research. Prof. Geogr. **56**, 160–173 (2004). doi:10.1111/j.0033-0124.2004.05602002.x

Hymel, K.M., Small, K.A., Dender, K.V.: Induced demand and rebound effects in road transport. Trans. Res. Part. B **44**, 1220–1241 (2010). doi:10.1016/j.trb.2010.02.007

INE: National Population and Housing Census. Spanish National Institute of Statistics (2011)

Institute of Transportation Engineers: Trip Generation Manual, 9th edn. ITE, Washington, DC (2012)

James, G., Witten, D., Hastie, T., Tibshirani, R.: An Introduction to statistical learning: with applications in R. Springer, New York (2014)

Kim J-H, Chung J-H, Kim I, Lim U(2009) Spatial Dependence in Trip Generation Models: Spatial Econometric Perspective. In: Transportation Research Board 88th Annual Meeting, p 23

Kitamura, R., Akiyama, T., Yamamoto, T., Golob, T.: Accessibility in a metropolis: toward a better understanding of land use and travel. Transp. Res. Rec. **1780**, 64–75 (2001)

Koenig, J.G.: Indicators of urban accessibility: theory and application. Transportation **9**, 145–172 (1980). doi:10.1007/BF00167128

Kohavi, R.: A study of cross-validation and bootstrap for accuracy estimation and model selection. Paper presented at the 14th International Joint Conference on Artificial Intelligence (IJCAI), Montreal, Quebec (1995)

Kwigizile, V., Teng, H.: Comparison of methods for defining geographical connectivity for variables of trip generation models. J. Transp. Eng **135**, 454–466 (2009). doi:10.1061/(ASCE)0733-947X(2009)135:7(454)

LeSage, J.P., Pace, R.K.: Introduction to spatial econometrics. Statistics, textbooks and monographs. CRC Press, Boca Raton (2009)

Lesage, J.P., Pace, R.K.: The biggest myth in spatial econometrics. Available at SSRN: http://www.ssrncom/abstract=1725503, 10.2139/ssrn1725503 (2010a)

LeSage, J.P., Pace, R.K.: Spatial Econometric Models. Handbook of Applied Spatial Analysis: Software Tools, pp. 355–376. Methods and Applications. Springer, Berlin (2010b)

Lim, K.-K., Srinivasan, S.: Comparative analysis of alternate econometric structures for trip generation models. Transp. Res. Rec. **2254**, 68–78 (2011)

Lopes, S., Brondino, N., Rodrigues da Silva, A.: GIS-based analytical tools for transport planning: spatial regression models for transportation demand forecast. ISPRS International Journal of Geo-Information **3**, 565–583 (2014)

Martínez, F.J.: Towards a Land-use and transport interaction framework. In: Hensher, D.A., Button, K.J. (eds.) Handbook of transport modelling, pp. 145–164. Elsevier, Amsterdam (2000)

McNally, M.G.: The Four-Step Model. In: Hensher, D.A., Button, K.J. (eds.) Handbook of transport modelling (2nd Edition). Emerald Inc, Jodhpur (2008)

Millard-Ball, A.: Phantom trips: Overestimating the traffic impacts of new development. J. Transp. Land. Use. **8**, 31–49 (2015)

Minnen, J., Glorieux, I., van Tienoven, T.P.: Transportation habits: evidence from time diary data. Transp. Res. Part A **76**, 25–37 (2015). doi:10.1016/j.tra.2014.12.013

Ord, J.K., Getis, A.: Local spatial autocorrelation statistics: distributional issues and an application. Geogr. Anal. **27**, 286–306 (1995). doi:10.1111/j.1538-4632.1995.tb00912.x

Ortúzar, JdD, Willumsen, L.G.: Modelling transport, 3rd edn. J. Wiley, Chichester New York (2001)

Purvis, C., Iglesias, M., Eisen, V.: Incorporating work trip accessibility in nonwork trip generation models in san francisco bay area transportation research record. J. Transp. Res. Board. **1556**, 37–45 (1996). doi:10.3141/1556-06

Roorda, M.J., Páez, A., Morency, C., Mercado, R., Farber, S.: Trip generation of vulnerable populations in three Canadian cities: a spatial ordered probit approach. Transportation **37**, 525–548 (2010)

Scott, D.M., Dugundji, E.R., Páez, A.: The social dimension of activity, travel and location choice behavior. J. Transp. Geogr. **31**, 212–215 (2013). doi:10.1016/j.jtrangeo.2013.06.009

Sherwin, H., Chatterjee, K., Jain, J.: An exploration of the importance of social influence in the decision to start bicycling in England. Transp. Res. Part A **68**, 32–45 (2014). doi:10.1016/j.tra.2014.05.001

Thill, J.-C., Kim, M.: Trip making, induced travel demand, and accessibility. J. Geograph. Syst. **7**, 229–248 (2005). doi:10.1007/s10109-005-0158-3

Vickerman, R.W.: Accessibility, attraction, and potential: a review of some concepts and their use in determining mobility. Env. Plan. A. **6**, 675–691 (1974). doi:10.1068/a060675

Walters, J., Bochner, B., Ewing, R.: Getting trip generation right. Eliminating the Bias Against Mixed Use Development, American Planning Association (APA) (2013)

Wu, D., Sall, E., Newhouse, S.: Trip rates and accessibility: gleaning basic planning information from activity-based travel demand model. Transp. Res. Rec. **2303**, 81–88 (2012)

Zeileis, A., Kleiber, C., Jackman, S.: Regression models for count data in R. J. Stat. Softw. **27**, 1–25 (2008)

**Ruben Cordera** is researcher at the Transport Systems Research Group (GIST), University of Cantabria (Spain). His areas of research include Land Use—Transport interaction models and spatial econometrics methods.

**Pierluigi Coppola** is associate professor of Transport at the University "Tor Vergata" of Rome (Italy). His research interests include transport demand models, Land Use—Transport interaction models and static and dynamic traffic assignment models.

**Luigi dell'Olio** is professor of Transport Planning at the University of Cantabria (Spain). His research focuses on transit operations and planning, travel behaviour and Land Use—Transport Interaction models. He is Head of Transportation Demand Modelling Division of the Transport Systems Research Group (GIST).

**Ángel Ibeas** is professor of Transport Planning at the University of Cantabria (Spain). His research focuses on logistic, transit operation, travel behaviour and intelligent transportation systems. He is the director of The Transport Systems Research Group (GIST).