

Chapter 14

Analysis of the Spatial Distribution Pattern of Tourist Activity: An Application to the Volume of Travellers in Extremadura



Cristina Rodríguez-Rangel and Marcelino Sánchez-Rivero

14.1 Introduction

Spatial econometrics emerged as a branch of general econometrics due to the need for developing a set of techniques that would allow the adequate treatment of data affected by the so-called spatial effects: spatial autocorrelation or dependence and spatial heterogeneity. The proliferation of georeferenced databases motivates a greater need for knowing what is happening with those data in their spatial distribution, and especially whether this distribution involves any structure that should be known in order to better understand the relationships that occur between the variables in space. Anselin (2001) defines it as “a section of econometrics dedicated to the treatment of spatial interaction (spatial dependence) and spatial structure (spatial heterogeneity) in cross-section and panel data regression models”. As can be deduced from Anselin’s definition, there are two main effects that motivate the appearance of a subfield within traditional econometrics: spatial heterogeneity and spatial dependence or autocorrelation.

Spatial heterogeneity or lack of structural stability arose as a consequence of using different spatial units to explain a single phenomenon, and it can be solved with today’s techniques for the treatment of time series (Moreno & Vayá, 2004). Indeed, as the authors indicate, the effect of heterogeneity, although it is related to the unequal distribution of a variable in space, does not require the development of new techniques to be treated, since this can be achieved with techniques that have already been proposed by traditional econometrics.

C. Rodríguez-Rangel (✉) · M. Sánchez-Rivero
Faculty of Economics and Business Studies, University of Extremadura,
Avda Elvas, s/n, 06006 Badajoz, Spain
e-mail: mcrisrod@unex.es

M. Sánchez-Rivero
e-mail: sanriver@unex.es

© Springer Nature Switzerland AG 2019
A. Artal-Tur et al. (eds.), *Trends in Tourist Behavior*,
Tourism, Hospitality & Event Management,
https://doi.org/10.1007/978-3-030-11160-1_14

Thereby, the present work is focused on analysing the second spatial effect, i.e. spatial dependence or autocorrelation. The emergence of spatial econometrics is motivated by the fact that the multidirectionality of this effect cannot be treated with the traditional econometric techniques. Spatial dependence or autocorrelation is defined as the phenomenon that takes place when there is a relationship between what happens in a specific point in space and what occurs in other points of such space (Anselin, 1988).

Considering this definition, it can be inferred that the presence of spatial effects is expected, mainly, in those variables that measure aspects of economic activities that are especially linked to their development in a specific space. In regional economics, this effect has been studied in different variables, such as production, unemployment, available income, etc. Tourism stands out as an activity that is strongly related to the geographic space in which it is developed (Sánchez, 2008). Therefore, it is surprising to find very few studies in the literature to analyse the distribution patterns of tourist variables in space.

The aim of the present work is to study the distribution patterns of a variable that is usually associated with tourism, i.e. “the number of travellers”, to determine whether it is randomly distributed in space or, on the contrary, there is spatial autocorrelation or dependence. This will ultimately show that the modelling of any phenomenon related to tourism will require the use of the techniques developed by spatial econometrics for the treatment of data affected by the spatial effects.

To this end, the present work is distributed in the following manner: after this first introduction section that describes the objective of this investigation, the next section presents a review of the existing literature, highlighting the main contributions up to the present day in the field of spatial econometrics, especially in the techniques proposed for the exploratory analysis of spatial data. Then, another section describes the methodology used to achieve the objectives of the study. After the methodology, the main results obtained from the analysis conducted are presented and, lastly, a final section includes the main conclusions and implications.

14.2 Literature Review

The analysis of spatial data has gained great interest in the last few decades, especially in those fields of regional economics that are strongly related to their development in a specific geographic space (Button & Kulkarni, 2001; Chasco & Vicéns, 2000; López, Palacios, & Ruiz, 2001; Moreno & Vayá, 2000, 2004).

Tourism stands out among these activities for being clearly affected by its location; therefore, it is obviously one of the fields in which studies are being conducted on the use of techniques proposed by spatial econometrics (Barros & Matias, 2007; Ma, Hong, & Zhang, 2015; Pavlyuk, 2010, 2013; Sanchez, 2008; Sánchez, Sánchez, & Rengifo, 2013, 2018; Zhou, Maumbe, Deng, & Selin, 2015).

The first works carried out in this regard can be attributed to Student, who in the year 1914 conducted the first studies in this topic, with the aim of understanding how spatial effects can influence the validity of statistical methods.

Kmenta (1971) stated that the hypothesis of the independence of observations in cross-section data was the most questionable one, especially in specific cases. It was in the field of geography where researchers began to wonder about the lack of reliability of this hypothesis of traditional statistics. Thus, Cliff and Ord (1973, 1981) published their pioneer studies about the lack of independence that usually occurs between the observations of cross-section units.

An important milestone in the future evolution of these new techniques is their compilation under the specific name of “spatial econometrics”, which was proposed by Paelinck and Klaasen (1979).

The first studies conducted in this field were focused on the proposal of techniques to detect the presence of spatial autocorrelation between the observations of a sample. In this line, the formal indices proposed by Moran (1948) and Geary (1954) constitute the first tools to allow diagnosing the presence, or absence, of spatial autocorrelation between the observations of a variable.

Despite these first efforts, the great development of spatial econometrics did not take place until the 1980s, with the works of Anselin (1980, 1988), Arbia (1989), Bloommestein (1983) and Cliff and Ord (1981), which are considered to be the basic studies that lay the foundation for the methodology of spatial econometric analysis.

These works constitute the basic pillars on which spatial econometrics develops; in fact, as Chasco (2003) pointed out, the book “Spatial Econometrics: Methods and Models”, by Anselin (1988), has been considered as the reference proceedings book for studies conducted in this topic since the 1990s.

Later, other studies have been published in journals of regional economics that posed specific contributions for the advancement of this subfield of econometrics, such as the series of articles by Anselin and Florax (1995), and Anselin and Rey (1997).

Anselin and Florax (1995) considered that a convergence was taking place between different factors that were promoting an increasing interest for spatial econometrics, such as: a greater interest for the role of space and spatial interaction in social networks, the availability of large georeferenced economic databases and the development of an efficient and inexpensive technology that allows the application of these techniques, through computerised geographic information systems, to carry out the analysis of spatial data.

Despite this trend, it cannot be considered that nowadays spatial econometrics had managed to become a reference in applied analysis; this fact is confirmed by its low presence in proceedings books of econometrics, which barely mention these techniques. As Moreno and Vayá (2000) indicate, the number of specific publications in the present time is still very low in this topic, especially in the Spanish scope.

Chasco (2003) studied the aspects that are hindering the dissemination of these techniques among the research community, concluding that the convergence of factors such as the priority to develop prediction techniques, the scarcity of microterritorial statistical information and the absence of useful and inexpensive software,

were contributing to the poor acceptance of spatial econometric techniques among researchers.

Now that some of these limitations are being overcome, on the one hand, a large variety of specific software, GIS (geographic information systems), can be found in the market, which include the use of spatial statistic techniques. Among the most commonly used, it is worth highlighting: SpaceStat, developed by Luc Anselin (1992, 1995a), ArcGIS and S+SpatialStats, among others.

On the other hand, the possibilities offered by new technologies to obtain georeferenced data, such as those of applications that allow geolocalization in smartphones, “geotagging” in social networks, and GPS technology, among others, help researchers to analyse the movements of tourists in their destinations (Shoval and Ahas, 2016), obtaining, also, data with a high level of reliability and precision.

All this leads to think of a favourable evolution of spatial econometric techniques, now that some of the barriers that initially hindered their development are being overcome.

On the other hand, the present study does not suggest that the techniques proposed by spatial econometrics should now be used in every case. However, given their suitability and validity when a variable is affected by some of the so-called spatial effects, first of all, it should be analysed whether this is happening, and, secondly, when the existence of this effect in the analysed variable is confirmed, this should be treated with the methods proposed to that end. In this sense, the analysis of a variable’s spatial autocorrelation becomes a fundamental part of a first exploratory phase of any analysis. It is considered that there is autocorrelation when a relationship is confirmed between what takes place in a specific point of a given space and what occurs in other points of that space (Anselin, 1988).

For the diagnosis of autocorrelation, a series of formal indicators have been proposed, which allow confirming the presence, or absence, of this spatial effect in a variable. Among the most commonly used, it is worth highlighting the indices proposed by Moran, with the global I test (1948), the Geary’s c test (1954) and the G(d) test of Getis and Ord (1992).

When analysing the spatial autocorrelation of a variable with any of these proposed indices, three different scenarios can take place. The first scenario is the lack of spatial autocorrelation, that is, those cases in which the variable analysed is distributed in space following a random pattern. The second scenario is the detection of a positive spatial autocorrelation pattern. In this scenario, the presence of a specific phenomenon in a region leads to its expansion to other nearby regions (Moreno and Vayá, 2004). In the specific case of tourism, the presence of this type of autocorrelation poses the presence of similar values of the tourist variable between nearby destinations, which means that there is a “contagion” effect (Sánchez, 2008). This would be the case of tourist attractions, or the lack of these, that cause the attraction of tourists in nearby locations. The last possible scenario is the presence of negative spatial autocorrelation, when the presence of a phenomenon in a region prevents or hinders its appearance in neighbour regions (Moreno and Vayá, 2004). In the field of tourism, this is known as the “absorption” effect (Sánchez, 2008).

As was previously stated, the exploratory phase of any analysis will require verifying whether the study variable is, or is not, affected by spatial autocorrelation, in order to determine the most suitable techniques to treat it. The aim of the present work is to analyse the distribution pattern of the variable “number of travellers” in Extremadura. To this end, the next section describes the methodology used to reach this objective.

14.3 Methodology

The methodology used in the present work lies in exploratory spatial data analysis (ESDA), which emerged as a specific part of exploratory data analysis (EDA) with the aim of focusing on the specific treatment of spatial data.

Therefore, it is defined as the set of techniques that allow describing spatial distributions, identifying atypical localizations (spatial outliers), discovering schemes of spatial association (spatial cluster) and suggesting spatial structures, as well as other forms of spatial heterogeneity (Anselin, 1999). As can be understood, ESDA is characterised by combining statistical analysis with a graphical-geographic-cartographic approach. Thereby, the development of specific modules within GISs has posed a great advancement for these techniques. Within this set of more general techniques that is ESDA, the present work is focused on analysing the phenomenon of spatial dependence or autocorrelation. To this end, the ArcGIS software was used, which, under a geostatistical perspective, allows conducting the analysis of spatial dependence or autocorrelation employing the most commonly used formal indices.

The study variable was the number of travellers that visited Extremadura in June 2015, with the data provided by the Tourism Observatory of Extremadura, using a sample of 270 establishments, of which 131 were hotels and the rest were non-hotel accommodations. The study of spatial autocorrelation or dependence of the mentioned variable in the territory of Extremadura was analysed from a double perspective: global and local. The contrast of spatial dependence in the global perspective was used to identify spatial tendencies or structures in a specific geographic space. To achieve this, the indicators proposed by Moran (1948) and Getis and Ord (1992) were employed. These indicators were the first formulations proposed in the literature as statistical measurements of the effect of spatial autocorrelation. Moreover, they are characterised by their capacity to summarise a general scheme of dependence in a single indicator (Moreno & Vayá, 2000). Both contrasts pose an objective statistical criterion that allows confirming or rejecting the presence of spatial tendencies or structures in the distribution of a variable. In both tests, the null hypothesis to be contrasted was the absence of spatial dependence, that is, the randomness of the variable’s distribution in the selected territory.

Moran's I test (1948) is given by the following formula:

$$I = \frac{N}{S_0} \frac{\sum_{i,j} w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^N (y_i - \bar{y})} \quad i \neq j. \quad (14.1)$$

where;

w_{ij} is the element of the matrix of spatial weights that correspond to the pair (i, j) ;

s_0 is the sum of the spatial weights $\sum_i \sum_j w_{ij}$;

\bar{y} mean or expected value of the variable;

N number of observations.

Upon row standardisation of the matrix of spatial weights $S_0 = N$, index I adopts the following expression:

$$I = \frac{\sum_i \sum_j w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^N (y_i - \bar{y})}. \quad (14.2)$$

According to Cliff and Ord (1981), when the sample size is large enough, it is distributed as a normal standard $N(0, 1)$. The inferential process uses the standardised values (z) of each of them, obtained by the quotient of the difference between the initial value and the theoretical mean divided by the standard deviation, as shown in the following formula:

$$z = \frac{I - E[I]}{SD[I]}. \quad (14.3)$$

The interpretation of the values obtained in the test was carried out in the following manner: non-significant values of the I test led to accept the null hypothesis of the variable's random distribution in the study space. On the other hand, significantly positive values of the variable (above 1.96 at 5% significance level) indicated the presence of positive spatial autocorrelation, that is, they indicated that it was possible to identify values of the variable, high or low, spatially grouped in space to a greater extent than expected if they were following a random pattern. Significantly negative values of the variable (below -1.96 at 5% significance level) indicated the existence of negative spatial autocorrelation, that is, the detection of a non-grouping pattern of similar values (high or low) of the variable, more obvious than expected in a random spatial pattern.

To complete the global analysis of the variable's distribution, the set of indicators proposed by Getis and Ord (1992) were used, which stand out for employing a different criterion to measure spatial autocorrelation, based on the indices of distance and spatial concentration.

The calculation of this index requires the definition of a critical distance (d); from such distance, an influence radius is established, from which it is determined which units are neighbours, based on whether or not they are within the influence radius determined by the critical distance.

Its expression is as follows:

$$G(d) = \frac{\sum_{i=0}^n \sum_{j=0}^n w_{ij}(d) y_i y_j}{\sum_{i=0}^N \sum_{j=0}^N y_i y_j} \text{ for } i \neq j \quad (14.4)$$

where two pairs of spatial units i and j are neighbours if they are within a given distance d , with w_{ij} being 1 when this is the case, or 0 in the opposite case.

The statistical significance is verified through the standardised index z , which is distributed asymptotically according to a normal $N(0, 1)$. The interpretation of this test for those cases that showed statistical significance was the following: a positive (or negative) z value, above 1.96 in absolute value, indicated a tendency of similar high (or low) values to concentrate. Once the distribution pattern of the variable “travellers in the region”, the analysis was completed with the study of local spatial autocorrelation. One of the main limitations of these global autocorrelation tests is that they are unable to detect local spatial structures, i.e. hotspots or coldspots that can or cannot expand to the global pattern structure (Anselin, 1993, 1995b; Getis and Ord, 1992; Moreno and Vayá, 2000; Openshaw, 1993; Tiefeldsdorf and Boots, 1997; Vayá and Suriñach, 1996).

To overcome this limitation, the local spatial autocorrelation tests were developed. The aim of these tests is to detect specially high or low values (hotspots or coldspots) of a variable with respect to its mean values. They are characterised by being calculated for each of the spatial units to be analysed, thus they allow detecting which of these units concentrate higher or lower values than expected in a homogenous distribution. In the study of local spatial autocorrelation, two different scenarios can take place, in contrast with global spatial autocorrelation, as stated by Vayá and Suriñach (1996). In the first scenario, a distribution pattern of concentration or dispersion of values is not detected in a specific space at the global level, while there are small clusters in which high (or low) values of the variable are concentrated. In the second scenario, in the presence of a global distribution pattern, some spatial units contribute to a greater extent to that global indicator. Thereby, the analysis of autocorrelation at the local level is a good complement for the study of global distribution.

Local Indicators of Spatial Association (LISA) proposed by Anselin (1995a, 1995b) and the set of G_i indices of Getis and Ord (1992) and Ord and Getis (1995), are the most commonly used indicators for the study of spatial autocorrelation at the local level. Anselin (1995b) proposed a set of local indicators of spatial association whose purpose is, on the one hand, to determine significant local spatial groups, i.e. clusters, and, on the other hand, to detect spatial instability, understood as the presence of atypical values. Among the indicators proposed by this author, it is worth highlighting Moran’s local I_i index, which is expressed as follows:

$$I_i = \frac{z_i}{\sum_i z_i^2 / N} \sum_{j \in j_i} w_{ij} z_j \quad (14.5)$$

where z_i is the normalised value of spatial unit i , and j_i is the set of neighbour spatial units near i . Under the hypothesis of random distribution, the expectancy of the index is:

$$E_A(I_i) = -\frac{w_i}{N - 1} \tag{14.6}$$

where: w_i is the sum of all the elements of the row of unit i .

The hypothesis that standardised I_i is distributed as a normal $N(0, 1)$ was assumed. The interpretation of the standardised index was performed in the following manner: a positively high z-score value (above 1.96 at 5% significance level) indicated the presence of clusters of high, or low, values of the variable. On the other hand, a significantly negative value (below -1.96 at 5% significance level) indicated the existence of spatial outliers. Getis and Ord (1992) proposed their set of G_i indicators for the analysis of local spatial autocorrelation.

First of all, they proposed the G_i index, which has the following formula:

$$G_i(d) = \frac{\sum_{j=1}^N w_{ij}(d)Y_j}{\sum_{j=1}^N Y_j} \quad j \neq i \tag{14.7}$$

where Y is the variable of interest (not normalised) and $W_{ij}(d)$ are the elements of the contiguity matrix W for an established d distance. Then, the authors proposed an alternative to their index, which includes the observation for which the index is calculated, that is, the previous $j \neq i$ restriction is removed. This new index is expressed as follows:

$$G_i^*(d) = \frac{\sum_{j=1}^N W_{ij}(d)Y_j}{\sum_{j=1}^N Y_j} \tag{14.8}$$

Both indices have two important restrictions. First of all, they can only be used with positive natural variables, and, second of all, they require symmetric contiguity matrices that are not standardised by rows. In order to overcome both limitations, Ord and Getis (1995) respecified their indices with the following expressions:

$$New\ G_i(d) = \frac{\sum_{j=1}^N W_{ij}Y_j - W_i\bar{y}(i)}{S(i)\{[(N - 1)S_{1i}] - W_i^2\}/(N - 2)}^{1/2} \quad j \neq i \tag{14.9}$$

$$New\ G_i^*(d) = \frac{\sum_{j=1}^N W_{ij}Y_j - W_i^*\bar{y}}{\{[(N S_{1i}^*) - W_i^{*2}]/(N - 1)\}^{1/2}} \tag{14.10}$$

where; $S(i)^2 = \frac{1}{N-1} \sum_j (Y_j - \bar{y}(i))^2$; $\bar{y}(i) = \frac{1}{N-1} \sum_j Y_j$; $S_{1i} = \sum W_{ij}^2$ para $j \neq i$; $S_{1i}^* = \sum_j W_{ij}^2$.

As can be observed, these indices were obtained from the standardisation of the previous ones. Once such standardisation was performed, the results obtained

were interpreted in the following manner: a significantly positive (or negative) value indicated the presence of clusters of high (or low) values.

It is important to point out the differences in the interpretation of the obtained results between the two contrasts explained. While the set of G_i indices detect positive spatial autocorrelation, understood as the presence of groups of high values, and negative spatial autocorrelation, as groups of low values, Moran's I_i index allows identifying also spatial outliers. In this case, the diagnosis of positive spatial autocorrelation is understood as the presence of groups of similar values, either high or low, and negative spatial autocorrelation involves the existence of dissimilar values grouped in space. Therefore, as was the case for global indices, the combined calculation of both indices contributes to complete and enrich the study of local spatial autocorrelation. Once the different contrasts to be used in the present study have been analysed, the next section presents the main results obtained from the analysis of spatial autocorrelation, at a local and global level, of the variable "number of travellers that visited Extremadura".

14.4 Results

The present study was based on a sample of 270 establishments in the region of Extremadura (Spain), which provided their data of travellers who used their facilities in July 2015. The exploratory analysis of the distribution pattern of the variable "number of travellers" in Extremadura was conducted, firstly, from a general perspective, including the total of travellers that visited the region regardless of the type of accommodation used. In a second phase, this first analysis was completed with the study of the distribution of travellers based on the type of accommodation used, considering two large groups: hotels and non-hotel accommodations. The reason why it was decided to conduct this separation was the difference in the accommodation capacity of each of the establishment types to be analysed, which could influence the results depending on the variable of interest used.

For the thorough analysis of the distribution pattern of the variable in the whole territory, the most commonly used global contrasts of spatial autocorrelation were employed: Moran's I and Getis and Ord $G(d)$. To carry out the analysis, the ArcGIS 10.3 software was used. This software works from a geostatistical perspective and in its Spatial Statistics Tools module it allows analysing spatial autocorrelation with the most commonly used formal indices. To conduct the analysis from a global perspective, it was decided to specify the neighbourhood relationship according to the criterion of reverse distance, using euclidean distance as a method and proceeding to the row standardisation of the matrix of spatial weights. Moreover, in both contrasts, the null hypothesis to verify is the random spatial distribution of the variable in the whole territory. The following table shows the results obtained from the global Moran's I test and Getis and Ord's general G using the ArcGIS software for the analysis of travellers in Extremadura.

Table 14.1 Results of global Moran's *I* test and Getis and Ord's *G* in the total number of travellers in Extremadura

Global Moran's I		General G summary	
Moran's index	0.180913	Observed general G	0.006051
Expected index	-0.003717	Expected general G	0.003717
Variance	0.001806	Variance	0.000000
z-score	4.344949	z-score	4.131847
p-value	0.000014	p-value	0.000036

Source Own elaboration using ArcGIS 10.3

As can be seen in the obtained results in the Table 14.1, the null hypothesis of random distribution of the variable in the whole study region is rejected with 1% significance level. Considering also Moran's index and z-score, it can be concluded that the distribution of travellers in the entire territory shows a concentration of similar values of that variable. In other words, the values of the variable "travellers" tend to concentrate in high or low values in the study region. To confirm the data obtained, and also to amplify their information, the next step was to carry out the contrast proposed by Getis and Ord with their general *G*(d) to analyse the distribution pattern of the number of travellers in the whole of the territory studied.

As was the case with Moran's test, the p-value obtained indicated that the hypothesis that the travellers are randomly distributed in the territory of Extremadura had to be rejected, with 1% significance level. From the analysis of the *G* value and z-score, it was also concluded that the variable tends to concentrate in space in high values. Therefore, in view of the obtained results, it can be concluded that the travellers that visit Extremadura are not randomly distributed in the territory, that the differences found are statistically significant and that the variable tends to concentrate in high values in this region. In order to enrich the results obtained from this analysis, it was decided to separate the data into two large groups. On the one hand, hotels, for which a subsample of 131 establishments was created, and, on the other hand, non-hotel accommodations, whose subsample consisted of a total of 139 establishments.

The reason for this division was the inherent differences between the two types of establishment, which are mainly due to their different guest capacity. Therefore, the next step was to conduct the same analysis for each of the mentioned establishment types into which the sample was divided. Firstly, for hotels, the following results were obtained.

The results obtained in Moran's *I* test, Table 14.2, show that, in the case of travellers who stay in hotels, the null hypothesis of random spatial distribution of the variable must be rejected, with 1% significance level. Furthermore, the z-score shows a positive value, which indicates that the values of the variable tend to follow a concentration pattern of similar values in space. Likewise, the Getis and Ord's test confirmed the obtained results. As can be observed, the p-value suggests the need to reject the null hypothesis of random distribution of the number of travellers in

Table 14.2 Global Moran’s I test and Getis and Ord’s general G (d) for travellers in hotels

Global Moran’s I		G general summary	
Moran’s index	0.133746	Observed general G	0.009735
Expected index	-0.007692	Expected general G	0.007692
Variance	0.002885	Variance	0.000001
z-score	2.633144	z-score	2.493937
p-value	0.008460	p-value	0.012634

Source Own elaboration using ArcGIS 10.3

Table 14.3 Global Moran’s I test and Getis and Ord’s general G (d) for travellers in non-hotel establishments

Global Moran’s I		General G summary	
Moran’s index	0.022905	Observed general G	0.008571
Expected index	-0.07246	Expected general G	0.007246
Variance	0.002568	Variance	0.000008
z-score	0.594973	z-score	0.479609
p-value	0.551861	p-value	0.631506

Source Own elaboration using ArcGIS 10.3

the whole territory, with 5% significance level. Moreover, the interpretation of the z-score indicates that travellers in hotels tend to concentrate spatially in high values.

In order to confirm whether the distribution pattern detected up to this point according to the spatial autocorrelation tests was the same for the two establishment types into which the sample was subdivided, the present analysis of global distribution was completed with the study of the variable “number of travellers that stayed in non-hotel accommodation establishments”.

In view of the results obtained, Table 14.3, in the contrasts conducted, in both cases, the null hypothesis of random distribution of the variable cannot be rejected. Therefore, it is not discarded that travellers who stay in non-hotel establishments are randomly distributed in the analysed territory. Considering the different results obtained, it can be asserted that the concentrated distribution pattern of travellers in the region is motivated by the behaviour of the travellers who choose hotels as the accommodation option. As has been described in the previous section, the study of autocorrelation at the local level helps explaining the results obtained in the analysis at the global level, allowing to determine whether the concentration or dispersion pattern is governed by some specific spatial units, or detecting the concentration of units affected by spatial autocorrelation that were not identified in the global study.

Thereby, in order to complete this analysis, the next step was to conduct the study at the local level, using the local Moran’s I_i , and the New G_i^* to carry out the Getis and Ord’s analysis of hotspots and coldspots. First, Fig. 14.1 shows the results of

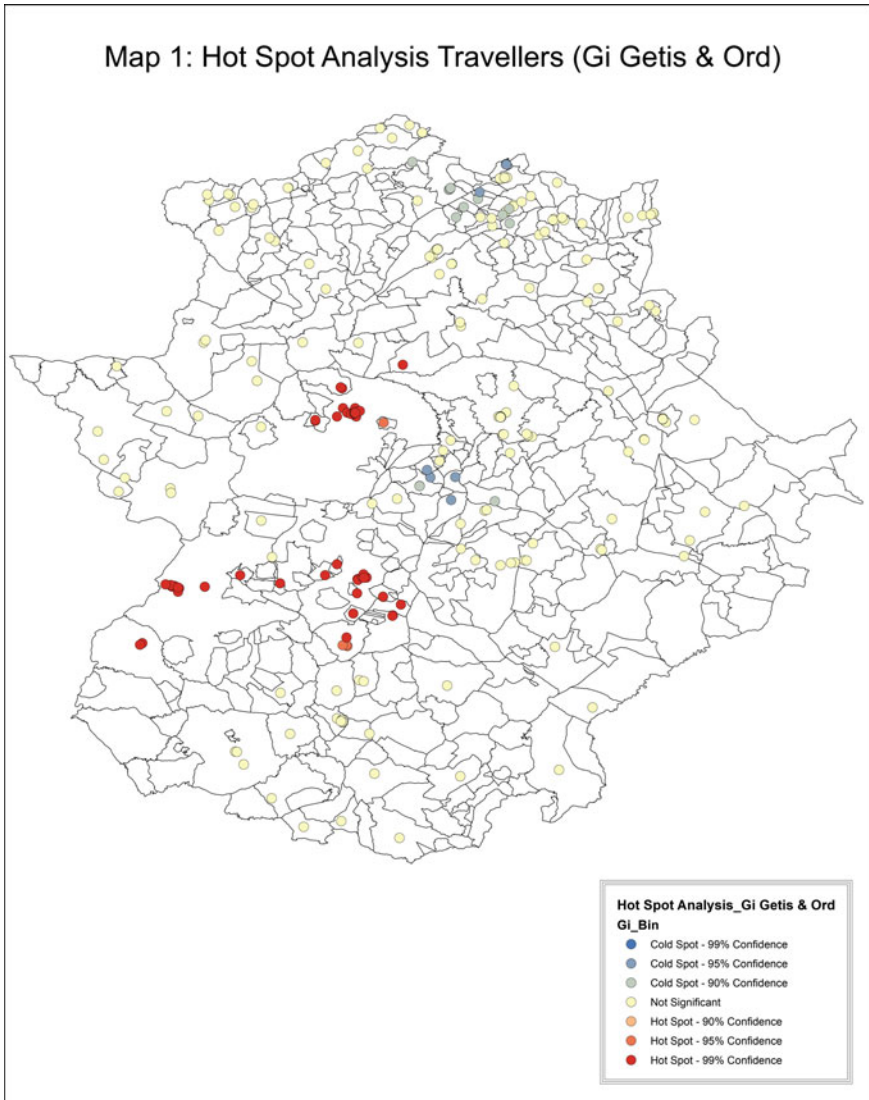


Fig. 14.1 Map hot spot analysis travellers (G_i Getis & Ord). *Source* Own elaboration using ArcGIS 10.3 (Color figure online)

the local Getis and Ord's test, indicating the hotspots and coldspots of high and low values of the number of travellers in the different points of the territory for all the establishments of the sample. In all the cases analysed, the fixed distance band criterion was used to establish the neighbourhood relationship, using the euclidean distance as a reference.

As can be seen in the map, there are some hotspots and coldspots at the local level in specific areas of the territory. Thus, there are hotspots between the two main cities: Merida and Badajoz. Their influence area reaches the closest towns, where the existence of hotspots is detected, with 99% confidence level, in which the concentration of travellers is higher than expected in a random distribution pattern. The same case was diagnosed in the third main city of the region, Caceres, where the existence of hotspots of the variable was detected, whose influence reaches the closest towns.

On the other hand, there were also units in which the variable analysed showed significantly low values. Thus, the existence of a series of coldspots was detected in the centre of Extremadura, in the area of Abertura, Almoharin and Montanechez towns. More surprising was the case of the north of the province of Caceres, which despite the strong tourist tradition rooted in the study areas, it showed the presence of coldspots of values of the variable when the number of travellers was jointly analysed, i.e. without separating the two types of establishment.

The next step was to break down the number of travellers according to the two large groups of accommodation options, hotels and non-hotel establishments, to verify whether the obtained results remained the same or whether this division would unravel new evidence about the local distribution pattern. To this end, the analysis conducted in each of the two types of establishment was replicated. Figure 14.2 shows the results obtained from the Getis and Ord's analysis of hotspots and coldspots when considering as the variable of interest the number of travellers who stayed in hotels in Extremadura.

Figure 14.2 shows how the location of hotspots and coldspots changes the distribution when only considering the travellers who stay in hotels, thus it becomes obvious that the type of establishment affects the distribution of the variable. The analysis identified the presence of spatial units in which the variable was concentrated in high values, mostly around three of the main cities of the region, Caceres, Merida and Badajoz. As can be observed, the presence of high values is detected with different significance levels, which also occurred in the combined analysis. Furthermore, the map shows that in the area of La Serena (in blue) there are cold points of values of the variable.

The main differences in the analysis of the distribution of the variable "travellers in hotels", with respect to the previous combined analysis of both accommodation options, is the larger number of hotspots in the main cities and the disappearance of coldspots in the northern area of the region. In order to delve further into the obtained results, the next step was to analyse the distribution of the travellers who stayed in non-hotel establishments, whose results are shown in Fig. 14.3.

Finally, when analysing the hotspots and coldspots in the distribution of the variable considering only the travellers who stayed in non-hotel establishments, the results reveal the inherent differences between the two types of accommodation analysed. Thus, in the distribution of travellers in non-hotel establishments there are hotspots in the north of the region, which are close to the coldspots identified in the combined analysis of the entire sample. These differences may be due to the large

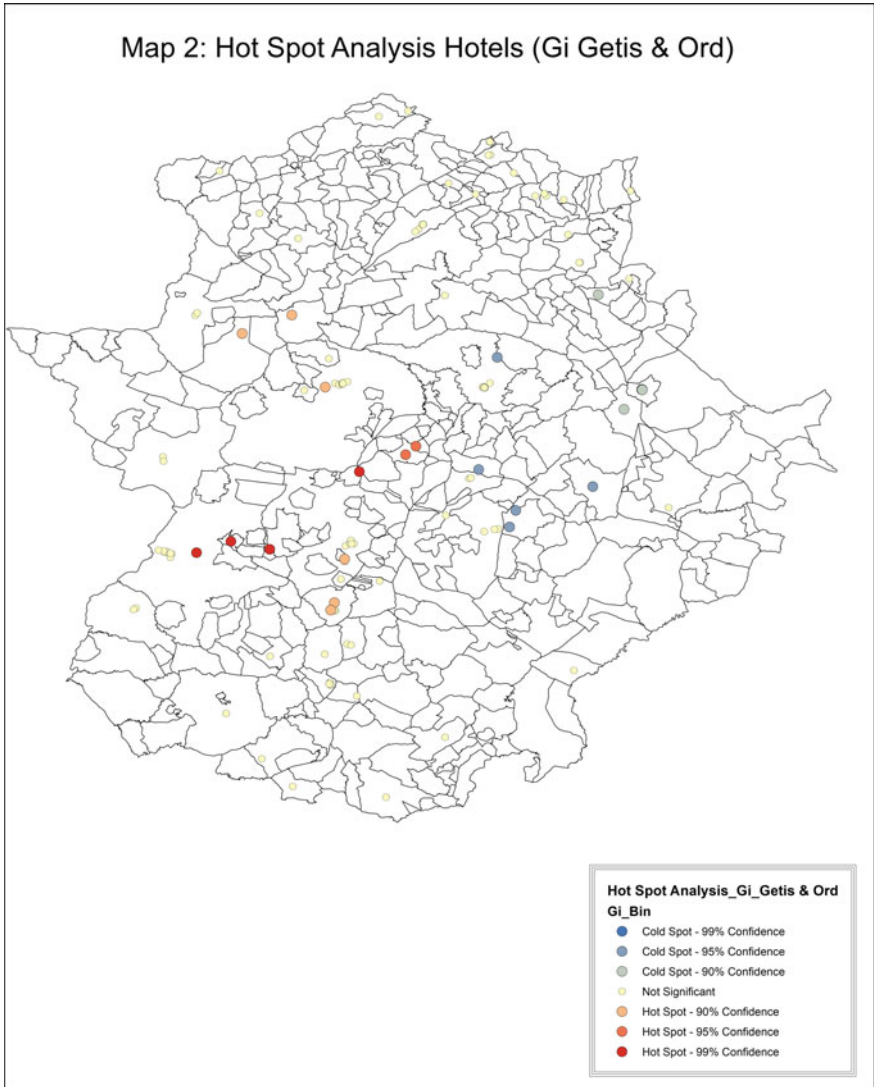


Fig. 14.2 Map hot spot analysis hotels (G_i Getis & Ord). *Source* Own elaboration using ArcGIS 10.3 (Color figure online)

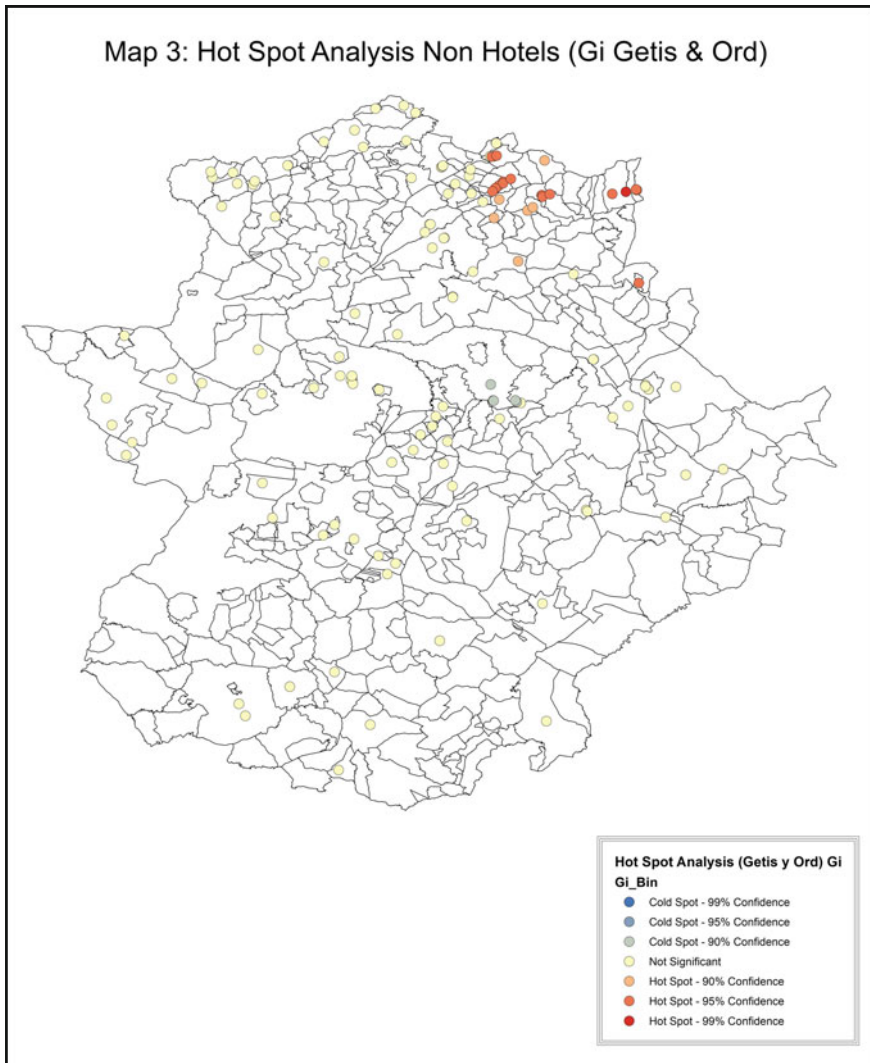


Fig. 14.3 Hot spot analysis non hotels (G_i Getis & Ord). *Source* Own elaboration using ArcGIS 10.3 (Color figure online)

number of non-hotel establishments in this area, which constitute the main type of accommodation offered.

The results obtained in Getis and Ord's new G_i^* test are complemented with the local Moran's I_i autocorrelation test, which allowed to identify spatial outliers and clusters of high and low values. To this end, the analysis of local spatial autocorrelation was conducted using Moran's I_i index for the total number of travellers who visited the region. This was achieved by specifying the neighbourhood relation-

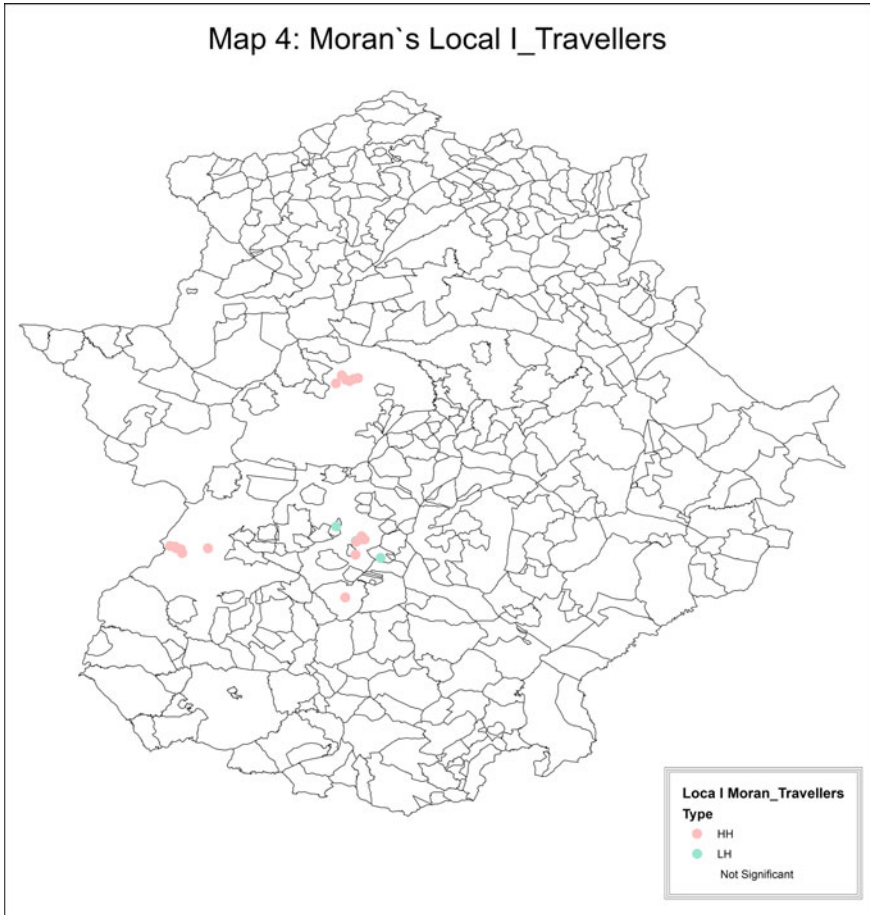


Fig. 14.4 Moran's local I travellers. *Source* Own elaboration using ArcGIS 10.3 (Color figure online)

ship with the same criteria that were used for the global correlation, that is, reverse distance, with the method of euclidean distance and the row standardisation of the matrix. The obtained results are shown in Fig. 14.4.

The map shows the existence of several clusters of high values in three of the main cities of the region, Badajoz, Caceres and Merida, which also occurred in the diagnosis performed with Getis and Ord's G_i^* index. On the other hand, the analysis also detected spatial outliers of low-high values, in the vicinity of Merida. Once again, the analysis was conducted dividing the sample into the two accommodation types chosen by the travellers, in order to determine whether the separate analysis could add more information to the obtained results.

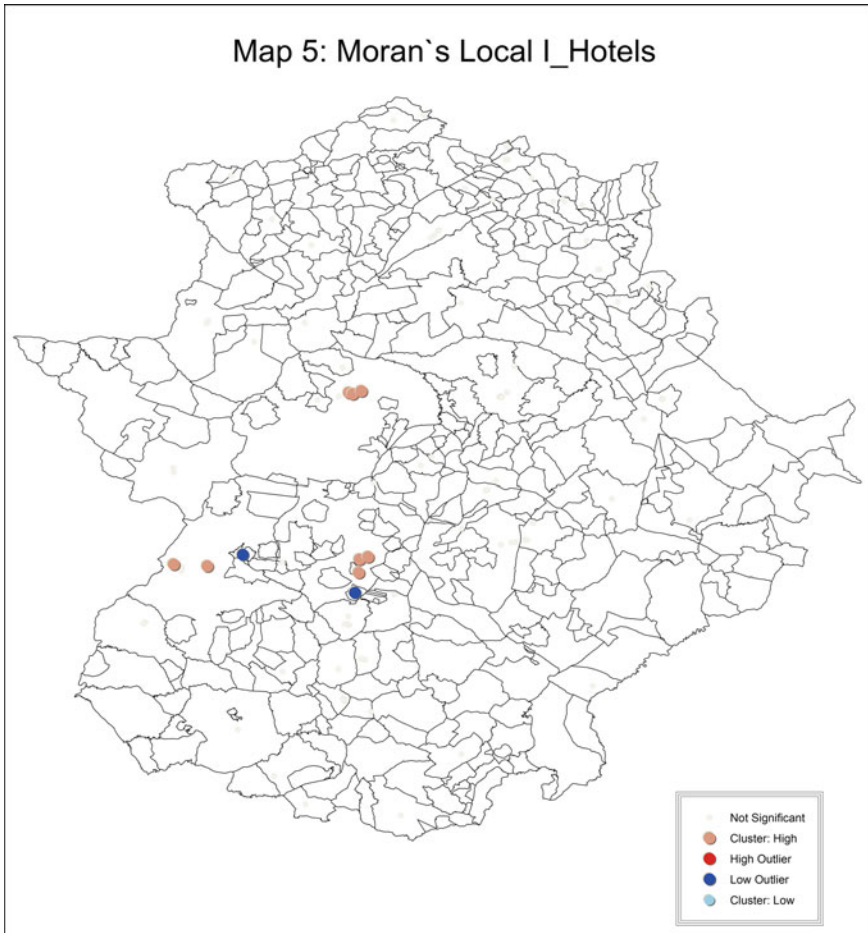


Fig. 14.5 Moran's local I in hotels. *Source* Own elaboration using ArcGIS 10.3 (Color figure online)

First, following the same order, the analysis of the distribution of travellers who stayed in hotels was conducted, whose results are shown in Fig. 14.5.

The map shows the similarities of the obtained results with respect to those identified in the analysis conducted with the travellers considering both accommodation options. The analysis confirmed the existence of clusters of high values in three of the main cities of the region, as well as the presence of outliers of low values surrounded by high values in towns near the cities of Badajoz and Merida.

To conclude the present analysis of spatial autocorrelation at the local level, the local Moran's I test was conducted for the travellers who stayed in non-hotel establishments. The results are shown in Fig. 14.6.

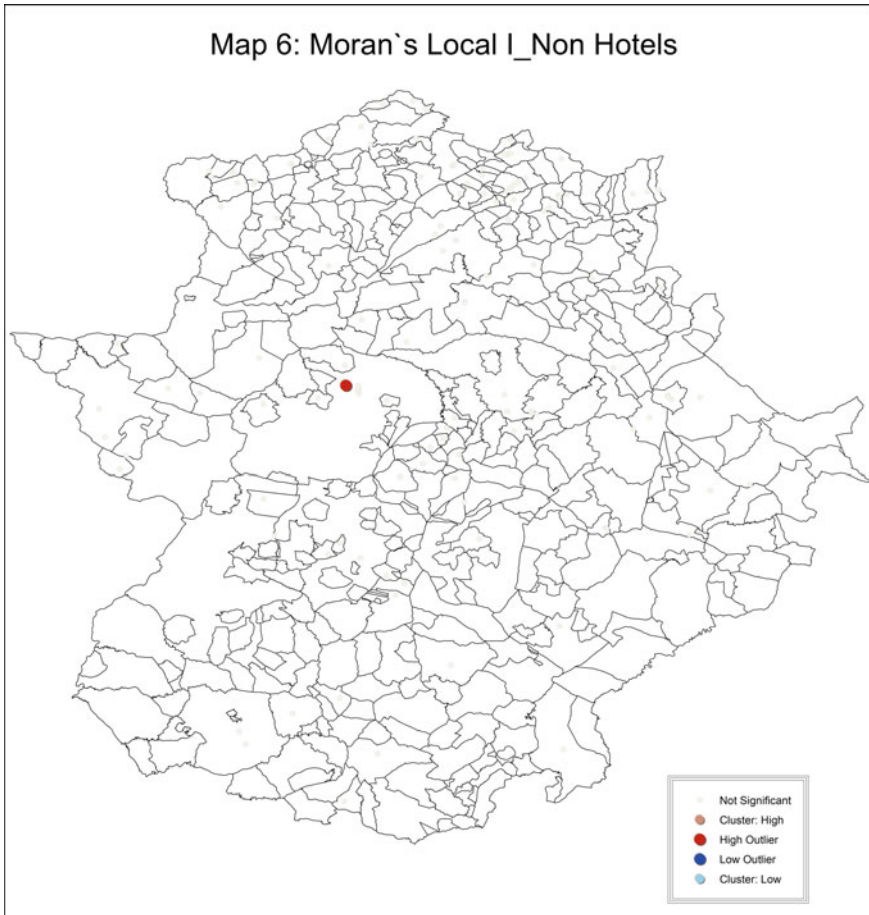


Fig. 14.6 Moran's local I_i in non hotels. *Source* Own elaboration using ArcGIS 10.3 (Color figure online)

Lastly, the map shows the results obtained in the local Moran's I_i autocorrelation test, with only one spatial outlier of high values surrounded by low values in the city of Cáceres.

14.5 Conclusion and Implications

The evolution of GIS technologies, along with the greater availability of georeferenced data, makes it possible for the spatial analysis of data to have a greater dissemination in different fields of social science. Thus, the previous analysis of the

distribution of the variables, especially those strongly related to their development in a specific geographic territory, is essential. Tourism stands out among them for being strongly related to the area in which it is developed.

The results of the spatial autocorrelation study conducted for the variable “number of travellers”, both for the combination of the two types of accommodation chosen by them and separately, showed that this variable does not have a random distribution, neither in the entire territory analysed (global spatial autocorrelation), nor in the individualised study of the different spatial units (local spatial autocorrelation). Thereby, the next step was to obtain more detailed information on how the variable was distributed in the region.

First, the results of the tests conducted for the study of the distribution pattern of the variable in the entire territory analysed showed that this distribution tends to concentrate in high values in space. Separating the sample by the type of accommodation chosen allowed detecting that this pattern is governed by the distribution of travellers who stayed in hotels, since in the case of those who stayed in non-hotel accommodations the hypothesis of random distribution of the variable cannot be rejected.

Secondly, the study of autocorrelation at the local level revealed that the non-uniform distribution of the variable in space could be due to the contribution of the values reached in three of the main cities of the region, Badajoz, Caceres and Merida, over the rest of the values. The high values reached in these specific points could be the ones contributing strongly to the fact that the global pattern is not uniform in the whole of the territory.

On the other hand, from the results obtained in this analysis, it can also be concluded that, although at the global level the hypothesis of random distribution of the variable “travellers in non-hotel establishments” was not rejected, it shows hotspots in the region. Specifically, it was observed that in the north of Extremadura, in the area of La Vera and El Valle de Ambroz, the variable shows higher values than expected in a random distribution of this variable.

In conclusion, and regardless of the particular values obtained, it is concluded that the study variable “number of travellers” does not follow a random pattern. In other words, each of the spatial units is spatially interdependent. This means that once the presence of a spatial effect is detected in the exploratory phase, such as spatial autocorrelation, it must be taken into account that further modelling, in which such variable is included, as well as any confirmatory analysis, will require considering the techniques proposed by spatial statistics for the treatment of this effect.

For further research, the authors consider that it would be interesting to use a wider time frame. It must be taken into account that the spatial relationships detected took place with the values of the variable in a specific month (July), thus it cannot be asserted that these are constant in time. A panel data analysis, including all the other months of the year, would greatly enrich the analysis conducted. Likewise, it would significantly benefit the analysis to repeat it with other variables that are typically associated with tourism, such as the degree of occupation, overnight stay, average stay, etc. The analysis of the spatial distribution of these variables that help

describing the evolution of an activity could also help to understand how they are related.

References

- Anselin, L. (1980). *Estimation methods for spatial autoregressive structures*. Regional science dissertation and monograph Series, Ithaca, NY.
- Anselin, L. (1988). *Spatial econometrics: Methods and models*. The Netherlands: Kluwer Academic Publisher.
- Anselin, L. (1992). *SpaceStat tutorial. A workbook for using SpaceStat in the analysis of spatial data* (Technical Reports S-92-1). Santa Barbara: University of California; National Center for Geographic Information and Analysis.
- Anselin, L. (1993). *The Moran scatterplot as an ESDA tool to assess local instability in spatial association*. GISDATA Specialist Meeting on GIS and Spatial Analysis, Amsterdam, The Netherlands.
- Anselin, L. (1995a). *SpaceStat version 1.80: Users' guide*. Regional Research Institute West Virginia University: Morgantown.
- Anselin, L. (1995b). Local indicators of spatial association (LISA). *Geographical Analysis*, 27(2), 93–115.
- Anselin, L. (1999). The future of spatial analysis in the social sciences. *Geographic Information Sciences*, 5(2), 67–76.
- Anselin, L. (2001). Spatial econometrics. In Baltagi (Ed.), *A companion to theoretical econometrics* (pp. 310–330). Oxford: Basil Blackwell.
- Anselin, L., & Florax, R. (1995). New directions in spatial econometrics: Introduction. In L. Anselin & R. J. G. M. Florax (Eds.), *New directions in spatial econometrics* (pp. 3–18). Springer.
- Anselin, L., & Rey, S. (1997). Introduction to the special issue on spatial econometrics. *International Regional Science Review*, 20, 1–8.
- Arbia, G. (1989). *Spatial data configuration in statistical analysis of regional economics and related problems*. Dordrecht: Kluwer.
- Barros, C. P., & Matías, A. (2007). Efficiency in a chain of small hotels with a stochastic production frontier model. In A. Matías, P. Nijkamp, & P. Neto (Eds.), *Advances in modern tourism research: Economic perspectives* (pp. 107–129). Heidelberg: Physica-Verlag.
- Bloommestein, H. (1983). Specification and estimation of spatial econometric models: A discussion of alternative strategies for spatial economic modelling. *Regional Science and Urban Economics*, 13, 130–251.
- Button, K., & Kulkarni, R. (2001). *Spatial and distance statistics of the trucking and warehousing industries using GIS tools*. 41th Congress of the European Association of Regional Science (ERSA) Zagreb.
- Chasco, M. C. (2003). *Econometría espacial aplicada a la predicción-extrapolación de datos microterritoriales*. Madrid: Consejería de Economías e Innovación Tecnológica de la Comunidad de Madrid.
- Chasco, C., & Vicens, J. (2000). *EU-membership impacts in the Spanish province income convergence: A spatial autocorrelation perspective*. Oviedo: Anales de Economía aplicada XIV Reunión de ASEPELT.
- Cliff, A., & Ord, J. (1973). *Spatial autocorrelation*. London: Pion.
- Cliff, A., & Ord, J. (1981). *Spatial processes, models and applications*. London: Pion.
- Geary, R. (1954). The contiguity ratio and statistical mapping. *The Incorporated Statistician*, 5, 115–145.
- Getis, A., & Ord, J. (1992). The analysis of spatial association by use of distance statistics. *Geographical Analysis*, 24, 189–206.

- Kmenta, J. (1971). *Elements of econometrics*. New York: MacMillan.
- López, F. A., Palacios, M. A., & Ruiz, M. (2001). Modelos explicativos del desempleo en términos de localización: una aplicación a las provincias españolas. *Anales de Economía Aplicada*, XV Reunión de ASEPELT. Santiago de Compostela.
- Ma, T., Hong, T., & Zhang, H. (2015). Tourism spatial spillover effects and urban economic growth. *Journal of Business Research*, 68, 74–80.
- Moran, P. (1948). The interpretation of statistical maps. *Journal of the Royal Statistical Society*, 10, 243–251.
- Moreno, R., & Vayá, E. (2000). *Técnicas econométricas para el tratamiento de datos espaciales: la econometría espacial* 44. Universitat de Barcelona.
- Moreno, R., & Vayá, E. (2004). Econometría Espacial: nuevas técnicas para el análisis regional. Una aplicación a las regiones europeas. *Investigaciones regionales* 1, 83–106.
- Openshaw, S. (1993). Some suggestions concerning the development of artificial intelligence tools for spatial modelling and analysis in GIS. In M. M. Fisher & P. Nijkamp (Eds.), *Geographic information system, spatial modelling and policy evaluation* (pp. 17–33). Berlin: Springer Verlag.
- Ord, J. K., & Getis, A. (1995). Local spatial autocorrelation statistics: Distributional issues and an application. *Geographical Analysis*, 27(4), 286–306.
- Paelinck, J. H. P., & Klaasen, L. H. (1979). *Spatial econometrics*. Saxon House.
- Pavlyuk, D. (2010). Multi-tier spatial stochastic frontier model for competition and cooperation of European airports. *Transport and Telecommunication*, 11(3), 57–66.
- Pavlyuk, D. (2013). Distinguishing between spatial heterogeneity and inefficiency: Spatial stochastic frontier analysis of European airports. *Transport and Telecommunication*, 14(1), 29–38.
- Sánchez, M. (2008). Análisis Espacial de Datos y Turismo: Nuevas Técnicas para el Análisis Turístico. Una aplicación al caso extremeño. *Revistas empresariales. Segunda Época*, 2, 48–66.
- Sánchez, J. M., Sánchez, M., & Rengifo, J. I. (2013). La evaluación del potencial para el desarrollo de turismo rural. Aplicación metodológica sobre la provincia de Cáceres. *Geofocus*, 13(1), 1–20.
- Sánchez, J. M., Sánchez, M., & Rengifo, J. I. (2018). Patrones de distribución de la oferta turística mediante técnicas geoestadísticas en Extremadura (2004–2014). *Boletín de la Asociación de Geógrafos Españoles*, 76, 276–302.
- Shoval, N., & Ahas, R. (2016). The use of tracking technologies in tourism research: The first decade. *Tourism Geographies*, 18(5), 587–606.
- Student. (1914). The elimination of spurious correlation due to position in time or space. *Biometrika*, 10, 179–180.
- Tiefelsdorf, M., & Boots, B. (1997). A note on the extremities of local Moran's I_i and their impact on global Moran's I . *Geographical Analysis*, 29(3), 249–257.
- Vayá, E., & Suriñach, J. (1996). Contrastes de autocorrelación espacial: una aplicación al ámbito de las provincias españolas. X Reunión ASEPELT, Albacete (Castilla la Mancha).
- Zhou, Y., Maumbe, K., Deng, J., & Selin, S. (2015). Resource-based destination competitiveness evaluation using a hybrid analytic hierarchy process (AHP): The case study of West Virginia. *Tourism Management Perspectives*, 15, 70–80.