

Geostatistics and Digital Social Networks - A study of tourism dynamics in “Alta and University of Coimbra” (UNESCO world heritage site)

José Gomes dos Santos* and **
 Liliana Raquel Simões de Azevedo*
 Joaquim António Saraiva Patriarca***
 Luis Carlos Roseiro Leitão** and ****

*Geography and Tourism Department, Faculty of Arts and Humanities, University of Coimbra, Coimbra – PT

**Center of Studies on Geography and Spatial Planning (CEGOT), Coimbra – PT

***Institute for Systems Engineering and Computers at Coimbra, Coimbra – PT

****Institute for Nature Conservation and Forestry (ICNF) – PT

{jgs966, lrsazevedo, joaquimaspatriarca}@gmail.com e {luis.leitao@icnf.pt}

ABSTRACT – Spatial modeling in Geographic Information Systems (GIS) always involves choices. The existence of constraints, either of a financial nature or related to the specifics of the software itself, to the algorithms, the uncertainty and even the reliability of the data, the purposes and the applications of the studies, make this a kind of guiding compass for GIS analysts. Building on a previous exercise of data acquisition (check-ins) based on two Digital Social Networks (DSN — Facebook and Foursquare) and on the awareness of the use of voluntary geographic information generated by tourists sharing their topophilic ties through DSN, the present analysis aims to evaluate the contribution of modern techniques of spatial analysis applied to tourism in the “Alta and University of Coimbra” area. Concepts and procedural tasks related to density determination, cluster analysis and identification of patterns associated with regionalized variables have thus been implemented with the purpose of evaluating and comparing the results obtained through the application of two techniques of spatial analysis, Kernel Density Estimation (KDE) and Optimized Hot-Spot Analysis (OHSA) & Inverse Distance Weighting (IDW) Interpolation.

Key words: Geostatistics, Spatial Modeling, KDE, OHSA, IDW, Tourism, Coimbra

1. INTRODUCTION

Modeling and Spatial Analysis as geostatistical methodologies are currently used in such varied fields as the communication strategies of numerous economic and scientific/research activities, both in the state and private sector, and applied to many different branches of knowledge, from the biomedical to the spatial sciences and, of course, the geosciences. They resort to a set of tools that allow for the construction of geospatial models and the inclusion of predictive capabilities conducive to the shaping of a system aimed at explaining both natural and social phenomena. The present study builds on a previous, broader research project that sought to compare and assess the reliability of three geostatistical algorithms, two of which in association (complement), applied to the study of tourism dynamics in the “Alta” and University of Coimbra area, *i.e.*, the uptown area and University of Coimbra main campus, in Portugal’s Center region. The study seeks to correlate variables, attributes and their respective interactions, and we concur with Getis’s (1991) notion that the family of spatial interaction models is a special instance of the broader models used in spatial autocorrelation.

2. STATE OF ART OVERVIEW AND FURTHER CONTEXT

According to Matheron (1965, *apud* Andriotti, 1988), who is usually credited with the founding of Geostatistics and the development of related concepts, “Geostatistics is the application of the formalism of Random Functions to the reconnaissance and estimation of natural phenomena”. Following on the conceptual contributions to Geostatistics made by Matheron in the mid-1960s and developed in the work of more recent authors like Önsoy & Bocquillon (2015, revised reissue), Andriotti (*op.cit.*) argues that this area of knowledge can be viewed as a practical application of the Regionalized Variable Theory. A “regionalized variable” – a term coined by Matheron – is a dipole of (seemingly contradictory) features and formed by variables whose behavior is at once random and structured. It is random in that the values of one’s measurements can vary considerably within the same sample, yet the behavior pattern of the variable cannot fail to evince a regionalized dimension structured according to some kind of spatial law, assuming that (as can be eas-

ily understood) no sample values are entirely independent from their respective geographic location. With the necessary adaptations, the present study applies Getis & Ord's 1992 paper on "sudden death syndrome by county in North Carolina" and its potential correlation with the "price of housing units sold by zip-code district in the San Diego metropolitan region." It is also based on a kind of first corollary of the "First Law of Geography" (Tobler, 1970), according to which the behavior of a regionalized variable is more likely to show similarities in "the neighborhood" of a given point (a known, geolocated point, about which we have attribute data), than in a geographically distant location. What this postulate means is that, in accordance with Probability Theory and with Andriotti (*op. cit.*), the values of two spatially close samples must be correlated. This is why it will be impossible to study numerical values as independent of each other, which is to say, in light of classical statistical methods alone. Regionalization as something that is reflected in the structured nature of phenomena will thus be best articulated in the language of Random Functions. However, one must point out that these concepts got it wrong in initially assuming an overly naturalistic bent, because today's knowledge in the area of Geostatistics lends itself to a treatment similar to that of other types of anthropic variables, as is the case with the areas of Health, Criminology, Road Accident Monitoring, and Tourism, among others.

The relevance of this type of data and of the use of these techniques in urban planning, based on tourism demand of sites of great historical and scenic value, is attested to by other recent studies, notably by García- Palomares, Gutierrez and Mínguez (2015) and by Zhihui *et al.* (2016). As far as the former is concerned, we would like to single out the data acquisition processes, as they bear close similarities to the methodological procedures of the present study (check-in data mining), namely with regard to the use of GIS-processed data on urban tourist destinations collected through photo-sharing services. But the work carried out by Zhihui *et al.* also shares procedural and methodological affinities with our own, namely with regard to its comparative analysis of patterns that identify potential tourist hot spots and its application to European metropolises of data obtained from social networks.

3. METHODOLOGY AND OPERABILITY

Working with spatial models involves choices. Then there are always constraints, either of a financial nature or related to the specifics of the software itself, to the algorithms being used, to the level of risk and uncertainty, etc. Spatial modeling encompasses all these components and we should be aware of this fact if we wish to be in a position to discuss the best solutions for each case. To these initial thoughts one should add the words of George E. P. Box and Norman Richard Draper when they wrote in 1987: "Remember that all models are wrong; the practical question is how wrong they have to be to *not* be useful. Essentially, all models are wrong, but some are useful (*apud* Rocha 2012, 323).

Although the main goal of this comparative analysis applied to tourism in the "Alta" and University of Coimbra area is not to interpret data in light of the classic conceptions of descriptive statistics, we sought to maintain some data that characterize the sample universe and make it possible to anticipate a number of properties among the variables at play, whose dominant feature is a strong independence and, as a consequence, a considerable degree of randomness. But the availability of modern spatial analysis techniques, such as density determination, cluster analysis, or pattern identification for turning native vector data into intelligent rasters, led us to test the data resulting from comparing two methods of density analysis, one of which in combination with a geostatistical interpolator – **Kernel Density Estimation (KDE) and Optimized Hot-Spot Analysis (OHSA) & Inverse Distance Weighting (IDW) Interpolation**. This is definitely the domain of "Geostatistics" and the exploratory and confirmatory techniques that lie at the "core of a spatial analysis module for a GIS" (Anselin & Getis, 1992), and on that basis we should be able to identify patterns as well as to **estimate, delimit** and possibly **predict** areas with **homogeneous characteristics** in terms of behavior.

3.1. Data preparation for cluster visualization: Point Density and Cluster and Outlier Analysis (*Anselin Local Moran's I*)

The starting point of the present study consisted in identifying areas with a higher density of points shared by tourists in DSNs (check-ins). Given that some of the Points of Interest (POI) are in very close proximity, we decided it would be useful for our work to identify the areas with the highest density (Point Density). We opted for a pixel value of 10m and set the search radius at 140m, rounding off the proposed default value. The sample was later reclassified into 25m classes. The final result is shown in Figure 1. A cluster of points can easily be seen to stand out from the entire neighborhood, forming a spatial unit that can be defined as the whole area comprising the various (and major) sites of the University of Coimbra and Museums, plus the historical monuments surrounding both the Old and New Cathedrals ("Sé Velha" and "Sé Nova"). It should be pointed out that we focused on an 86 point (POI) Extent, *i.e.*, locations checked in by DSN users (Azevedo, 2014), even if in the case at hand, and given that the methodology only includes

“Falls inside” points, two of them (“Choupal” and Coimbra-B train station) ended up excluded from the cartographic representation. For more complete information, we have used the *Anselin Local Moran's I* statistic for cluster-specific mapping¹, according to a workflow (also included in ArcMAP 10.3)² that followed the path, ... \ **Arctoolbox** \ **Spatial Statistics tools** \ **Mapping clusters** \ **Cluster and Outlier Analysis (Anselin Local Moran's I)**, Figure 2.

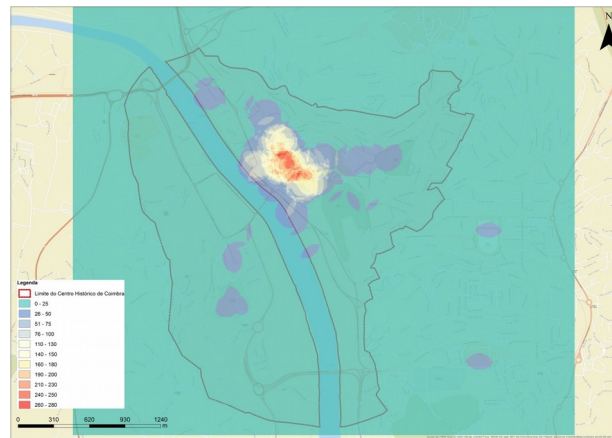


Figure 1 – Point density map.



Figure 2 – Workflow for identification of clusters in point themes (adapted from the original: ESRI Help).

In the Configuration Wizard we created a “Checkin_Final” layer (point theme), with the attribute SumFF (number of check-ins in each POI). We selected *Inverse distance* for the *Conceptualization of Spatial Relationships* field, and *Euclidean Distance* for the *Distance method* field. As can be readily gathered from the results shown in Figure 3, object identification remains very limited, and so the influence of location and distance on the decision to visit the remaining POIs is left unclear, although it is possible that the reason lies in such factors as neighborhood, proximity, and iconographic standing. However, we still fail to understand the reasons underlying the large variance in the number of Facebook and Foursquare check-ins (SumFF, an attribute included in the configuration). On the other hand, it is not possible to identify actual hot spots and cold spots, since the entire “Alta” area operates as a cluster in which no POI stands out from the neighborhood but is rather very much a part of it.

When we open the theme attribute table *Checkin_final_Clusters_Outlier* (Figure 4), we can see that new fields with “z-scores” and “p-values” have been created³. The results thus obtained show the existence of a considerable number of points whose check-in attribute value, their location and physical proximity notwithstanding, are not indicative of cluster formation.

¹ (...) new general class of Local Indicators of Spatial Association – LISA (Anselin, 1995).

² According to the *ESRI Help Toolset*, “Given a set of weighted features, [cluster and outlier analysis] identifies statistically significant hot spots, cold spots, and spatial outliers using the Anselin Local Moran's I statistic.”

³ “Z-scores are simply standard deviations. If, for example, a tool returns a z-score of +2.5, you would say that the result is 2.5 standard deviations. Both z-scores and p-values are associated with the standard normal distribution. The p-value is a probability. For the pattern analysis tools, it is the probability that the observed spatial pattern was created by some random process. When the p-value is very small, it means it is very unlikely (small probability) that the observed spatial pattern is the result of random processes, so you can reject the null hypothesis.” (Source: ArcGIS Help Desktop 10.3).

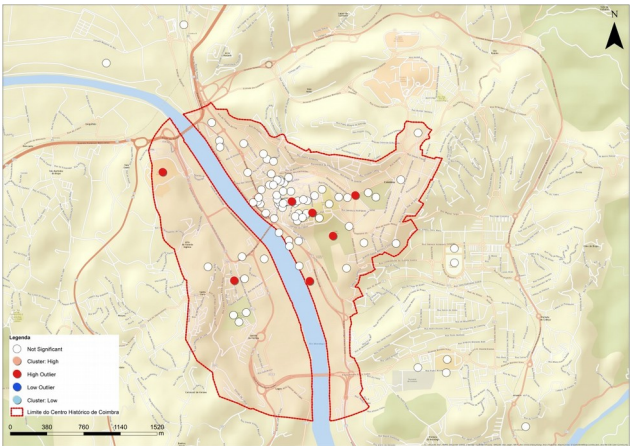


Figure 3 – Cluster and outlier analysis (*Anselin Local Moran's I*).

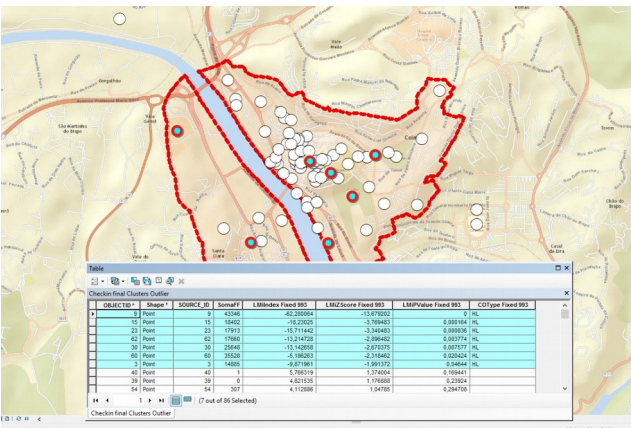


Figure 4 – Table data for *High Outliers* according to the *Cluster and outlier analysis statistic (Anselin Local Moran's I)*.

Given the check-in distribution (and the relative randomness of that distribution), one’s attention is drawn to the existence of 7 points (high outliers). How significant are they? These data correspond indeed to the lowest p-values and negative z-scores, showing that each of the objects in question could be part of a neighborhood if the values of the nearest objects were less irregular (z-scores and p-values). That not being the case, these values are statistically atypical. A text (HL) field (COType) was nevertheless created for these 7 red points (see Figure 3). This means that each of the latter corresponds to an object characterized by a large number of occurrences (check-ins) but surrounded (in what we may term a first buffer) by points with lower values, while further away there are other points with higher values. Such a distribution cannot be described as clusters, which in our opinion is why none are to be found in Figures 3 and 4.

4. NONPARAMETRIC KERNEL DENSITY ESTIMATOR (KDE) vs A COMBINATION OF OHSA DENSITY OPERATOR + IDW INTERPOLATOR: OPERATIONALIZATION AND PRESENTATION OF RESULTS

4.1. Kernel Density Estimation (KDE)

In some open-source GIS software (such as QGIS), the KDE function is – mistakenly, in our view – referred to as “heatmaps”. Without going into too much detail, as that would fall outside our present scope, we believe that its closest equivalent might be confused with *HotSpot Map*, which is different from KDE. We should point out that KDE is particularly useful in terms of reading, interpreting and analyzing thematic maps, and that it is the result of converting a vector theme to a raster theme which in turn can be changed back to vector (if deemed helpful). Anderson Medeiros (Webinar Mundogeo, 3 July 2015), sees at least two advantages in the use of Kernel maps:

- a) Whenever there is an excessive concentration of points, visual analysis can end up being hindered. Thus, for example, a given point in a given area may in fact represent more than one occurrence (as can be the case with illnesses, crime, etc.);
- b) Representation is not limited to predefined areas, as is the case with neighborhood or county polygons.

In order to build a KDE map one needs a point mesh, and for each point one needs to determine whether the distance to their location is greater or lesser than the radius of the predefined circle. One of the problems to be tackled, and for

which we offer a solution, is to determine the radius value. We need to estimate the Kernel function for each interior point, as well as the accumulated value of point concentration, with the result being converted to a raster data model. Density will tell us about the existence of clusters, but not whether they carry any statistical significance.

Another frequently asked question is to know the number of points required (after aggregation) to achieve reliable results. Data representativeness depends on the work scale, among other factors, the general idea being that “The more points you have, the more reliable your results will tend to be.” In the present case we followed the path, ... \ **Arctool-box \ Spatial Analyst tools \ Density \ Kernel density**. In the Wizard configuration (Figure 5) we entered the theme under analysis and the attribute SumFF. We changed to 10m the default value of the cell size resolution (16.6530470571816), so as to obtain a less pixelated, visually smoother output surface. In the next field we set the radius value for searching neighborhood. The final cartographic result will depend greatly on the value entered in this field. We performed a number of tests, which led to the finding that there exist major differences in the final layouts.

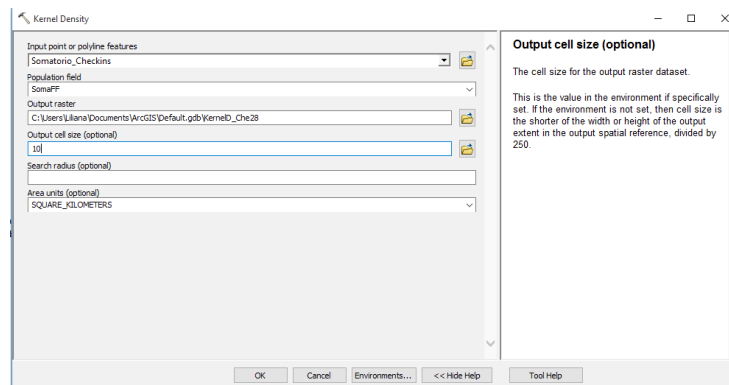


Figure 5 – Wizard Configuration in KDE.

According to the methodologies covered in the bibliographical sources used in the course of our research, it became clear that AcrMap contains a function that offers an expeditious way of solving this problem, which, as mentioned above, depends, among other things, on the scale of the work to be done. The tool in question is *Calculate Distance Band from Neighbor Count* (Figure 6), which allows us to accurately determine the radius value to be entered in this Wizard field.

The *Calculate Distance Band from Neighbor Count* function generates a *Logfile* (Figure 7), which in fact can also prove very useful for implementing the *HotSpot Analysis* tool, as it provides information with regard to the minimum, average and maximum distance required to ensure that all points have, at least, one neighbor. What this obviously means is that the maximum value thus obtained must be the one to be entered in the *Search Radius* field seen in Figure 5.

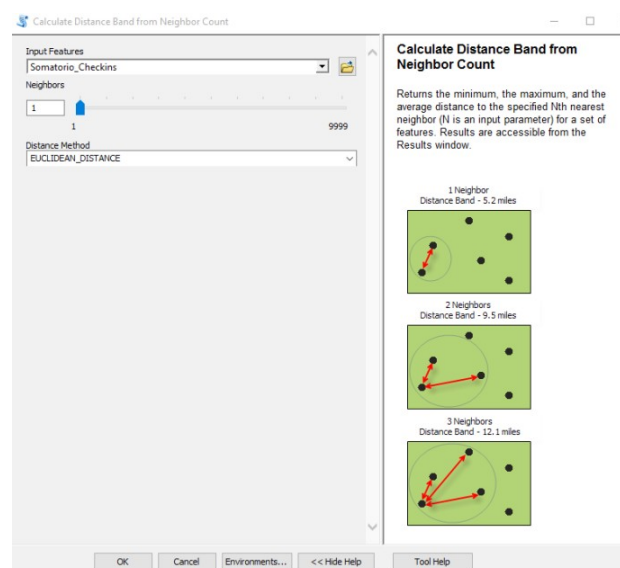


Figure 6 – Wizard of the *Calculate Distance Band from Neighbor Count* tool.

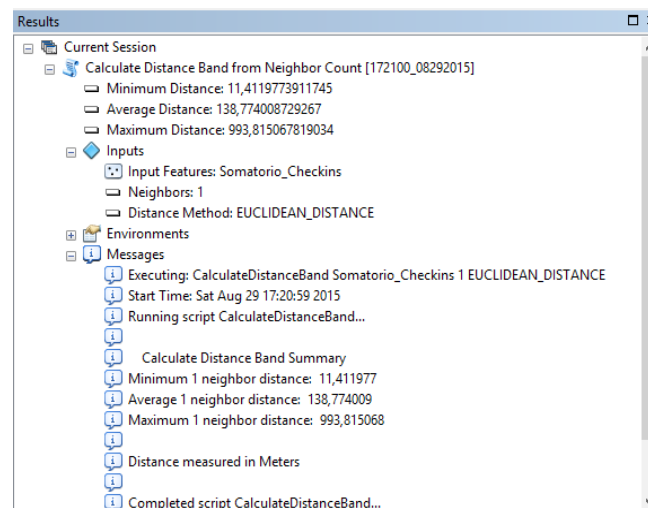


Figure 7 – Logfile generated by the *Calculate Distance Band from Neighbor Count* tool.

Once the Wizard was set up, a surface was generated, as shown in Figure 8. This is a continuous surface, where occurrences (check-ins) seem to gain relevance. As was to be expected and is suggested by Figure 8, the “Alta” area – *i.e.*, the University and the city’s most symbolic historical and religious monuments – includes the highest density of check-ins. But this type of cartography can be significantly improved when combined with basemap data appropriate to each case study, and with the inclusion of administrative limits and of labels whenever justified. Finally, we would like to note that this type of cartographic representation is not restricted to administrative purposes, but rather allows for the depiction of a number of events of a non-administrative nature but related to other kinds of phenomena instead (illnesses, crime, traffic accidents).

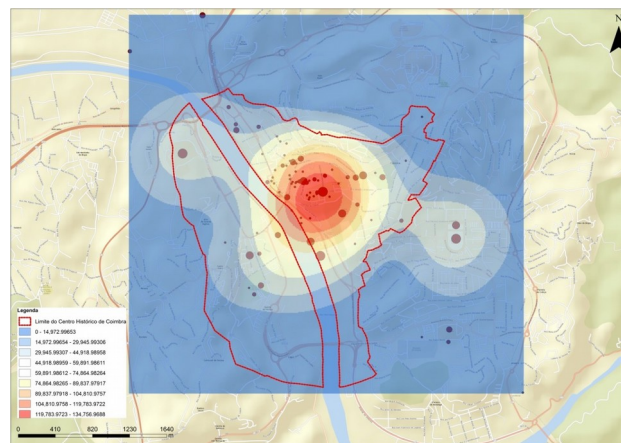


Figure 8 – Kernel Density Estimation.

4.2. Optimized Hot-Spot Analysis (OHSA) & Inverse Distance Weighting (IDW) interpolation

When using the *HotSpot Analysis* function (*HSA*) we are faced with processes leading to the aggregation of data and values (the presence of clusters) that can be either high or low, and certainly are not lacking in statistical significance (measured by confidence intervals). We have to deal with the probability that distribution is not entirely random but follows some spatial pattern instead, and we also have to consider the proximity and neighborhood factors. These concepts become especially relevant for the implementation of this analytical technique and for spatial modeling, perhaps because they contribute more and better information, at least in visual terms, than is the case with other methodologies. But in spite of making it possible to overcome the difficulties resulting from classic choropleth mapping (think of a sample with thousands of points, each representing thousands of occurrences, as is the case with calls to the Medical Emergency Number), the representation of such an amount and variety of information can prove a highly entropic exercise and thus turn into yet another obstacle instead of a decision support tool. On the other hand, it is possible to complement the *HSA*-processed vector data (in the form of points or polygons) with interpolators in order to build surfaces (in a raster data model) that can be more or less smooth depending on the value assigned to the pixel, thereby generating cartographic information for analyzing patterns capable of explaining the behavior of

such phenomena as the association we have been establishing with regard to the number of check-ins at each of Coimbra's 86 tourist POIs. One of the interpolators most frequently used as a complement of HSA is *Inverse Distance Weighting (IDW)*. There is a nuance to our algorithm – **Optimized Hot Spot Analysis (OHSA)** – that allows the less experienced user, or the one less knowledgeable in the field of geostatistics, to execute a workflow with preconfigured parameters, thereby reducing error and uncertainty. This tool is based on the **Getis-Ord G_i^*** statistical estimation procedure (where G_i^* should be pronounced as “**G Eye star**”; for more information on the procedure, read the ArcGIS Help). And as we said above, the qualifier “Optimized” is an allusion to the fact that the tool itself preconfigures some of the parameters of the data being processed. Bearing in mind that we have been working with one variable (POI) and with the spatial implications of one of its quantitative attributes (occurrences – check-ins), and given also that it is our aim to establish a correlation between those attributes and the distance factor, the OHSA tool was initially tested for the POIs according to the following path, ... \ **Arctoolbox \ Spatial Statistical tools \ Mapping clusters \ Optimized Hot Spot Analysis**.

In the Wizard configuration we placed the Checkin_Final layer and instead of inserting an attribute we just accepted the method suggested by default, *i.e.*, *Incident Data Aggregation Method (Count_Incidents_Within_Fishnet_Polygons)*. As a result, the cells thus generated indicated a POI distribution that confirms the existence of a cluster in the “Alta” and University area (Figure 9).

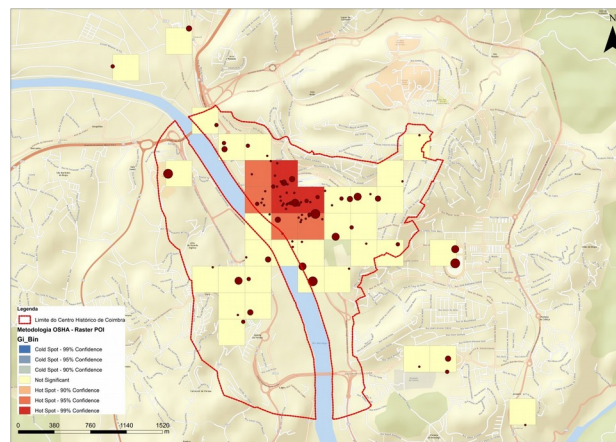


Figure 9 – Matrix representation of the “Alta” and University cluster, according to the OHSA methodology (Raster).

Since we are not dealing with a continuous surface here, the next step consisted in using a different value aggregation method – *Snap_Nearby_Incidents_to_Create_Weighted_Points* –, to permit the creation of a single weighted point from nearby incidents. We thus go back to working with vector points (Figure 10). The weight for each point is the result of the number of incidents (check-ins) for each nearby point in the defined neighborhood. We can safely confirm the existence of the cluster in the “Alta” and University area, as further attested to by the number of related check-ins. In a return to the raster data model (Figure 11), we then used the IDW interpolator for assessing continuous surfaces to look into the existence of tendencies and patterns associated with this particular distribution.

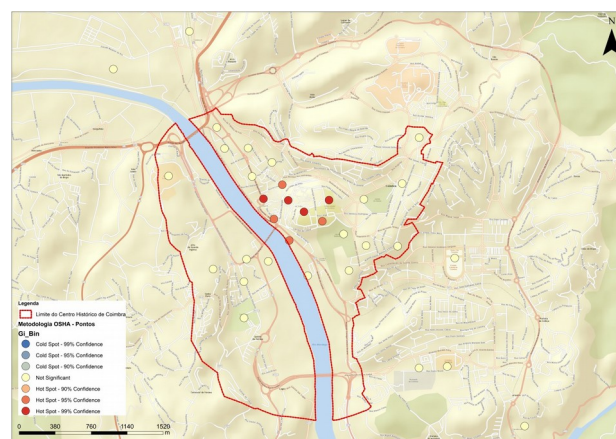


Figure 10 – Vector representation of the “Alta” and University cluster, according to the OHSA methodology (points).

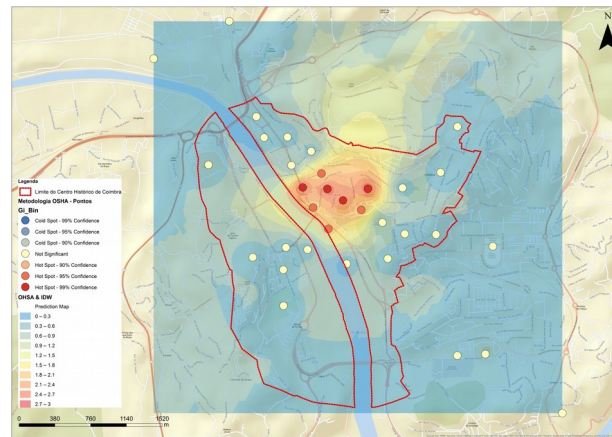


Figure 11 – OHSA & IDW surface cartogram (*Geostatistical Analyst tool*)

Let us briefly pause to reflect on a very frequent but, in our view, erroneous concept. We allude to the fact that a high number of occurrences is commonly equated with the presence of a Hot Spot. For clarification on the existence of hot and cold spots we need an analysis of the Gi-Bin field data (which in this OHSA technique is executed and displayed by default, as mentioned above), because it assigns confidence intervals to the results. We understand why the mistake is so often and hastily made, but the fact is that the relation is not always that straightforward. As we said before, to talk of a Hot Spot is to talk of **neighborhood relations**; there is no Hot Spot unless a particular feature is aggregated with other features with similar occurrences, but that **stand out from a set of neighboring features**. If all the features in a given area have high values (occurrences) and look similar, they will not stand out from the neighborhood and therefore will not constitute real Hot Spots. To give an example, in Figure 11 there are points (POI) that, once aggregated, form clusters characterized by a high level of confidence (95%) but which, after IDW interpolation, fall within areas whose level of confidence in terms of identifying the existence of a Hot Spot is not necessarily the same; in the case at hand, what we have is a value somewhere between 1.8 and 2.1, which is still far from the maximum required for being able to say with certainty that one is in the presence of Hot Spots – as they stand out from their neighborhood. *Mutatis mutandis*, the same interpretation applies to Cold Spots (absent from Figure 11).

4.3. Discussion of results

An analysis of the data and of the results obtained by implementing the modeling processes at our disposal shows the existence of optimal tourist areas in the University-“Quebra Costas” axis as well as in the city’s Green Spaces, namely in the “Parque Verde” (Green Park) area. However, in view of the check-ins (86 POI) of the two DSNs under consideration, the present comparative methodological exercise allows us to conclude that, in the case of KDE interpolation, it is possible to estimate the intensity of an event/occurrence in an area where no real occurrence has been generated. This is a probabilistic, nonparametric estimator based on the ranking of the data, its only premise being that when we assess the occurrence of events in space, their location must be considered as random. For that reason, and since the occurrences do not follow a heterogeneous distribution, it does not reflect their spatial distribution in the best manner possible, but rather indicates, without the required accuracy – hence reliability –, the areas where the occurrences show a greater concentration. In our case, it is imperative to adequately estimate the radius value for nearest neighbor search. KDE showed the existence of a cluster in the University area but did not permit us to accurately determine the value of the radius, and that means that the surface thus generated is bound to vary according to the value defined in the wizard.

The OHSA technique, on the other hand, showed with a high degree of confidence the existence of a pattern of spatial distribution very similar to KDE – in other words, the presence of a spot, an optimal tourist area in the “Alta” and University part of the city.

With regard to our present study, both techniques used (KDE and OHSA/IDW – the latter **carried out in Geospatial analyst**[Confirmar que é a designação correta]) seem to yield very similar spatial distributions, *i.e.*, it is possible to discern real Hot and Cold Spots. Given that the entire “Alta” area is a vast cluster made up of POIs that cannot be distinguished from the neighborhood, we sought other solutions, aimed at depicting and cartographically representing, if possible, a different kind of situation, in which a given POI is part of a neighborhood rather than standing apart from it. However, we should further note that the use of the same OHSA/IDW operators – whose parameters have a different configuration, and which can also be accessed **from Spatial Statistic Tools** – can produce

results and cartographic layouts that, although distinct, are still compatible with the doubts we raised earlier having to do with the fact that the POIs cannot really be distinguished from the neighborhood.

5. CONCLUSIONS

As a first conclusion we can say that, whatever the methodology used in a geospatial modeling process, one has to ponder carefully the choice of algorithms and configuration parameters. Reliability, exhaustiveness, consistency and accuracy of the input data are also key factors in ensuring the quality of the work.

Another conclusion, and also food for thought, is the possibility that the data is inadequately organized and the number of POIs under consideration (36, after aggregation) is too small to allow conclusions to be drawn with a greater degree of confidence in the techniques being used.

Finally, we would like to stress the fact that the identification of hot spots and the concepts of proximity and neighborhood effect can be invaluable in terms of applying GIS and spatial modeling based on geostatistical operators to the analysis of tourism dynamics. We are very much aware of the capabilities of these methodologies and we believe that they should be used as an analytical tool to support tourism management, because determining “where” there occur more visits and a higher concentration of occurrences (check-ins) will make it possible to develop strategies for expanding the range of tourism offer and thus allow tour operators to be “where” the tourists are. On the other hand, “where” there is no tourism demand one should try and seek the underlying reason for that fact, in order to find creative solutions for promoting tourism in those specific city locations.

REFERENCES

- ANDRIOTTI, J., 1988. Introdução à Geoestatística. Estudos Tecnológicos – **Acta Geológica Leopoldensia**, São Leopoldo, v. 11, nº27, p. 5-81.
- ANSELIN, L., GETIS, A., 1992. Spatial Statistical Analysis and Geographic Information Systems. **The Annals of Regional Science**, Vol. 26, p. 19-33, DOI: 10.1007/BF01581478.
- ANSELIN, L., 1995 . Local Indicators of Spatial Association-LISA. **Geographical Analysis**, Vol. 27, nº2, Ohio State, University Press, p. 93-115,.
- ArcGIS Help Desktop 10.3(ESRI)*
- AZEVEDO, L. 2015 . **Geolocalização, Geocolaboração e Redes Sociais. Estudo de caso aplicado ao turismo no Centro Histórico de Coimbra**. Tese de Mestrado, 194 p., Universidade de Coimbra, Coimbra.
- GARCÍA-PALOMARES, Juan Carlos, GUTIERREZ, Javier e MÍNGUEZ, Carmen. 2015. Identification of tourist hot spots based on social networks: A comparative analysis of European metropolises using photo-sharing services and GIS. **Applied Geography**, 63, p. 408-417, Elsevier, também disponível em: [https://www.eeducation.psu.edu/geog884/sites/www.eeducation.psu.edu/geog884/files/palamoresetal\(2015\).pdf](https://www.eeducation.psu.edu/geog884/sites/www.eeducation.psu.edu/geog884/files/palamoresetal(2015).pdf).
- GETIS, A., 1991. Spatial Interaction and Spatial Autocorrelation: A Cross-Product Approach. **Environment and Planning A** 23, p. 1269-77.
- GETIS, A., ORD, K., 1992. The Analysis of Spatial Association by Use of Distance Statistics. **Geographical Analysis**, 24, p. 189-206.
- ONSOY, H., BOCQUILLON, C., 2015. Interpolation de variables régionales et cartographie automatique: modèle triangulaire hydrologique (MTH)/Interpolation of regional variables and automatic drawing of isopleths, **Hydrological Sciences Bulletin**, 25/3, 297-306.
- ROCHA, F., 2015 . **Sistemas complexos, modelação e geosimulação da evolução de padrões de uso e ocupação do solo**, Dissertação de Doutoramento, 987 p. Universidade de Lisboa, Lisboa.

TOBLER, W. 1970. A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, Vol. 46, Supplement: Proceedings. International Geographical Union. Commission on Quantitative Methods, Clark University, DOI: 10.2307/143141, p. 234-240.

ZHIHUI Gu, YAN Zhang, YU, Chen e XIAOMENG Chang, 2016. Analysis of Attraction Features of Tourism Destinations in a Mega-City Based on Check-in Data Mining—A Case Study of Shenzhen, China. ISPRS, **International Journal of Geo-Information**. p. 1-13, MDPI, disponível também em: <http://www.mdpi.com/2220-9964/5/11/210>.

This work was supported by the European Regional Development Funds, through the COMPETE 2020 – Operational Programme ‘Competitiveness and Internationalization’, under Grant POCI-01-0145- FEDER-006891; and by National Funds through the Portuguese Foundation for Science and Technology (FCT) under Grant UID/GEO/04084/2013

Co-financed by:

