

G-SIVAR: A GLOBAL SPATIAL INDICATOR BASED ON VARIOGRAM

Cláudia Cristina Baptista Ramos Naizer¹ - ORCID: 0000-0001-8326-7082

David Souza Rodrigues¹ - ORCID: 0000-0002-8334-5260

Jorge Ubirajara Pedreira Junior¹ - ORCID: 0000-0002-8243-5395

Cira Souza Pitombo¹ - ORCID: 0000-0001-9864-3175

¹Escola de Engenharia de São Carlos-USP, Departamento de Engenharia de Transportes, São Carlos – São Paulo, Brasil.

E-mail: ccbnaizer@gmail.com; david.souza.rodrigues@usp.br, jorge.ubirajara@ufba.br, cirapitombo@usp.br.

Received on 05th March 2019

Accepted on 04th July 2019

Abstract:

Among the exploratory spatial data analysis tools, there are indicators of spatial association, which measure the degree of spatial dependence of analysed data and can be applied to quantitative data. Another procedure available is geostatistics, which is based on the variogram, describing quantitatively and qualitatively the spatial structure of a variable. The aim of this paper is to use the concept of the variogram to develop a global indicator of spatial association (Global Spatial Indicator Based on Variogram – G-SIVAR). The G-SIVAR indicator has a satisfactory performance for spatial association, with sensibility for anisotropy cases. Because the indicator is based on geostatistics, it is appropriate for quantitative and qualitative data. The developed indicator is derived from theoretical global variogram, providing more details of the spatial structure of the data. The G-SIVAR indicator is based on spatial dissimilarity, while traditional indexes, such as Moran's I, are based on spatial similarity.

Keywords: Spatial association, Moran, Pseudo-significance test, Semivariogram.

How to cite this article: NAIZER, C. C. B. R.; RODRIGUES, D. S.; PEDREIRA-JUNIOR, J. U.; PITOMBO, C. S. G-SIVAR: a global spatial indicator based on variogram. *Bulletin of Geodetic Sciences*. 25(4): e2019022, 2019.



This content is licensed under a Creative Commons Attribution 4.0 International License.

1. Introduction and background

Spatially correlated data exists in various areas of study, such as: epidemiology (Goovaerts, 2006; Goovaerts, 2009; Sousa et al., 2017), geology (Lee et al., 2007; Orton et al., 2016; Tamayo-Mas et al., 2016); environment (Pearce et al., 2009; Park, 2013) and urban and transport planning (Pitombo et al., 2015; Sidharthan et al., 2011; Xie and Yan, 2013; Lindner and Pitombo, 2018). Thus, current approaches emphasize the need to include spatial components in the analysis and modelling of such data as traditional non-spatial modelling presupposes data independence (Sener et al., 2010).

Using spatial models to estimate spatially dependent variables requires prior and exploratory investigation underlying the spatial dependence of these data. Among the exploratory spatial data analysis techniques, we highlight the use of indicators of spatial association. Various authors (for example, Moran, 1950; Geary, 1954; Getis and Ord, 1992; Anselin, 1995) developed indicators of spatial analysis in order to prove the existence of spatial autocorrelation and to quantify relations of spatial dependence. Most indexes refer to the similarity between values of the same variable in different locations and are applicable to quantitative variables.

The use of these indicators is well known in the literature in several areas of knowledge. Sokal and Oden (1978) point out that the commonly used method for nominal data corresponds to join counts, whereas for quantitative and ordinal data, global analyses are used through Moran I and Geary's c indices. The Moran I index is strongly affected by the difference between two neighbouring values in relation to the mean, and is more sensitive to extreme values, whereas Geary's c index is affected by the absolute difference between neighbours, composing a function based on distance (Legendre and Fortin, 1989). Comparing the techniques, Moran's I index has a more powerful hypothesis test than that presented in Geary's c index (Cliff and Ord, 1981).

Getis and Ord (1992) developed the G-statistics family, indicating their use in addition to global indicators, as a local form of analysis, identifying spatial association pockets and outliers. Anselin's (1995) Local Indicator of Spatial Association (LISA) not only identifies local spatial association, but also relates local values to the global measure of spatial association, thus evaluating the influence of outliers.

Currently, the most used approach for global analysis corresponds to the Moran index and to the LISA for local approach. The calculation of such indicators assumes that the variation of the values of the variable in space is the same for all directions (isotropy). Furthermore, these tools assume a normal distribution of data, they are not applicable to qualitative data and they are not capable of identifying and modelling the spatial structure of the data graphically.

Among spatial analysis techniques is also Geostatistics, which starts from measuring the dissimilarity of the values of a variable at different points of different locations, through a graphical tool known as variogram. While the spatial indicators, commonly known in the literature, present the result in the form of a number, geostatistics reflects the spatial autocorrelation phenomenon by calculating the experimental variogram graph, which relates semivariance and distance values between points, as well as its adjustment to a theoretical function. The technique is applicable to quantitative and qualitative variables.

Thus, this article addresses exploratory spatial data analysis. In this study, the semivariogram concepts help in the composition of the indicator of spatial association. A methodological procedure is then proposed for the composition of a global indicator of spatial association.

This paper combines the practicality of the use of indicators, which characterize the spatial association through a number, with the tool of the variogram, which allows the applicability to quantitative and qualitative data. Geostatistics is able to investigate spatial dependence with respect to directions, enabling a more complete analysis of anisotropic data. Thus, the contribution of this study is the combination of concepts in order to create a more comprehensive indicator than those currently used.

The developed indicator corresponds to the G-SIVAR, a short for Spatial Indicators based on Variographic Concepts or SIVAR.

This article presents seven sections besides this introduction. Section 2 describes basic concepts of the geostatistics tool. Section 3 indicates the methodological procedure developed for the proposal of the G-SIVAR indicator, while Section 4 describes the study region and database. Section 5 presents the results and discussions. Finally, section 6 addresses the conclusions and contributions obtained from this study.

2. The Experimental and Theoretical Semivariogram

The geostatistics analyses spatial data sets through a stochastic interpretation. Unlike Classical Statistics, observations are interpreted together to characterize the spatial pattern of the phenomenon faithfully (Wackernagel, 2003; Journel and Huijbregts, 1978).

Matheron (1971) developed the theory of regionalized variables, which are numerical functions with spatial dependence and cannot be explained by deterministic functions. Nowadays, it is possible to apply geostatistical techniques to quantitative variables (continuous and discrete) and qualitative (nominal, ordinal and binary) variables.

The semivariogram can represent the quantitative variation of the regionalized variables by calculating the semivariance according to the distance between points, measuring the dissimilarity between the data. The lower the value of the semivariance, the more similar the data between the analysed points. Therefore, if there is spatial dependence, it is natural that the value of semivariance grows with distance and stabilizes at a Sill, from which there is no more spatial dependence (Burrough, 1986). The semivariogram is given by Equation 1.

$$\gamma(h) = \frac{1}{2n} \sum_{N(h)} [Z(x+h) - Z(x)]^2 \quad (1)$$

Where: $\gamma(h)$ is the semivariogram function in h , h is the distance between variables, $Z(x)$ is the random variable Z in x and N is the number of elements accounted for.

To develop the semivariogram, the values of the angle (direction), distance between points (lag) and band and angular tolerances must be determined. The pairs of points used for the calculation include the point of origin of the vector and the points within the area delimited by the tolerances. All points are analysed for multiple distances of h , so that a graph can be made relating variance γ and distance between pairs (h), as shown in Figure 1. The same figure shows the adjustment of points to a theoretical function.

The remarkable points of the semivariogram are: Range (a), corresponding to the distance to which there is spatial correlation; Sill (C), corresponds to the value of the semivariance in range; Nugget effect (C_0), observed effect due to semivariogram discontinuities for small distances; Contribution (C_1), difference between the Sill and the Nugget effect.

The modelling of the experimental semivariograms is performed by analysing the best theoretical model that fits the experimental semivariogram points. The most usual models are: the pure nugget effect, spherical, exponential and gaussian.

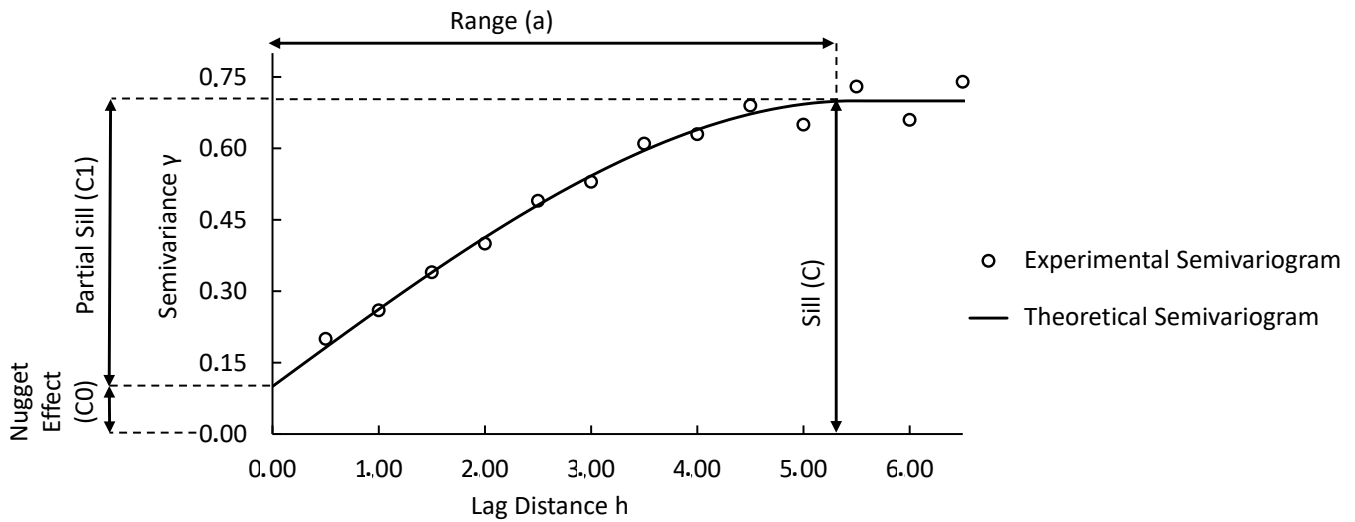


Figure 1: Theoretical semivariogram scheme. Adapted from Burrough (1986).

3. G-SIVAR: Global Spatial Indicator Based on Variogram

The G-SIVAR indicator follows the assumptions: (i) for each pre-defined distance neighbourhood, a semivariance value is obtained; (ii) this value is evaluated in order to verify if there is spatial autocorrelation in the analysed data; (iii) the indicator L_i is given as a function of the values of a given variable in the location i and the values of the neighbours i for this variable. Neighbours are considered from a certain distance and direction (Anselin, 1995).

$$L_i = f(y_i, y_{j_i}) \quad (2)$$

In which: L_i is the indicator, y_i is the variable in i and y_{j_i} are the values of the neighbours for variable i .

The developed indicator should be such that it is possible to infer the statistical significance of the spatial pattern of location “ i ”. A location “ i ” is spatial dependent if the value of the indicator is larger than a critical value (δ_i), regarding a level of significance (α_i). Thus, this requires that the probability of the indicator of being higher than the critical value is lower than the level of significance used.

$$\text{Prob}[L_i > \delta_i] \leq \alpha_i \quad (3)$$

Where: Where L_i is the indicator, δ_i is a critical value and α_i is the significance level chosen or the pseudo-significance level (for example, the result of a randomization test).

It is noteworthy that, when there is spatial dependence, the variables calculated by the indicator are not random variables. Figure 2 shows a flowchart of the proposed methodological procedure for the composition and obtaining of the global indicator G-SIVAR.

Steps 1 and 2:

First, the coordinates were standardized, so as to make it possible to compare the results between different variables. In order to determine the distance that includes the most relevant points for the spatial analysis and to automate the selection of the lag between points, we used the bandwidth concept from the Geographically Weighted Regression (GWR) technique (Hoon Cho, Lambert and Chan, 2010; Fotheringham et al., 2002).

The technique is used to model non-stationary spatial phenomena locally, adjusting a regression model for each point, using the parameters of the neighbouring points as input. Points of interest for the model are determined from the bandwidth. The selection criteria used were the minimizations through cross validation (CV) and the Akaike information criterion (AIC), and the lowest value was adopted (Fotheringham et al., 2002).

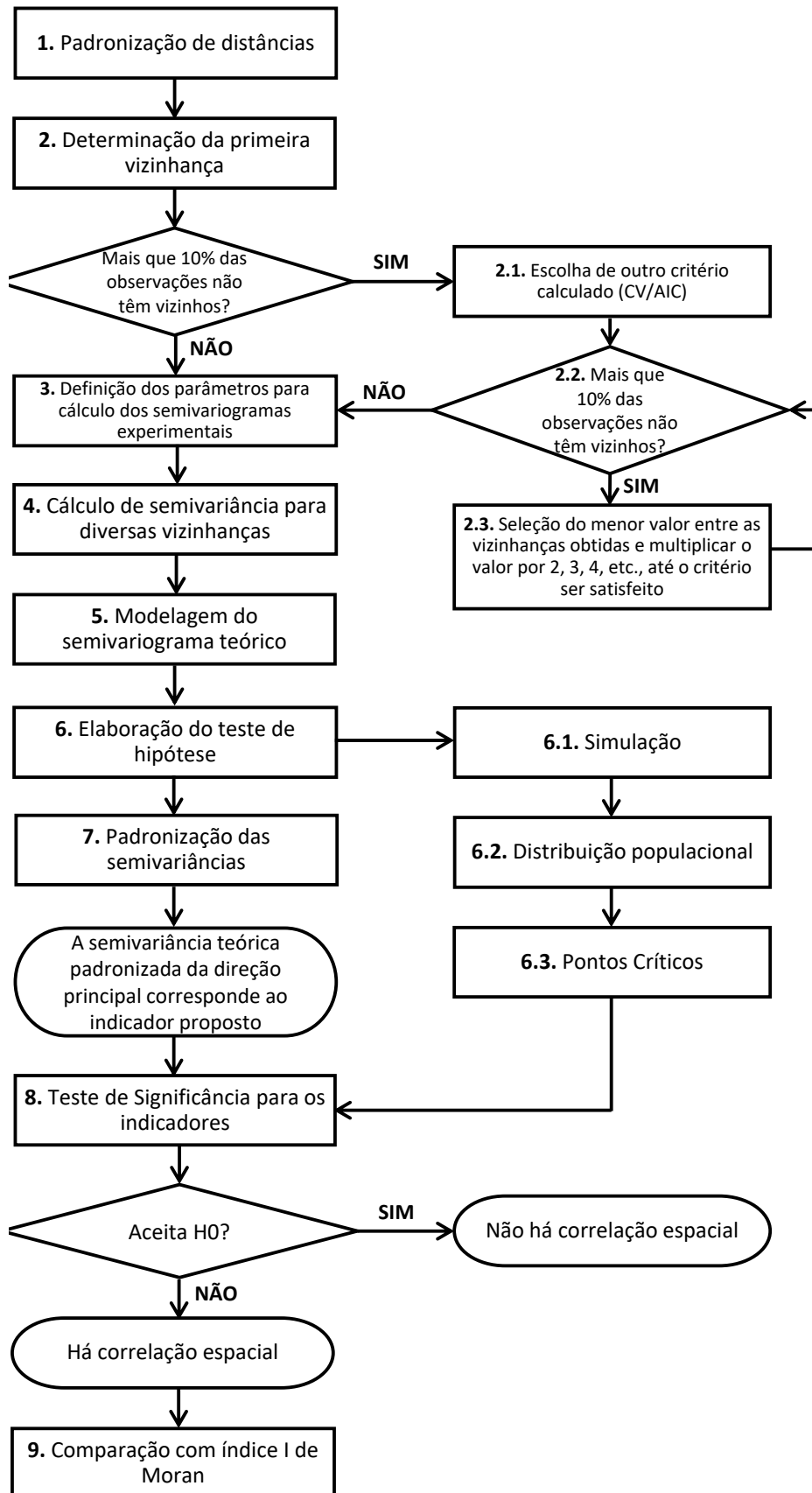


Figure 2: Flowchart of the proposed method for the Global Indicator of Spatial Association.

Steps 3, 4 and 5:

The variographic analysis begins by defining the parameters to calculate the experimental semivariograms. The lag parameters, lag tolerance, angular tolerance, band and analysis directions are defined. When the experimental semivariograms are calculated, the main and orthogonal directions are determined.

Then, the theoretical semivariogram is determined, in which a model is adjusted that considers the experimental semivariogram of the main direction and the orthogonal direction. If the spatial structure is isotropic, the same semivariogram function will be adjusted for both directions. Otherwise, there is anisotropy and the range value used in the theoretical semivariogram equation is different for the main and orthogonal directions.

Steps 6, 7 and 8:

To determine if these values are spatially correlated, a hypothesis test is developed. The hypothesis test used corresponds to a pseudo-significance test, as applied to Moran's indexes (1950). To estimate the significance of the proposed index it needs to be associated with a known probability distribution, such as the normal distribution. The basic hypotheses of the test are formulated as: (1) Null Hypothesis: the variable is spatially random; (2) Alternative Hypothesis: the variable is spatially dependent, there is a spatial pattern.

The objective of this methodological step is to generate a distribution of values of the proposed global indicator for the case of spatial randomness. This distribution is essential for determining the hypothesis test to evaluate the statistical significance of the indicator.

To do this, a database is created using the same geographical coordinates as the original, but with random data between the minimum and maximum limits of the original database. Then, omnidirectional experimental semivariograms and theoretical semivariograms for the simulated data were calculated and adjusted.

Since the generated data is random, the spatial dependence cannot be observed. When this occurs, the semivariogram tends to values corresponding to the mean semivariance of the data, forming a horizontal line near this value. In this situation, the graph is interpreted as the pure nugget effect.

After 100 repetitions, the normality of the distribution of theoretical semivariance is checked by the Kolmogorov-Smirnov test. If the distribution is similar to normal, the corresponding critical values can be found for the desired level of significance.

The G-SIVAR indicator corresponds to the theoretical semivariance of the standardized main direction and this is shown in Equations (4,5,6,7). As we can see, it is a set of possible equations representing the different possible adjustments: Exponential, Spherical, Gaussian and Pure Nugget Effect. The standardization should be performed in such a way that the maximum possible value for the semivariogram level is equivalent to the typical random data level, i.e. the mean value of the semivariance obtained for the simulated data (spatially random data) for the hypothesis test.

Exponential	$G_SIVAR(h) = \frac{C_0 + C \left[1 - \exp\left(-\frac{h}{a}\right) \right]}{C_0 + C} \quad (4)$
Spherical	$G_SIVAR(h) = \begin{cases} \frac{C_0 + \frac{C}{2} \left[3\left(\frac{h}{a}\right) - \left(\frac{h}{a}\right)^3 \right]}{C_0 + C} & \text{para } h < a \\ 1 & \text{para } h \geq a \end{cases} \quad (5)$
Gaussian	$G_SIVAR(h) = \frac{C_0 + C \left[1 - \exp\left(-\left(\frac{h}{a}\right)^2\right) \right]}{C_0 + C} \quad (6)$
Nugget Effect	$G_SIVAR(h) = 1 \quad (7)$

Where: $G_SIVAR(h)$ is the Global Spatial Indicator Based on Variogram to the neighbourhood h , C_0 is the Nugget Effect of the Theoretical Variogram, C is the Contribution of the Theoretical Variogram, h is the Distance and a is the Range of the Theoretical Variogram.

Having the values of the G-SIVAR indicator and the critical points, assuming that the distribution of semivariance of the random data is similar to the normal one, a test of significance is performed, evaluating whether or not to accept the hypothesis of spatial randomness for the values obtained for the indicator. The methodological steps for G-SIVAR calculation are available in a R - code for any interested user (<https://jupedreirajr.github.io/g-sivar/>).

4. Data

The database used corresponds to the average number of car trips per household for a cut from the City Center of São Paulo - SP, Brazil, from the Mobility Survey 2012 (Metrô, 2012). The data were aggregated into squares of 500m x 500m, represented by their centroids.

The city of São Paulo is the capital of the state of São Paulo, located in the southeastern region of Brazil, Latin America. For the year 2017, the estimated population corresponds to 12,106,920 inhabitants, being the municipality with the largest population of the country (IBGE, 2017). The region of interest is shown in Figure 3.

The database has 905 observations. The variable average number of car trips ranges from 0 to 10, with an average of 2.285 and a standard deviation of 1.906. Figure 4 shows the distribution of the study variable in the central region of São Paulo.



Figure 3: Location of the region of interest in the city of São Paulo, Brazil.

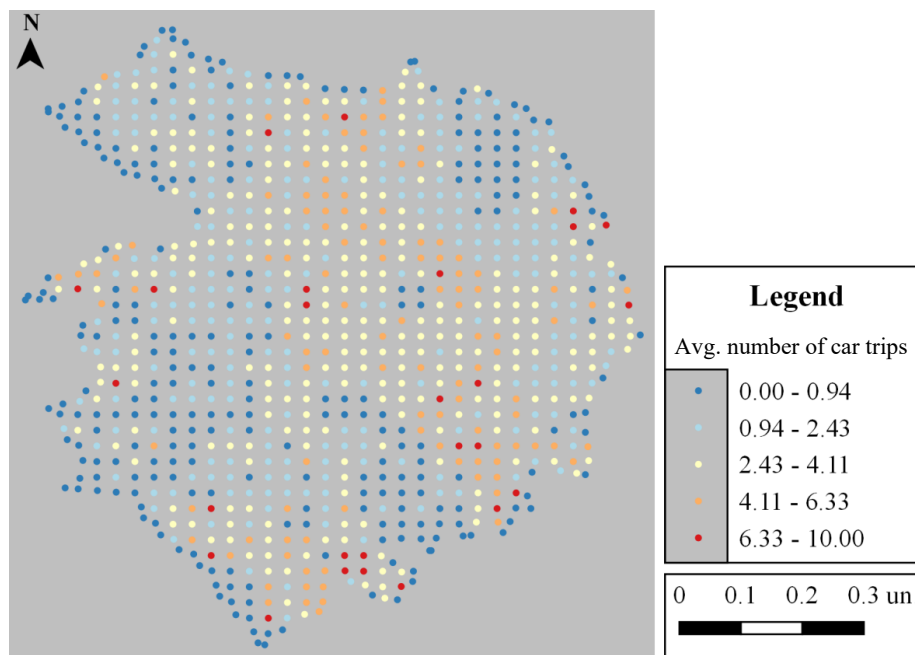


Figure 4: Average number of car trips per square of 500 x 500 m represented in the form of centroids.

5. Results and discussion

The lag corresponding to the bandwidth of 0.044, determined by minimizing the Akaike information criterion, was used.

The variographic analysis of the variable average number of car trips was performed in the 0°, 30°, 60°, 90°, 120° and 150° directions. The parameters were set as follows: lag of 0.044, lag tolerance of 0.022, angular tolerance of 22°, band tolerance of 0.022. It can be observed that the main direction corresponds to 120° and the orthogonal direction to 30°.

The theoretical semivariogram was adjusted through an Exponential function, the nugget effect (C_0) corresponds to 2.000 and the contribution (C_1) to 1.800. Anisotropy was observed in the data set. The variograms are shown in Figure 5 and the experimental and theoretical semivariance values for the main and orthogonal directions of the first and most relevant neighbourhoods are presented in Table 1.

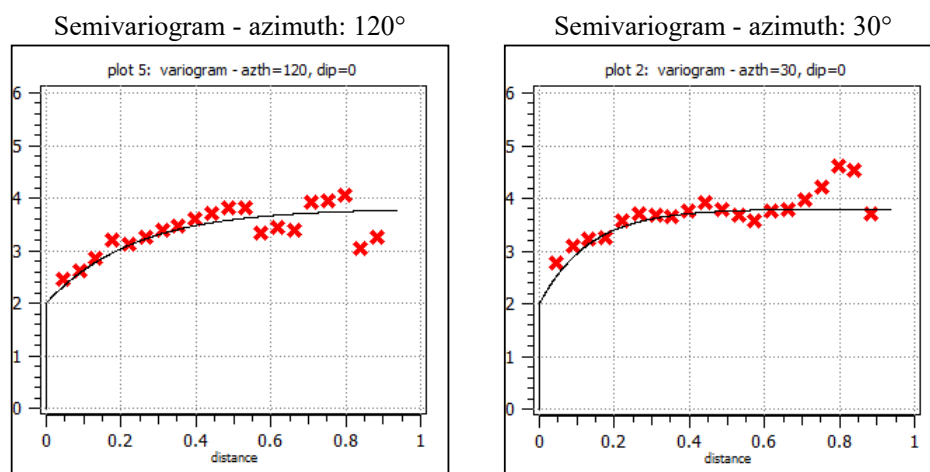


Figure 5: Experimental and theoretical semivariograms for main and orthogonal direction.

Table 1: Summary of experimental and theoretical semivariance values for main and orthogonal directions.

Viz.	Dist. (h)	Semivariance (γ)			
		Experimental		Theoretical	
		Main direction (120°)	Orthogonal Direction (30°)	Main Direction (120°)	Orthogonal Direction (30°)
1	0.044	2.462	2.771	2.357	2.188
2	0.088	2.608	3.082	2.643	2.357
3	0.133	2.860	3.236	2.873	2.508
4	0.177	3.198	3.266	3.057	2.643
5	0.221	3.128	3.574	3.204	2.764
6	0.265	3.258	3.695	3.322	2.873
7	0.310	3.378	3.688	3.417	2.970

As previously described, a database was generated with the same coordinates as the original database. The values adopted for the variable average number of car trips correspond to random numbers between zero and ten, the minimum and maximum values of the original database. The semivariograms developed are omnidirectional, considering spatial tendencies that the data can present. The random data presented no spatial structure, producing semivariograms with a constant trend represented by the pure nugget effect model. This procedure was repeated 100 times, generating new random data within each repetition.

The distribution of theoretical semivariance has an average of 8.292 and a standard deviation of 0.262 and is similar to normal according to the Kolmogorov-Smirnov test. Thus, the values corresponding to the critical points can be obtained by comparing the analysed distribution with the standard normal distribution.

For a 95% confidence level, the limit obtained through a unilateral analysis of the standard normal distribution corresponds to the semivariance of 7.861. If the value of the obtained indicator is above the established limits, the null hypothesis must be retained with a given level of significance, that is, the data have no spatial association. Otherwise, if H_0 is rejected, there is spatial correlation.

Considering that the mean value of the semivariance of the simulated data corresponds to 8.292, the data can be standardised. Thus, the values presented in Table 2 were obtained, and the values of theoretical standardized semivariance for the main direction correspond to the proposed indicator.

Table 2: Comparison of indicator values, z and null hypothesis status (for 95% significance) for Moran I Index and G-SIVAR.

Neigh,	Dist,	Moran				G-SIVAR			
		Index	Z	p-value	H0	Index	Z	p-value	H0
1	0.044	0.347	20.273	0.000	Rejects	0.620	-12.017	0.000	Rejects
2	0.088	0.234	26.101	0.000	Rejects	0.696	-9.633	0.000	Rejects
3	0.133	0.171	28.780	0.000	Rejects	0.756	-7.722	0.000	Rejects
4	0.177	0.148	28.217	0.000	Rejects	0.804	-6.190	0.000	Rejects
5	0.221	0.123	27.130	0.000	Rejects	0.843	-4.962	0.000	Rejects
6	0.265	0.106	26.589	0.000	Rejects	0.874	-3.978	0.000	Rejects
7	0.310	0.087	24.564	0.000	Rejects	0.899	-3.189	0.001	Rejects

Equation 8 shows the calibrated G-SIVAR for the study region and the main direction.

$$G_SIVAR(h) = 0,526 + 0,474 \left[1 - \exp\left(-\frac{h}{0,200}\right) \right] \quad (8)$$

Where: h is the Distance (neighbourhood)

It can be observed that G-SIVAR is a measure of dissimilarity and its values range from zero to one. The closer to one, the greater the dissimilarity between the pairs of points, considering a direction and tolerances, in addition to a distance, determined by the neighbourhood. This measure is associated with a hypothesis test that measures how significantly the indicator represents spatial association.

The last step in the methodology, illustrated in Figure 2, is to compare the G-SIVAR to an indicator commonly used in the literature with the purpose of validating the procedure described in this article. In this paper, the comparison between G-SIVAR and the Moran index was performed for the case of global spatial association. Table 2, above, shows the relationship between G-SIVAR and the Moran index in order to validate the results found.

The Moran Index I value for the variable average number of car trips corresponds to approximately 0.347 for the first neighbourhood, reflecting a moderate spatial correlation between the data. The index values decrease as the distance considered increases, all of them are rejected for 95% confidence in the Moran hypothesis test, that is, there is a spatial correlation between the data for the first seven neighbourhoods.

In this example, it is clear that the indices are complementary as Moran's index measures similarities while G-SIVAR measures dissimilarities. Values that correspond to the high spatial correlation are close to one for Moran and zero for G-SIVAR. The values obtained for the proposed and Moran index almost complement each other in terms of totalling the unit value, as shown in Figure 6.

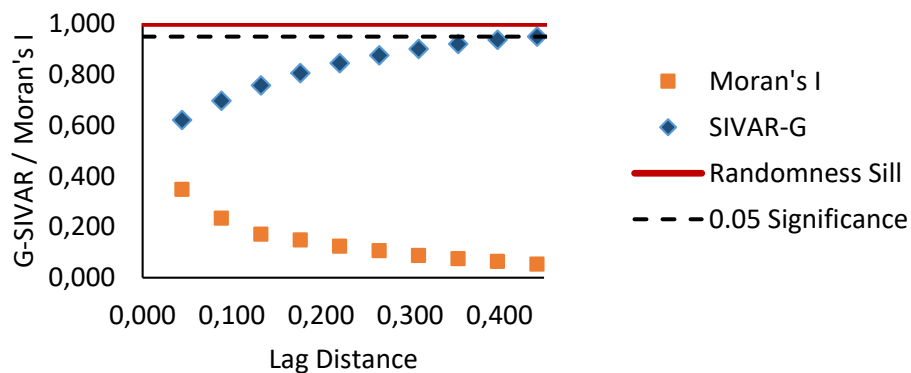


Figure 6: Comparison between the Moran indices and the proposed indicator and exposure of the critical limit to the 95% confidence level.

6. Conclusions

The G-SIVAR indicator can identify and quantify the spatial association of spatially distributed data adequately and it is consistent with the Moran index, showing the advantage of being applicable to quantitative and qualitative data, as well as providing a more detailed analysis of the structure of the variable.

In addition, the G-SIVAR indicator was sensitive to the chosen direction, and was able to identify anisotropy of the data and return different values of spatial association depending on the analysed direction. In the example presented, this was one practical benefit of using G_SIVAR instead of Global Moran. The proposed global indicator is more suitable when dealing with spatial dependence associated to a direction.

It is worth remembering that using the bandwidth of the GWR is a way of calculating the ideal distance of

analysis of the set of neighbours, but it is not a rule. As in the case of the Moran, the proposed indicator can be calculated for any distance that the user wants; the smoothing parameter is only a theoretical suggestion of distance.

ACKNOWLEDGEMENT

This research was carried out with the support from the National Council of Scientific and Technological Development (CNPq427525/2018-7; 303645/2015-6) and São Paulo Research Foundation (FAPESP 2018/08755-4 AND FAPESP 2016/07563-9).

AUTHOR'S CONTRIBUTION

Cláudia Cristina Baptista Ramos Naizer: method proposition and text writing

David Souza Rodrigues: code development

Jorge Ubirajara Pedreira Junior: code development and text revision

Cira Souza Pitombo: method proposition and text revision.

References

- Anselin, L. and Griffith, D. A. 1988. Do spatial effects really matter in regression analysis?. *Papers in Regional Science*, 65, p.11-34.
- Anselin, L. 1995. Local indicators of spatial association—LISA. *Geographical analysis*, 27, p.93-115.
- Burrough, P. A. 1986. *Principles of geographical information systems for land resources assessment*. New York: Oxford Press, p.155-65.
- Cliff, A. and Ord, K. 1972. Testing for spatial autocorrelation among regression residuals. *Geographical analysis*, 4, p.267-84.
- Elzhov, T. V. Mullen, K. M. and Bolker, B. 2010. R interface to the Levenberg-Marquardt nonlinear least-squares algorithm found in MINPACK. *Plus Support for Bounds*. R package version 1.2-1. <https://CRAN.R-project.org/package=minpack.lm>
- Fotheringham, A. S. Brunsdon, C. and Charlton, M. 2002. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships/cA*. Stewart Fotherington, Chris Brunsdon, and Martin Charlton. Wiley, p.284.
- Geary, R. C. 1954. The contiguity ratio and statistical mapping. *The incorporated statistician*, 5(3), p.115-46.
- Getis, A. and Ord, J. K. 1992. The analysis of spatial association by use of distance statistics. *Geographical analysis*, 24, p.189-206
- Goovaerts, P. 2006. Geostatistical analysis of disease data: accounting for spatial support and population density in the isopleth mapping of cancer mortality risk using area-to-point Poisson kriging. *International Journal of Health Geographics*, 5(1), p.52.
- Goovaerts, P. 2009. Medical geography: a promising field of application for geostatistics. *Mathematical Geosciences*, 41(3), p.243-64.

- Hoon Cho, S., Lambert, D. and Chen, Z. 2010 Geographically weighted regression bandwidth selection and spatial autocorrelation: An empirical example using Chinese agriculture data. *Applied Economics Letters* 17(8), p.767-772.
- IBGE, Instituto Brasileiro de Geografia e Estatística, 2017. *Brasil em Síntese*, Panorama Cidade de São Paulo. [online] Available at: < <https://cidades.ibge.gov.br/brasil/sp/sao-paulo/panorama> > [Accessed 3 March 2018].
- Journel, A. G. and Huijbregts, C. J. 1978. *Mining geostatistics*. Academic press.
- Lee, S. Y. Carle, S. F. and Fogg, G. E. 2007. Geologic heterogeneity and a comparison of two geostatistical models: Sequential Gaussian and transition probability-based geostatistical simulation. *Advances in water resources*, 30(9), p.1914-32.
- Lindner, A. and Pitombo, C. S. 2018. A Conjoint Approach of Spatial Statistics and a Traditional Method for Travel Mode Choice Issues. *Journal of Geovisualization and Spatial Analysis*, 2, pp. 1-13.
- Mathereon, G. 1970. La Théorie des Variables Régionalisées et ses Applications, *Les Cahiers de Morphologie Mathématique de Fontainebleau*. Paris: École Nationale Supérieure des Mines de Paris, n°5 pp. 211.
- Metrô, Companhia do Metropolitano de São Paulo, 2012. *Pesquisa de Mobilidade 2012*. [online] Available at: < <http://www.metro.sp.gov.br/> > [Accessed 10 August 2016].
- Moran, P. A. 1950. Notes on continuous stochastic phenomena. *Biometrika*, 37, pp. 17-23.
- Orton, T. G. Pringle, M. J. and Bishop, T. F. A. 2016. A one-step approach for modelling and mapping soil properties based on profile data sampled over varying depth intervals. *Geoderma*, 262, pp. 174-86.
- Park, N. W. 2013. Spatial downscaling of TRMM precipitation using geostatistics and fine scale environmental variables. *Advances in Meteorology*, 2013.
- Pearce, J. L. Rathbun, S. L. Aguilar-Villalobos, M. and Naeher, L. P. 2009. Characterizing the spatiotemporal variability of PM_{2.5} in Cusco, Peru using kriging with external drift. *Atmospheric Environment*, 43(12), pp. 2060–69.
- Pitombo, C. S. Salgueiro, A. R. da Costa, A. S. G. and Isler, C. A. 2015. A two-step method for mode choice estimation with socioeconomic and spatial information. *Spatial Statistics*, 11, pp. 45-64.
- Sener, I. N. Pendyala, R. M. and Bhat, C. R. 2011. Accommodating spatial correlation across choice alternatives in discrete choice models: an application to modeling residential location choice behavior. *Journal of Transport Geography*, 19, pp. 294-303.
- Sidharthan, R. Bhat, C. Pendyala, R. and Goulias, K. 2011. Model for children's school travel mode choice: accounting for effects of spatial and social interaction. *Transportation Research Record: Journal of the Transportation Research Board*, 2213, pp. 78-86.
- Sousa, D. C. B. D. Pitombo, C. S. Rocha, S. S. Salgueiro, A. R. and Delgado, J. P. M. 2017. Violence in public transportation: an approach based on spatial analysis. *Revista de saude publica*, 51, pp. 127.
- Tamayo-Mas, E. Mustapha, H. and Dimitrakopoulos, R. 2016. Testing geological heterogeneity representations for enhanced oil recovery techniques. *Journal of Petroleum Science and Engineering*, 146, pp. 222-240.
- Wackernagel H. 2003 *Multivariate Geostatistics*, 3rd edn. Springer-Verlag, Heidelberg, pp. 388
- Xie, Z. and Yan, J. 2013. Detecting traffic accident clusters with network kernel density estimation and local spatial statistics: an integrated approach. *Journal of Transport Geography*, 31, pp. 64-71.