BCG - Boletim de Ciências Geodésicas - On-Line version, ISSN 1982-2170
http://dx.doi.org/10.1590/S1982-21702017000200019

Article

DETECTION OF INCONSISTENCIES IN GEOSPATIAL DATA WITH GEOSTATISTICS

Detecção de inconsistências em dados geoespaciais através da geoestatística

Adriana Maria Rocha Trancoso Santos[1]
Gerson Rodrigues dos Santos[1]
Paulo César Emiliano[1]
Nilcilene das Graças Medeiros[1]
Amy L. Kaleita[2]
Lígia de Oliveira Serrano Pruski[1]

[1]Universidade Federal de Viçosa – UFV – Av. Peter Henry Rolfs, s/n - Campus Universitário – Viçosa – MG – Brazil – CEP: 36570-900. Emails: adrianatrancoso1@gmail.com; gerson.santos@ufv.br; paulo.emiliano@ufv.br; nilcilene.medeiros@ufv.br; ligia.oserrano@gmail.com.

[2]Iowa State University – ISU – 3352 Elings Hall – 605 Bissell Road – Ames – IA – USA – ZIP: 50010. Email: kaleita@iastate.edu.

Abstract:

Almost every researcher has come through observations that "drift" from the rest of the sample, suggesting some inconsistency. The aim of this paper is to propose a new inconsistent data detection method for continuous geospatial data based in Geostatistics, independently from the generative cause (measuring and execution errors and inherent variability data). The choice of Geostatistics is based in its ideal characteristics, as avoiding systematic errors, for example. The importance of a new inconsistent detection method proposal is in the fact that some existing methods used in geospatial data consider theoretical assumptions hardly attended. Equally, the choice of the data set is related to the importance of the LiDAR technology (Light Detection and Ranging) in the production of Digital Elevation Models (DEM). Thus, with the new methodology it was possible to detect and map discrepant data. Comparing it to a much utilized detections method, *BoxPlot*, the importance and functionality of the new method was verified, since the *BoxPlot* did not detect any data classified as discrepant. The proposed method pointed that, in average, 1,2% of the data of possible regionalized inferior outliers and, in average, 1,4% of possible regionalized superior outliers, in relation to the set of data used in the study.

Keywords: Outliers, geoprocessing, LiDAR technology.

Resumo:

Provavelmente todo pesquisador já se deparou com observações que se "afastam" das demais, sugerindo a existência de inconsistências. O objetivo desse trabalho é propor um novo método de detecção de dados inconsistentes para dados geoespaciais contínuos baseando-se na Geoestatística, independentemente da causa geradora (erros de medição, execução e variabilidade inerente aos dados). A escolha pela Geoestatística está baseada em suas características ideais, como evitar erros

sistemáticos, por exemplo. A importância da proposta de um método de detecção de dados inconsistentes está no fato de que alguns métodos existentes utilizados em dados geoespaciais consideram pressuposições teóricas dificilmente atendidas. De igual forma, a escolha do conjunto de dados está relacionada com a importância da tecnologia LiDAR (Light Detection and Ranging) na produção de Modelos Digitais de Elevação (MDE). Assim, com a nova metodologia foi possível detectar e mapear dados discrepantes. Comparando-a com um método muito utilizado de detecção, *BoxPlot*, verificou-se a importância e funcionalidade do novo método, já que o *BoxPlot* não detectou nenhum dado como discrepante. O método proposto apontou, em média, 1,2 % dos dados de possíveis *outliers* inferiores regionalizados e, em média, 1,4 % de possíveis *outliers* superiores regionalizados em relação ao conjunto de dados estudados.

Palavras-chaves: Outliers; geoprocessamento; tecnologia LiDAR.

1. Introduction

It is very likely that most researchers have encountered data in which some observations are very different from the rest, suggesting any number of issues: that the data are naturally or legitimately erratic, or that the data generating mechanism is not the same, or that the unusual data indeed belong to another population. These observations are considered to be inconsistent, commonly called outliers, or discrepant data.

Many authors have contributed to this subject, such as (see, e.g., Anscombe, 1960; Grubbs, 1969; Beckman and Cook,1983, Rousseuw and Zomeren, 1990; Muñoz-Garcia et al., 1990; Barnett and Lewis,1994) among other pioneers. Some of these authors assert that the concern about disparate data is as old as the first attempts at analysis of a set of data, as in the case of comments of Bernoulli in 1777 about the existence of such data.

Recently, new methods for handling outliers have been developed to meet the demands of various areas of scientific knowledge, as in the case of (Hongxing, et al., 2001) for spatial data distributed in irregular grids, (Barua and Alhajj, 2007) for processing images, (Qiao et al., 2013) for data from satellites and (Appice et al., 2014) for geophysical data stream.

Studying the detection of outliers is important because the first step in data analysis consists in the evaluation of data quality. Although it is important to evaluate each observation in-depth, to discuss the impact of each in the analysis, and to consider the inclusion or exclusion of a given observation in the analysis during the outliers' detection phase, in some situations all posterior action can be rejected because of the decision taken in the beginning of the data analysis (Muñoz-Garcia et al., 1990).

The importance of proposing an inconsistent data detection method lies on the fact that a good portion of the methods adopted in the treatment of geospatial data consider theoretical assumptions rarely met and/or verified, ie. according to (Mood et al., 1974), the variables must be independent and identically distributed for a classical statistical treatment. Most variables from planned sampling are demonstrably dependent on space (not random) and difficult to prove as distribution, such as geospatial data. Thus, a new methodology that takes into account planned sampling (in the form of regular or irregular grids) and the spatial dependence structure between observed values, considering this structure throughout the analytical data process, will be of great scientific value (Yamamoto and Landim, 2013).

The choice for Geostatistics as the support methodology is based on its ideal characteristics, as from its historical inception in 1951. The mining engineer Daniel Gerhardus Krige and the statistician Herbert Simon Sichel, studying data of gold concentration found the existence of

outliers, leading them to adopt a similar procedure to the moving averages in order to avoid systematic errors (Silva et al., 2008).

According to (Yamamoto and Landim, 2013) and (Santos et al., 2011), to solve this problem it was necessary to create a method that consider all the information available of a particular area, so the estimation of variance is as little as possible. This method guarantees the minimum variance and was developed in 1971 and was given the name kriging by Georges Matheron in tribute to Daniel Krige.

Geostatistics was chosen in the outliers' analysis once, according to (Vieira, 2000), it was created to estimate dependent variables with accuracy (without trend), with minimum variance, taking into account spatial dependence structure of the samples in modeling and predicting phase.

The aim of this paper is to create a new detection method to identify and manage inconsistent data in continuous geospatial data based on Geostatistics, regardless of the cause the inconsistencies (measurement errors, runtime errors and variability inherent in data).

2.Materials and Methods

2.1 Data description

The study area is located in the state of Iowa, USA, in the County of Pottawattamie, comprising a portion of 34.32 ha in the city of Treynor. The study region is delimited by latitudes 41°10'23"N and 41°09'53"N, and longitudes 95°38'24"W and 95°38'47"W, as shown in Figure 1.

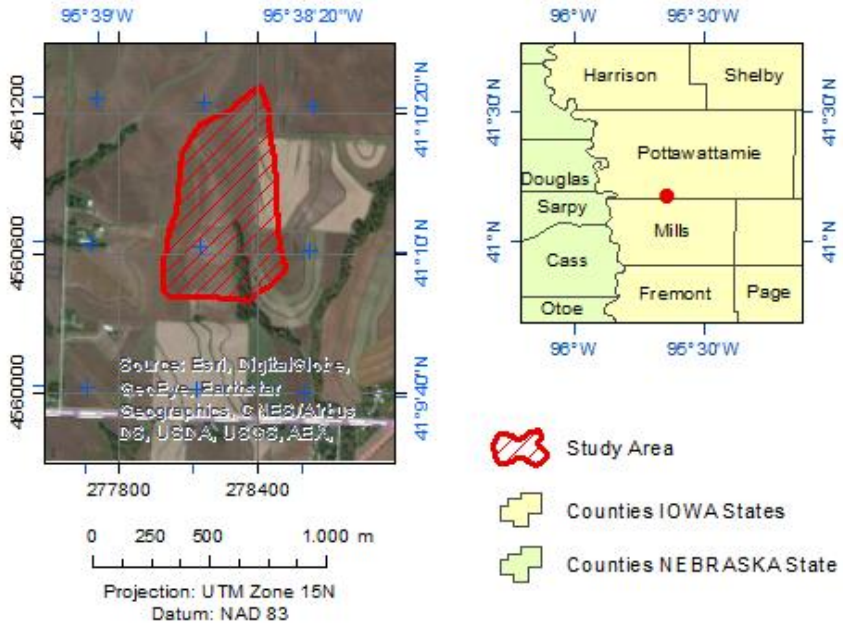

Figure 1: Overview of the study area, located in the state of Iowa, USA, in the county of Pottawattamie, comprising a portion of 34.3209 hectare in the city of Treynor.

As discussed in (Höhle and Höhle, 2009), for mapping areas, Digital Elevation Models (DEM) have been produced using mainly LiDAR technology (Light Detection and Ranging). With this technology, the high density of planialtimetric points provided is shown to be accurate and efficient.

The altimetry data used in this study, collected with LiDAR mapping, are referenced to geodetic system NAD 83 (North American Datum of 1983) and represented in projection UTM (Universal Transverse Mercator) zone 15N. Amounting 192,079 thousand points of known elevations within the study area, with a density of 0.55 points/m² and a spacing of approximately 1.7 and 1.2 m in the X and Y directions. Elevation values range from a minimum of 340.8 m to a maximum of 385.5 m.

2.2 Method proposition

The proposition of creating a new method for inconsistency detection for geospatial data was based on theoretical assumptions of residual from statistical modeling, according to (Rencher and Schaalje, 2008). Such residuals are characterized as white noise, that is, in its standardized form, it follows a Gaussian probability's distribution with zero mean and unit variance. In other words, standard normal distribution, $\mathcal{E}_p$"~Z(0;1), in which $\mathcal{E}_p$" are the standardized residuals, according to Vieira (2000).

To meet the theoretical assumptions of residual, we adopted the geostatistical analysis for the geospatial data, following the recommendations of (Vieira, 2000; Santos et al., 2011; Yamamoto and Landim, 2013), meaning that, a method that models with no trend and with minimum variance, taking into account the spatial dependence structure of the samples.

In order to obtain the residuals, as directed by (Druck et al., 2004), the regionalized variable chosen in this study was additively decomposed, into three components: a structural component associated with a constant average value or a constant trend; a random component spatially correlated; and a residual component, also called white noise, random noise or random walk. Therefore, considering the vector of spatial location x, where the variable is in a position (x = x), two (**x** = [x, y]) or more dimensions (**x** = [x, y, z, ...]), the regionalized variables Y, x, also called random function may be denoted as

$$Y(\mathrm{x}) = \mu(\mathrm{x}) + \varepsilon'(\mathrm{x}) + \varepsilon'' \qquad 1$$

in which: μ(x) is a deterministic function that describes the structural component Y in x; $\mathcal{E}$'(x) is a correlated stochastic term locally; and $\mathcal{E}$" is a white noise uncorrelated with normal distribution with zero mean and variance σ^2.

Thus, adopting the understanding of the decomposition of Equation 1, the geostatistics methodology was used to analyze the geospatial data with spatial dependence detected and characterized in order to get the residuals from this modeling. This methodology consists of classical exploratory analysis, spatial exploratory analysis, assumptions testing, variogram analysis, variogram modeling, cross-validation and kriging (Ferreira et al., 2013).

Adopting this geostatistical approach to geospatial data corresponds to the consideration each georeferenced sample point as a random variable, and generation of a random function, or commonly named, stochastic process (Cressie, 1993; Santos et al., 2011).

According to (Vieira, 2000; Yamamoto and Landim, 2013), the stationarity assumption of the variogram is assumed, ie. the variogram exists and is stationary for the variable in the study area in order to guarantee statistical validity in this methodology.

Adopting the theoretical variogram, according to Vieira (2000), $2\gamma(h)=E\{[Y(x) - Y(x+h)]^2\}$ and the estimator of (Kamimura et al., 2013), given by the equation

$$\hat{\gamma}(\mathrm{h}) = \frac{1}{2N(\mathrm{h})} \sum_{i=1}^{N(\mathrm{h})} \{[\mathrm{Y}(\mathrm{x}_i) - \mathrm{Y}(\mathrm{x}_i + \mathrm{h})]^2\} \quad (2)$$

in which: $N(\mathrm{h})$ is the number of pairs of measured values in (x_i) and (x_i+h) ; $Y(x_i)$ and $Y(x_i+h)$ represent all random variables separated by a vector **h** which generates the samples and, therefore, the variogram, the main spatial dependence detection mechanism of the geostatistical methodology, i.e., a graph of $\hat{\gamma}(\mathrm{h})$ in function of distance-vector **h**.

After the geostatistical analysis, residuals are consequently obtained. Therefore, the white noise characteristics were tested, namely, independence, normal distribution with zero mean and constant variance. With satisfactory results, the next step was to interval estimation of the residual. As the estimators are considered random variables, its estimates are usually distinct from the value of the parameter, i.e., commonly commits an estimation error. For this reason, it became necessary to construct confidence intervals with probability (1-α) (Ferreira, 2009).

For the interval estimation (CI - confidence interval) the standard normal distribution and α significance level of 1% are adopted. The level α is also called the level of the probability of committing a type I error, it means the probability of reject a null true hypothesis (which in this case will $H_0 : \varepsilon''_{p_i} = 0$) (Mood et al., 1974; Vieira, 2000; Casella and Berger, 2010).

Thus, all values that do not belong to the created CI, without bias, and with minimum variance, taking into account the spatial dependence structure, are possible inconsistent values.

In an innovative way, using the georeferencing data resources, we intend to point, to quantify, and to determine the location of the residuals with high probability of being classified as outliers.

For comparison and/or validation of the new method, comparisons of the new method were performed with one of the most robust, statistically, and current detection methods used, the boxplot (Hoaglin et al., 1983).

All the innovative part of the methodology was executed using the free software R (R Development Core Team, 2014), in which the geostatistical analysis was performed using the geoR package, developed by (Ribeiro Junior and Diggle, 2001). However, for the geoprocessing analysis, we used the software Arcgis 10.2 (ESRI, 2011). After georeferenciation and analysis of the data through the *Geostatistical Analyst* tool, the sample size was reduced, i.e., sub-sampled the data using irregular sampling grids generating three others samples with 75 %, 50 % and 25 % of the original data. These procedures were performed through the *Toolbox Random Selection.*

3 Results and Discussion

3.1 Data characterization

The original sample had 192,079 points. By reducing the sampling size to about 75 %, 50 % and 25 % of the original size, we counted with three extra sets of data with 144,059; 96,040 and 48,020 points, respectively, apart from the original. Figure 2 presents the three-dimensional representations of the four data sets of the study area.

With the four images in Figure 2 it is possible to observe a distinct change in the topography of the region and, at first, it is not possibly to point the actual caused the change, i.e., if this change exists in the study area or it is a problem, generating probably outliers.

3.2 Exploratory data analysis

According to (Yamamoto and Landim, 2013; Ferreira et al., 2013), it is important to check the spatial behavior of the data, as well as exploratory analysis.

With this respect, in Figure 3 the graphics of quartiles (graphics using the first quartile, median and third quartile as color divider) of the four data sets are represented.

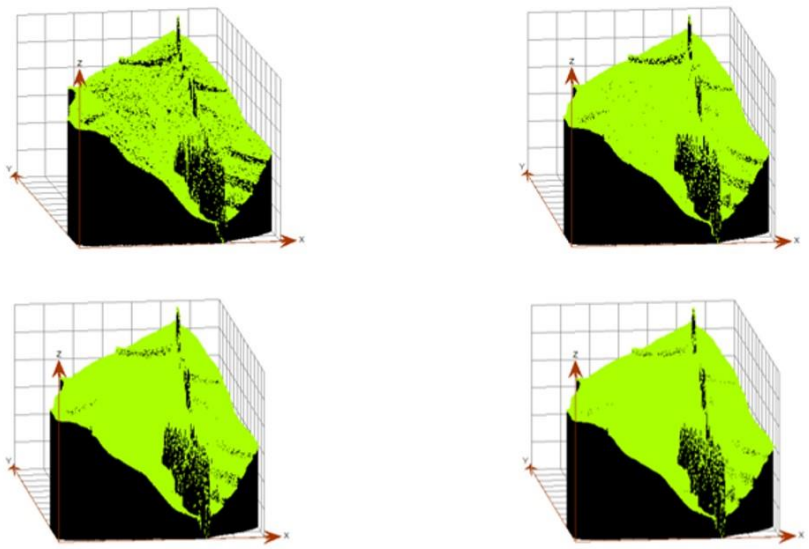

Figure 2: Overview of altimetry data by LiDAR Cloud of a small hydrographic basin of the region of the town of Treynor - Iowa – USA with 25% of the data (upper left), 50% (upper right), 75% (lower left and 100% (bottom right).

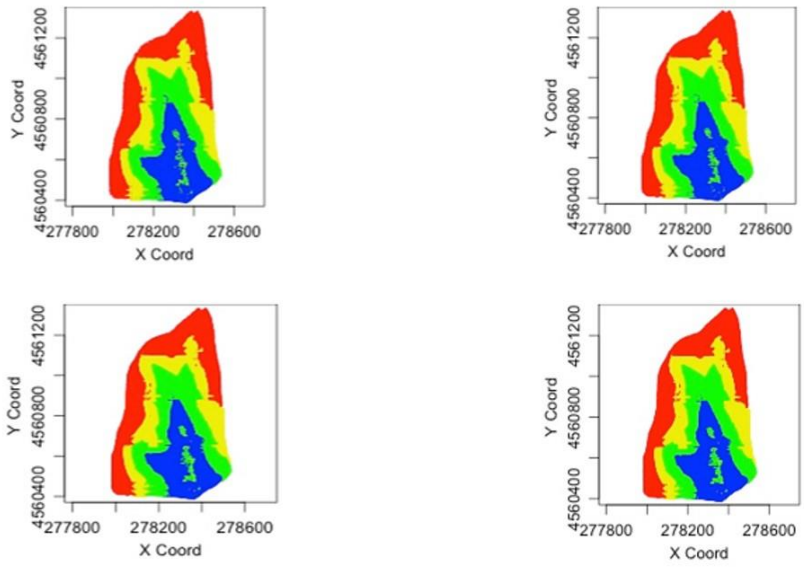

Figure 3: Data behavior presentation about the "intensity" of altimetry obtained by LiDAR Cloud of a small hydrographic basin of the region of the town of Treynor - Iowa - USA, using the graphic of quartiles. 25% of the data (upper left), 50% (upper right), 75% (lower left) and 100% (bottom right).

As shown in Figure 3, for the 4 sampling sizes, the color red refers to the highest quartile value of the data present at the ends of the images. Then one realizes that the immediately lower values, represented by the color yellow, are the values that arise between the median and the third quartile. After these are the values in the second quartile, represented by the color green, and finally, the values of the first quartile, or lower altitude values represented by the color blue and the central positions of images.

It is important mentioning that, once again, between the altimetry's lower values, possible discrepant values were also perceptible in Figure 3.

In order to establish the existence of discrepant data throughout the study area, mainly, in the most central part of the images, it is used the *BoxPlot* resource, as shown in Figure 4.

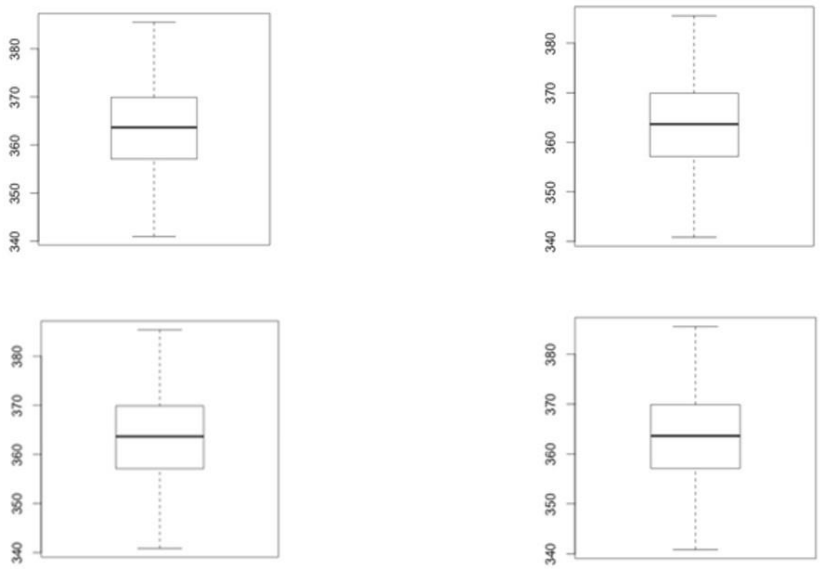

Figure 4: Presentation of the BoxPlot of the altimetry data.

Through Figure 4 it can be seen that this important and widely used outlier detection methodology was not possible to detect the perceived inconsistent data through Figures 2 and 3. Despite the strength of this methodology, according to (Tukey, 1977; Benjamini, 1988) apparently the data set has not discrepancies.

3.3 Geostatistical data analysis

As described in the paper methodology, in order to obtain the residuals with properties that meet the theoretical assumptions of Statistics, the data were analyzed by Geostatistics, as shown in Table 1.

On Table 1, we can see that the main analysis features have been preserved in the 3 sets data from the full sampling, namely, mean, variance, standard deviation, and the isotropic variogram model. Also notable is the small range of estimates of the variogram parameters.

Based on variogram analysis, Figure 5, it was possible to represent the behavior of the altimetry in the entire study area through interpolation via simple kriging, as recommended by (Santos et al., 2011).

As the kriging interpolates taking into account the spatial dependence structure characterized by the neighborhood points, it is possible to see that the Figure 5 shows the same perceived

characteristic of possible outliers. Thus, the kriging map represents the set of data, but cannot represent what actually happens in the study area, if it evidences the presence of *outliers.*

Table 1: Main informations about the geostatistical data analysis of altimetry.

Measure/Feature	**Estimates**			
Sampling Size (%)	25	50	75	100
Average (m)	363.27	363.67	363.26	363.26
Variance (m^2)	62.73	62.25	62.25	62.32
Standard Deviation (m)	7.92	7.89	7.89	7.89
Anisotropy	No	No	No	No
Variogram Model	Gaussian	Gaussian	Gaussian	Gaussian
Nugget Efect (m^2)	2.78	2.19	3.02	3.01
Sill (m^2)	67.37	62.19	69.57	67.04
Range (m)	336.75	310.80	350.87	339.39

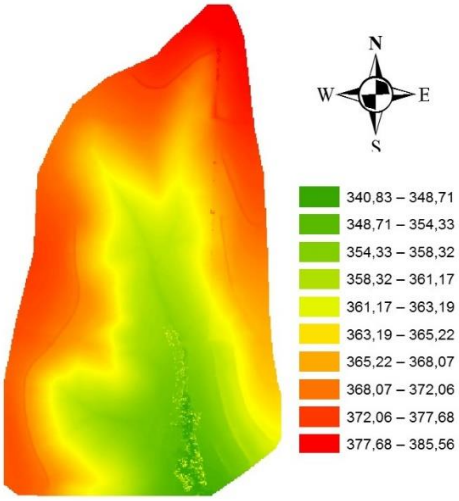

Figure 5: Simple Kriging of altimetry data, obtained by LiDAR Cloud, of a small hydrographic basin of the region of the town of Treynor - Iowa - USA.

3.4 Outliers detection using residual analysis

Residuals from adequate modeling have important features, among them, there is the spatial distribution without agglomerations, which is commonly called spatial homogeneity, as shown in Figure 6. Other features, besides this, of equal importance, are: independence, normality, zero mean and unit variance, if the residual is standard, as was the case in this study, the residual must display a standard normal distribution, i.e., normal with a zero mean and unit variance. All these features were found by statistical tests.

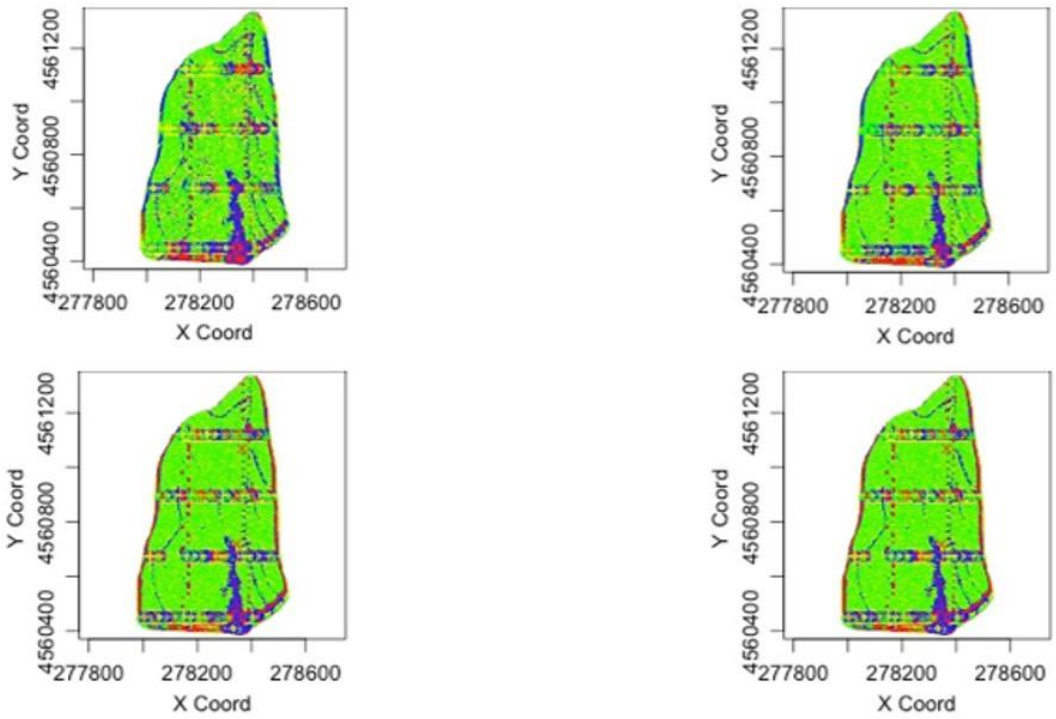

Figure 6: Presentation of the spatial homogeneity behavior of residuals from the cross-validation of a geostatistical analysis using the graphic of quartiles.

The behavior of Figure 6 differs from Figure 3 because of the spatial uniformity characteristics of the residuals, however, Figure 6 shows a further agglomeration (blue color in the south-central part of the region) which demonstrates the inconsistency problem.

Thus, as recommended by (Ferreira, 2009), once the residuals have all requirements of the statistical assumptions, it passes to the interval estimation of probability of 99%, as results shown in Table 2.

Table 2: Interval estimation with 99% of probability for the residuals from the cross-validation of a geostatistical analysis of altimetry data

Sampling Size (%)	$CI_{(99\%)}$ [Min ; Max] (m)
25	[-14.21; 14.22]
50	[-12.44; 12.44]
75	[-19.45; 19.48]
100	[-21.33; 21.37]

As the intervals estimated were bilateral, two types of outliers were detected, superior and inferior outliers. It is shown in Figure 7 the data with a high probability of being inferior outliers, as well as the location thereof.

It is also presented the data with a high probability of being upper outliers in Figure 8. Briefly, shows the percentage of data that were considered outliers in Table 3.

The results in Table 3 show that, even with the significant reduction in the data set size, the percentage of outliers detected statistically vary from 1.03 % to 1.72 %. Also in this sense,

assessing the figures of this section, it may be noted that since the beginning of geostatistical analysis it was possible to detect regions where data were disparate in relation to others, which proves the importance of methodological adoption of spatial statistical data in this nature.

On such importance, (Vieira, 2000; Ferreira et al., 2013; Yamamoto and Landim, 2013) state and/or show the importance of using Spatial Statistics for geospatial data, besides not ignore the classical statistics.

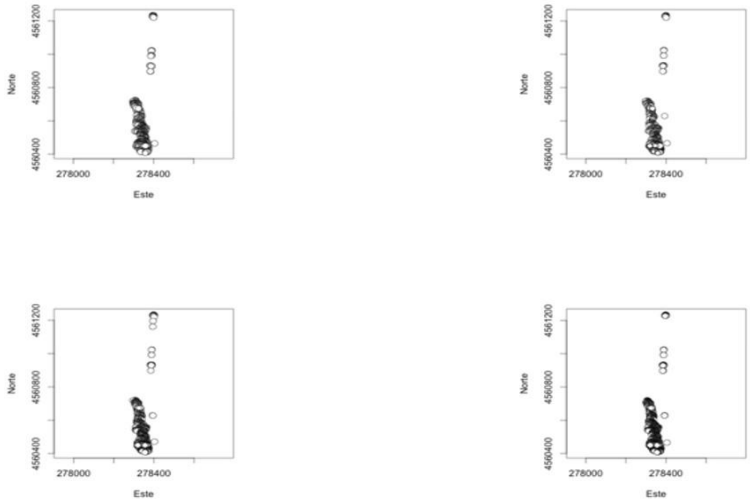

Figure 7: Presentation of the lower values with a high probability of being *outliers*, for 25% of the data (upper left), 50% (upper right), 75% (lower left) and 100% (bottom right).

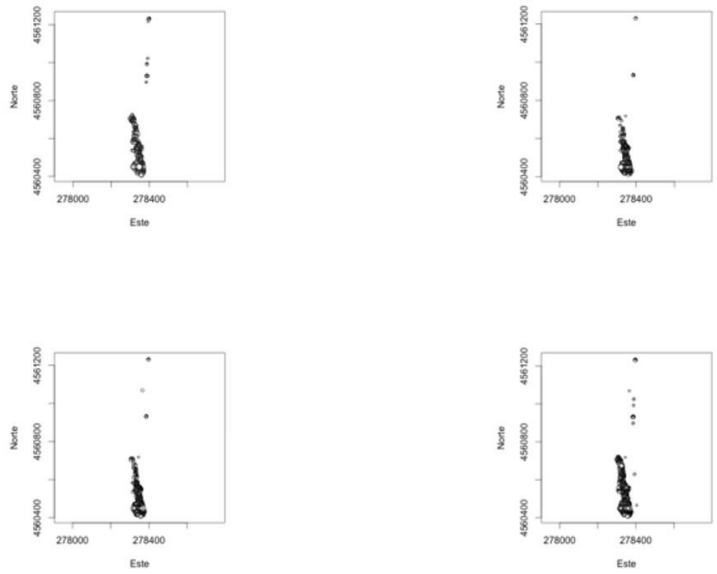

Figure 8: Presentation of upper values with a high probability of being outliers, for 25% of the data (upper left), 50% (upper right), 75% (lower left) and 100% (bottom right).

Table 3: Data summary percentage with high probability of being considered outliers for altimetry for 4 sets of work data.

Sampling Size (%)	Lower (%)	Upper (%)
25	1.67	1.72
50	1.03	1.22
75	1.07	1.23
100	1.12	1.33

4. Final considerations

Data sets in which some values differ from the rest can generate wrong conclusions about the sampled data and the population. These values are usually called outliers and all kinds of studies are subject to the occurrence these.

As observed in the paper and discussed by other authors, regardless of the cause generating these inconsistencies (measurement errors, runtime errors, inherent variability of the data, etc.) and the type of variables in study (georeferenced or not, univariate or multivariate), it is necessary to adopt correct methods of analysis, and not adopt general methods because the number of theoretical assumptions may differ strongly.

A new inconsistencies detection method for geospatial continuous data was obtained through the use of a geostatistical methodology. For example, utilizing the optimal characteristics of the geostatistical modelling for this kind of data was possible to obtain residuals (from the cross-validation) that presented all the theoretical assumptions required for it, and therefore, to detect the observed data that do not follow the regional and probabilistic behavior of the neighborhood. Furthermore, due to the georeferencing of the data, it was possible to identify the position of the inconsistencies in the studied region.

Comparing this new outlier detection technique with a widely used method of detection, Boxplot, it was found the importance and functionality of the new method, since the Boxplot did not detect any data as discrepant (even if it had detected, the Boxplot method does not consider the geospatial location).

As a recommendation for future work, it is suggested to further this method of detection creating solutions for the variables in that the residual confidence interval cannot admit the bilateralism of it, because you could lose the practical application of the method.

References

Anscombe, F.J. "Rejection of outliers". *Technometrics* 20 (1960): 123-147. doi:10.1080/00401706.1960.10489888.

Appice, A., Guccione, P., Malerba, D., & Ciampi, A. "Dealing. with temporal and spatial correlations to classify outliers in geophysical data streams". *Information Science* 285 (2014): 162-80. doi:10.1016/j.ins.2013.12.009.

Barnett, V., & Lewis, T. "Outliers in statistical data." *Biometrical Journal* 379 (1994): 256. doi:10.1002/bimj.4710370219.

Barua, S., & Alhajj, R. "High performance computing for spatial outliers detection using parallel wavelet transform." *Intelligent Data Analysis* 11 (2007): 707-730.

Beckman, R.J., & Cook, R.D. (1983). "Outliers". *Technometrics* 25 (1983): 119-149. doi:10.1080/00401706.1983.10487840.

Benjamini, Y., & Addison, W. "Opening the Box of a Boxplot." *Journal of the American Statistical Association* 42 (1988): 257-262. doi:10.2307/2685133

Casella, G., Berger, R.L. *Inferência estatística.* Cengage Learning (2010).

Cressie, N. *Statistics for spatial data.* Wiley-Interscience (1993).

Druck, S., Carvalho, M.S., Câmara, G., & Monteiro, A.V.M., (Ed.) *Análise Espacial de Dados Geográficos.* Brasília, Embrapa (2004).

Environmental Systems Research Institute (2011) – Esri. ArcGIS Desktop: Release 10. Redlands, CA: 2011.

Ferreira, D.F. *Estatística básica.* Lavras, Editora UFLA (2009)

Ferreira, I.O., Santos, G.R., & Rodrigues, D.D. (2013). "Estudo sobre a utilização adequada da krigagem na representação computacional de superfícies batimétricas." *Revista Brasileira de Cartografia* 65 (2013): 831-842.

Grubbs, F.E. "Procedures for detecting outlying observations in samples." *Technometrics* 11 (1969): 1-21. doi:10.1080/00401706.1969.10490657.

Hoaglin, D.C, Mosteller, F., & Tukey, J.W. *Understanding robust and exploratory data analysis* New York, J. Wiley (1983)

Höhle, J., & Höhle, M. "Accuracy assessment of digital elevation models by means of robust statistical methods." *ISPRS Journal of Photogrammetry and Remote Sensing* 64 (2009): 398-406. doi:10.1016/j.isprsjprs.2009.02.003.

Hongxing, L., Kennetch, C.J., & Morton, E.O. "Detecting outliers in irregularly distributed spatial data sets by locally adaptive and robust statistical analysis and GIS." *International Journal Geographical Information Science* 15 (2001): 721-741. doi:10.1080/13658810110060442.

Kamimura, K.M., Santos, G.R., Oliveira, M.S., Dias, Jr., M.S., & Guimarães, P.T.G. "Variabilidade espacial de atributos físicos de um Latossolo Vermelho-Amarelo sob lavoura cafeeira." *Revista Brasileira de Ciência do Solo* 37 (2013): 877-888. doi:10.1590/S0100-06832013000400006.

Mood, A.M., Graybill, F.A., & Boes, D.C. *Introduction to the theory of statistics.* Kogakusha, McGraw-Hill, (1974)

Muñoz-Garcia, J., Moreno-Rebollo, J.L., & Pascual-Acosta, A. "Outliers: a formal approach." *International Statistical Review* 58 (1990): 215-226. doi:10.2307/1403805.

Qiao, C., Haibo, H., & Hong, M. Spatial outlier detection based on iterative self-organizing learning model. *Neurocomputing* 117 (2013): 161-172. doi:10.1016/j.neucom.2013.02.007.

R Core Team. R: a language and environment for statistical computing. *R Foundation for Statistical Computing*, (2014) Vienna, W. Recuperado de http://www.Rproject.org/.

Rencher, A.C., & Schaalje, G.B. *Linear Models in Statistics*. New Jersey, John Wiley & Sons, (2008)

Ribeiro, J.P.J, & Diggle, P.J. GeoR: a package for geostatistical analysis. *R-News*. 1: 15-18.

Rousseuw, P.J., & Zomeren, B.C. "Unmasking multivariate outliers and leverage points." *Journal of the American Statistical Association* 85 (1990): 633-639. doi:10.1080/01621459.1990.10474920.

Santos, G.R., Oliveira, M.S, & Santos, A.M.R.T. Krigagem simples versus krigagem universal: qual o preditor mais preciso? *Revista Energia na Agricultura* 26 (2011): 49-55.

Silva, S.A., Lima, J.S.S., Souza, G.S., & Oliveira, R.B. "Avaliação de interpoladores estatísticos e determinísticos na estimativa de atributos do solo em agricultura de precisão." *Idesia* 26 (2008): 75-81. doi:10.4067/S0718-34292008000200010.

Tukey, J.W. *Exploratory Data Analysis* Princeton, Ed. Pearson (1977)

Yamamoto, J., & Landim, P. *Geoestatística: Conceitos e Aplicações* São Paulo, Oficina de Textos, (2013)

Vieira, S.R. "Geoestatística em estudos de variabilidade espacial do solo." *Tópicos em Ciências do Solo* 1(2000): 1-54.

Recebido em 15 de junho de 2016.

Aceito em 12 de setembro de 2016.