MULTIDISCIPLINARY REVIEWS **ISSN 2595-3982**
https://doi.org/10.29327/multi.2018007

REVIEW

Georreferenced statistics: approach and practical application in Brazilian animal science

Mariah Tenório de Carvalho Souza[a], Enmelly Rayane Azevedo da Rocha[b]

[a]Universidade Federal de Alagoas (UFAL/Campus Arapiraca) – Pesquisadora PDCR.
[b]Universidade Federal de Alagoas (UFAL/Campus Arapiraca) – Graduanda em Zootecnia.

ABSTRACT The objective of this review was to improve knowledge about data geoprocessing as well as an application of georeferenced statistics for use of animal science. The advancement of knowledge in the various areas that make up the agricultural sciences has evidenced the limitations of traditional methods of statistics, in the practical application of spatial variability and in the analysis of variables of the soil-plant-water - atmosphere system in large areas territories. The national and international literature reports several methods of forage mass determination, however, the greatest difficulty is found in accurate methods for determining the availability of forage in critical or drought periods in Caatinga areas of Brazilian semiarid regions. The methodology proposed by geostatistics attempts to extract, even with randomness of the data collected, the structural and probabilistic characteristics of the regionalized phenomenon. The interpolation method is called kriging and is based on the sample data the regionalized variable and the structural properties of the semivariogram obtained from these data. Isoline maps have level curves representing a set of regionalized points and may represent an isovalue. The studies with the litter of dead material on the soil (burlap), cacti and cycads are already represented in areas of Caatinga, however, greater importance should be given to the practical application of georeferenced statistics and the evaluation of spatial distribution applied to animal science for a better understanding of environmental characterization.

KEYWORDS: *Caatinga; georeferencing; isolines; kriging; zootechny*

Received June 13, 2018 *Accepted* June 18, 2018 *Published online* June 23, 2018

Cite this article:
Souza MTC, Rocha ERA (2018) Georreferenced statistics: approach and practical application in Brazilian animal science. *Multidisciplinary Reviews* 1: e2018007. DOI: 10.29327/multi.2018007

Estatística georreferenciada: abordagem e aplicação prática na ciência animal brasileira

RESUMO O objetivo desta revisão foi melhorar o conhecimento sobre geoprocessamento de dados, bem como a aplicação de estatísticas georreferenciadas para o uso da ciência animal. O avanço do conhecimento nas diversas áreas que compõem as ciências agrárias tem evidenciado as limitações dos métodos tradicionais de estatística, na aplicação prática da variabilidade espacial e na análise de variáveis do sistema solo-planta-água-atmosfera em territórios de grandes áreas. A literatura nacional e internacional relata diversos métodos de determinação da massa de forragem, entretanto, a maior dificuldade é encontrada em métodos acurados para determinar a disponibilidade de forragem em períodos críticos ou de seca nas áreas de Caatinga das regiões semiáridas brasileiras. A metodologia proposta pela geostatística tenta extrair, mesmo com a aleatoriedade dos dados coletados, as características estruturais e probabilísticas do fenômeno regionalizado. O método de interpolação é chamado de krigagem e baseia-se na variável regionalizada amostrada e nas propriedades estruturais do semivariograma obtido a partir desses dados. Os mapas de isolinhas têm curvas de nível representando um conjunto de pontos regionalizados e podem representar um isovalor. Os estudos de material morto no solo (estopa), cactos e cicadáceas já estão representados em áreas da Caatinga, no entanto, maior importância deve ser dada à aplicação prática de estatísticas georreferenciadas e à avaliação da distribuição espacial aplicada à zootecnia, além de uma melhor compreensão da caracterização ambiental.

PALAVRAS-CHAVE: Caatinga; georreferenciamento; isolinhas; krigagem; zootecnia

CONTATO Mariah Tenório de Carvalho Souza – mariah_tenorio@hotmail.com

MALQUE Publishing

Introduction

It is a fact that for decades many researchers and scientists have been misinterpreting environmental issues in the Caatinga areas of the semi-arid region of the Brazilian Northeast, often having an impact on an "adverse climate" and on "poverty in environmental resources" in these areas.

It is known that the climate in the semiarid region has as an important characteristic the irregularity of the rainfall regime, with two defined seasons: the rainy season, which corresponds to periods with rainfall, irregularly distributed, and the dry season, corresponding to critical periods of dry season with low voluminous offer for animals raised on pasture. But this does not prevent animal production from being successful in semi-arid environments with areas of native Caatinga.

It is important to note that the predominant vegetation in these areas are resilient and opportunistic; deciduous plants with a high regrowth capacity. The production of the burlap from this vegetation can contribute to the soil nutrient cycle and serves as a forage support for small ruminants in transitional periods (dry water) and dry season.

However, few studies are conducted in Caatinga areas to assess the spatial variability and behavior of native vegetation between the dry and rainy seasons, which can lead to the aggravation of desertification of these areas, occupied by small ruminants in extractivist models of creation.

Therefore, studies that aim at the evaluation of spatial distribution are justified by the possibility of obtaining new information to optimize the production in extractive systems of these areas. Thus, the aim of this review is to improve the understanding on geoprocessing of data in animal science and to address the application of georeferenced statistics for use in animal science.

Literature review

This review was elaborated with data from researches in data bank of scientific journals (Scielo, Elsevier) and books. The advance of knowledge in the different areas that make up the agrarian sciences has evidenced the limitations of the traditional methods of statistics, in the treatment of the spatial variability of variables of the soil-plant-water-atmosphere system.

Classical statistics techniques assume the principles of chance, repetition and local control, where all samples are random and independent of a simple probability distribution, without spatial continuity. Therefore, its application does not imply any knowledge of the current position of the samples or the relationship between the samples. Spatial statistics is the quantitative study of phenomena positioned in space, and assumes that values are associated with location in space, with continuous distribution of values and admits estimation processes for unsampled values.

In geostatistics, it is assumed that the distribution of the differences of variables between two sampled points is the same for the whole area and that this depends only on the distance between them and on the orientation of the points (Clark 1979). Thus, it is possible to admit that multivariate geostatistics is the application of mathematical and statistical methods to problems of Earth Sciences, with the main objective of simultaneously estimating a set of spatially correlated variables (corregionalized variables).

In this way, the estimation of the dependence between neighboring samples in space can be made through autocorrelation, which is very useful when sampling in one direction. When sampling involves two directions (x, y), the most indicated instrument in estimating the dependence between the samples is the semivariogram (Silva 1988).

The set of variables Z (x) measured in the whole area S can be considered a random function Z (x), since, according to Isaaks and Srivastava (1989), they are random, regionalized variables and it is assumed that the dependence between them specifies by some probabilistic mechanism (Figure 1).

The next, the semivariogram analyzes the degree of spatial dependence between samples within an experimental field, in addition to defining necessary parameters for the estimation of values for unsampled premises, through the technique of "kriging" (Salviano 1996). This being the case, the variogram is the basic tool, which allows quantitatively describing the variation in space of a regionalized phenomenon (Huijbregts 1975).

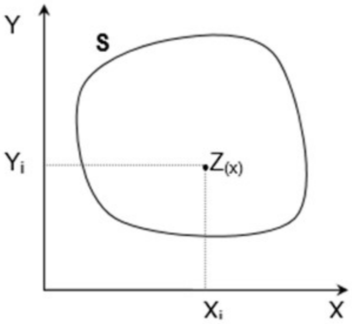

Figure 1. Regionalized random variable (Z) in an area (S). Source: Isaaks e Srivastava (1989)

In general, geostatistics calculates estimates within a context governed by a natural phenomenon with distribution in space and, thus, assumes that the values of the variables, considered as regionalized, are spatially correlated. Due to this characteristic, its application has been large mainly to make estimations and/or simulations of variables in non-sampled places.

Thus, the geostatistical methodology seeks to extract, from an apparent randomness of the collected data, the probabilistic structural characteristics of the regionalized phenomenon, that is, a correlation function between the values located in a certain neighborhood and direction in the sampled space.

Kriging

Known the semivariogram of the variable having spatial dependence between the samples, it is possible to do the interpolation of the values in any position in the field of study, without tendency to the minimum variance (Vieira 2000).

The interpolation method is called kriging and is based on the sample data of the regionalized variable and the structural properties of the semivariogram obtained from these data, which allows visualizing the behavior of the variable in the region through a map of isolated or of surface.

In the case that the value of the variables was selected in such a way that the obtained estimate is non-biased, the kriging estimator becomes the best non-biased linear estimator (BLUE = Best Linear Unbiased Estimator), since it presents minimal variance and does not tendentious to ensure that the sum of the weights is equal to unity.

According to Landim (2002), if a regionalized variable n (i) is collected at various points i, the value of each point will be related to values obtained from points located at a distance Δh and the influence will be greater the smaller the measure distance between points. (S), ... n (n), ... n (n), realizations of a regionalized variable, the non-biased estimate of semivariance is given by:

$\gamma(h) = 1/2n \sum\{v(i+h) - v(i)\}^2$

Such relations are shown when the function g (h) is placed in graph against Δh to originate the semivariogram and this does not present a tendency. This expresses the spatial behavior of the regionalized variable and shows, according to Figure 2:

a) the amplitude (range) (a), which indicates the distance from which the samples pass to have no spatial correlation and the relationship between them becomes random; any sample whose distance to the point to be estimated is less than or equal to the amplitude provides information about the point;

b) the sil (C), which indicates the value according to which the function is stabilized in the random field, corresponding to the distance "a"; shows the maximum variability between pairs of values, that is, the variance of the data and, consequently, the null covariance;

c) the continuity (Co), by the form of the variogram, in which for h = 0, g (h) already has some value. This situation is known as the nugget effect and is represented by Co; the nugget effect can be attributed to measurement errors or to the fact that the data were not collected at sufficiently long intervals, to show the underlying spatial behavior of the phenomenon under study;

d) anisotropy, when the semivariograms show different parameters for different sampling directions.

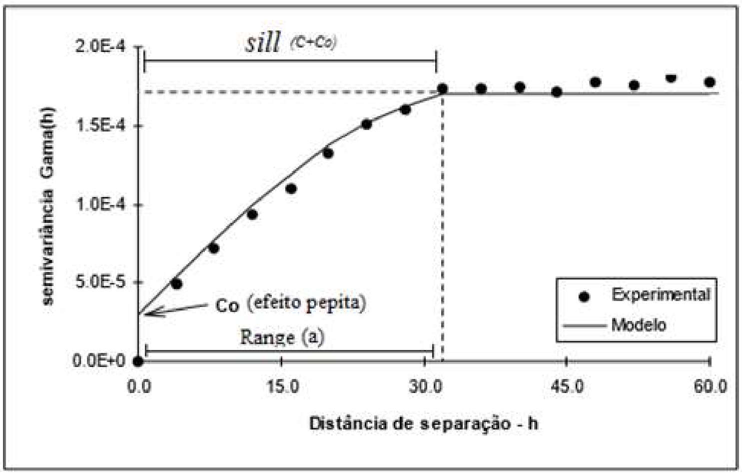

Figure 2. Experimental semivariogram and adjusted mathematical model. Source: Landim (2002)

Some models of variography follow, second Lembo and Magri (2002) (Figure 3):

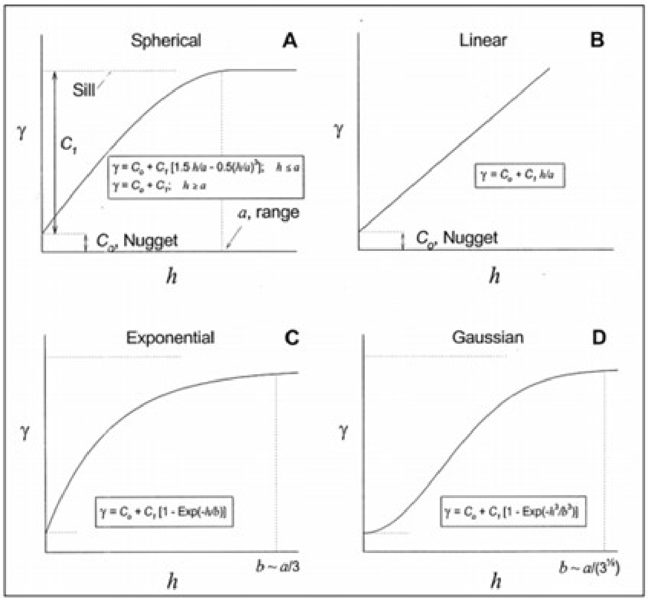

Figure 3. Models and components of the variogram: a) Spherical; b) Linear; c) Exponential; d) Gaussian. Source: Lembo and Magri (2002)

The kriging uses information from the variogram to find the optimal weights to be associated with the samples with known values that will estimate unknown points. In this situation, the method provides, in addition to the estimated values, the error associated with the estimation, which distinguishes it from the other interpolation algorithms. It is understood as a series of regression analysis techniques that seeks to minimize the variance estimated from a previous model, which takes into account the stochastic dependence between the data distributed in space. Therefore, among the estimation methods, commonly used, the geostatistical method of krigation can be considered as the best linear estimator without prejudice (as cited above), whose objective is to minimize the variance of the estimate.

The most common forms are simple irrigation and ordinary irrigation and among the non-linear methods the indicative kringage stands out. Simple irrigation is used when the average is assumed to be statistically constant for the whole area. The ordinary kriging, in turn, considers the floating or moving average throughout the area. The indicative kriging basically consists in the application of the ordinary kriging for the transformed variable. The concept of the indicative transformation is the most simple and friendly, since the indicative variograms are the easiest to model (Landim 2002).

Krajewski and Gibbs (1966), show the following comparison between interpolation methods (Table 1).

Table 1. Comparison between methods of interpolation of amostral dices.

Algorithm	Fidelity to original data	Smoothness of curves	Computation speed	General accuracy
Triangulation	1	5	1	5
Distance reverse	3	4	2	4
Surface / Trend	5	1	3	2
Minimal curvature	4	2	4	3
Kriging	2	3	5	1

Caption: 1 = better; 5 = worse. Source: Krajewski and Gibbs (1966); Thus, the kriging is shown once again as the most accurate method.

Sampling

A basic requirement in sampling for the purposes of analyzing spatial dependence using geostatistical methods is that the observations, that is, that the samples are georeferenced. It is not necessary to use geographic coordinates, but there is some type of reference.

Guimarães (2004) exemplifies types of references in geostatistics:

a) Samples taken over time: each observation is referenced with respect to time (for example Study of annual precipitation in region "X");

b) Samples collected along a straight line: in a particular agricultural culture, each observation is referenced by a single point in space (for example, samples collected in trance);

c) Samples taken in an area: each observation will be identified by an ordered pair of coordinates belonging to space (eg, samples collected in an area X).

Regardless of the type of sample collected for analysis, the greater the number of points sampled, the greater will be the number of pairs for the calculation of semivariaries and, theoretically, the greater the precision of the semivarian estimates.

It can be said that the number of observations will depend on the objectives of the work, the scale (that is, the dimension), among the other factors that must be evaluated by the researcher.

Another important point to be addressed in georeferenced statistics is the number of samples that should be used for good experimental precision.

Some authors recommend that at least 100 sample points be used, however, this is not a rule and yes recommendation, there are works with good results of adjustment of semivariograms using only 45 sampling points (Guimarães 2004).

Principles of geostatistical analysis

In parametric statistics, the principles of chance, repetition and local control are adopted (with the exception of completely random delineation) and the sample data must precede from population parameters. In addition, this methodology considers, in its development and application, the following assumptions: normality of the variable; the independence of errors and the homoscedasticity of the variance (homogeneity of the variance). In nonparametric statistics, as there is an absence of population parameters, the analysis must be based on the measurement scales: nominal, ordinal, interval and ratio, in the "k-samples" to determine the "p-value" p-value). In geostatistics, the principles of regionalized variables are adopted and fundamentally based on the following assumptions: ergodicity, stationarity and intrinsic hypothesis.

In addition, in spatial statistics, it is assumed that the samples must have values associated with the location in space (and/or time), with the continuous distribution of the values and unsampled points must be estimated (with the use of the interpolation of the data). Therefore, it is adopted as spatial statistics procedures:

a) Exploratory analysis of the data;
b) The calculation of the experimental variogram;
c) Modeling;
d) The kriging: estimation and interpolation;
e) The simulation.

Landim et al (2002) assume that:

- Ergodicity: all possible events were performed within a certain domain; and that the mathematical hope of these possible realizations (of the variables) is the same.

- Stationarity: in the region where the estimations are intended, the phenomenon is described as homogeneous;

- Intrinsic hypothesis: the differences between values show a slight increase, that is, the differences are locally stationary.

Thus, the three basic types of distribution are aggregate, uniform and random. The simplest view of a spatial pattern can be obtained by adopting an orientation from any individual and asking the following question: given the location of an individual, what is the probability that another individual is near (Krebs 1989). There are three possibilities:

a. The probability is high: pattern added (Figure 4 a);
b. The probability is low: random pattern (Figure 4 b);
c. The probability is not affected: uniform pattern (Figure 4 c).

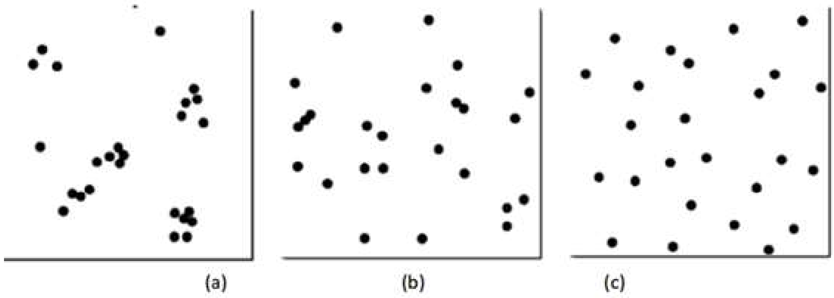

Figure 4. Three main types of spatial distribution: (a) aggregate, where all individuals tend to be distributed in groups; (b) random, where all individuals are located one independent of the other; (c) uniform, where individuals appear regularly distributed in space. Source: (Krebs 1989).

Trend of surface analysis

Another method for data mapping and isoline formation is by trend surface analysis. Nothing else is a polynomial regression statistic that traces trend lines from sample data. Trend surfaces can be used to interpolate values, extrapolate the sequence of data, infer about the presence of trends or estimate characteristics of interest on the data (Vieira 1998). Depending on the number of available data, surfaces of any size can be calculated.

The trend surface analysis is a method by which a continuous theoretical surface is adjusted by least squares regression criteria, to the values of the dependent variable (Zi), having as independent variables the north-south coordinates (y) and east-west (x). For the application of this technique, the mathematical equation used for the adjustment of the surface is based on non-orthogonal polynomials, with the adjustment being increased by the addition of additional terms to the polynomial equation.

In addition, according to Bernardi et al. (2001) the analysis of trend surfaces can be a very useful tool for the evaluation of environmental impacts derived from anthropic action. Well, in addition to presenting regional trends are able to highlight localized anomalies. Considering that the relationships found in nature are graduated, the values of a variable can be partially predicted by neighboring points, being dependent on each other.

Isoline maps

An isoline is a line that unites a set of points with the same value (isovalues), while an isolation map is simply the representation of a surface by means of "isovalue" curves. One of the best examples of isolation maps is the contour map where each level curve is an insulator that represents the set of points with the same altitude. The other types of isolation maps are the various types of maps of climatic distributions (isothermal, isobar, isohyetic, isoneas), water salinity (isoaline), precipitation (isohyetics), various types of densities (population density, the percentage of cultivated land, etc.), among many other cases. Actually, an isolation map can be used to represent any feature that can be quantified and represented by a set of points.

Practical application of spatial distribution

The studies on the productivity and the accumulation of burlap that are carried out are used in their majority of classic methods of statistical analysis of data, which, in general, suppose that the realizations of the random variables are independent of each other (Snedecor and Cochran 1967). However, natural phenomena commonly occur with some structuring in the variations between neighbors and may present some degree of spatial dependence (Guimarães 1993). Therefore, the distribution of the vegetation of a certain place can present spatial continuity (Miller et al. 2007) and, consequently, classical statistical methods may not be good estimators for studies of productivity and accumulation of burlap.

Aquino (2013), studying the spatial analysis of the sacking (leaf and total fraction) by the kriging method in the bush region, verified that it was possible to visualize regions from which it can be defined as zones of management, which can help in the decision making regarding the transposition of the burlap, the collection of seed bank for the recovery of areas, among others. This result is very important to deal with sustainable work in native areas and environmental preservation.

Barbosa (2017), when evaluating the spatial distribution of cacti species in Caatinga areas, observed that some native species behave in an aggregate manner, while others present uniform distribution within the Caatinga areas. In this way, the management can be differentiated as regards the native species present in the areas.

Another way to evaluate the spatial distribution (or variability) in forest areas or biomes (eg Caatinga) is through the floristic/phytosociological survey, however, in this case, the elaboration of contour maps is not mandatory (with isolated), one since the calculations are made to evaluate the density (relative and absolute) and frequency among plant species, among other parameters. The evaluation of some of these parameters can be found in studies done by Santos (2012), Barbosa (2017), Araújo (2010), Parente (2009), among others. These studies are based on the evaluation of the plots through the sampling of the variables in areas treated in pickets of 10m² or 1 hectare (10,000m²) for the arboreal/shrub layer or in plots of 1 m², equidistant for the evaluation of the stratum herbaceous (Rodal et al 1992).

The elaboration of the contour maps has been widely used for the evaluation of the physical attributes of the soils of a certain region and, often, used to characterize the spatial patterns of the attributes of the soil (Zhu and Shao 2008, Brocca et al 2007, Cichota et al 2006, Hébrard et al 2006, Souza et al 2004, among others). But recently, these maps have been used to evaluate other variables that can be regionalized, such as the distribution of vegetative strata in certain degraded areas, distribution of CO_2 released by the soil in areas of Caatinga, distribution of the sacking, among others.

Thus, more and more geo-referenced statistics are applied to several areas of study in order to obtain new technical information and contribute even more to reach optimal levels of productivity allied to sustainability.

Final Considerations

Given the previous observations, it is easier to understand the importance of the practical application of georeferenced statistics with the evaluation of the spatial distribution of population samples for a better understanding of production systems and environmental characterization, to maintain a sustainable production without causing damage to the environment.

References

Aquino PSR (2013) Análise espacial da produtividade e acúmulo de serapilheira em mata de galeria. Dissertation, Universidade de Brasília, DF.

Araújo KD (2010) Análise da vegetação e organismos edáficos em áreas de caatinga sob pastejo e aspectos socioeconômicos e ambientais de São João do Cariri. Thesis, Universidade Federal de Campina Grande, PB.

Barbosa AS, Andrade AP, Júnior LRP, Bruno RLA, Medeiros RLS, Neto MAB (2017) Estrutura populacional e espacial de Mandacaru (*Cereaus jamacaru* DC) em duas áreas de Caatinga do agreste da Paraíba-Brasil. Ciência Florestal 1:325-324.

Bernardi JVE, Fowler HG, Landim PMB (2001) Um estudo de impacto ambiental usando análises estatísticas espacial e multivariada. Holos Environment 27:315-324.

Brocca L, Morbidelli R, Melone F, Moramarco T (2007) Soil moisture spatial variability in experimental areas of central Italy. Journal of Hydrology 333:356-373.

Cichota R, Hurtado ALB, Lier QJV (2006) Spatio-temporal variability of soil water tension in a tropical soil in Brazil. Geoderma 133:231-243.

Clark I (1979) The semivariogram - Part I. London.

Genú AM (2004) Geoestatística multivariada. Texto didático. Universidade de São Paulo pp. 2-17.

Gumarães EC (2004) Geoestatística Básica e Aplicada. Texto didático. Universidade Federal de Uberlândia pp. 1-73.

Guimarães EC (1993) Variabilidade espacial da umidade e da densidade do solo em um Latossolo Roxo. Dissertation, Universidade Estadual de Campinas.

Hébrard O, Voltz M, Andrieux P, Moussa R (2006) Spatio-temporal distribution of soil surface moisture in a heterogeneously farmed Mediterranean catchment**.** Journal of Hydrology 329:110-121.

Huijbregts CJ (1975) Regionalized variables and quantitative analysis of spatial data. In: Davis, JC and McCullagh, MJ (ed) Displayand anaysis of spatial data. New York, John Wiley. pp.38-53.

Isaaks EH, Srivastava RM (1989) An Introduction to Applied Geostatistics. New York: Oxford University Press.

Krajewski SA, Gibbs BL (1966) Understanding Contouring: A pratical Guide to Spatial Estimation and Contouring Using a Computer and Basics of Using Variograms: Gibbs Associates. Oxford, NY.

Krebs CJ (1989) Ecological Methodology. New York, NY.

Landim PMB, Sturaro JR, Monteiro RC (2002) Krigagem Ordinaria para Situações com Tendência Regionalizada. Accessed in: May 13, 2011

Lembo AJ, Magri A (2002) Geostatistics. Spatial Modeling and Analysis – CSS. Cornell University, Ithaca, NY, pp 620.

Miller J, Franklin J, Aspinall R (2007) Incorporating spatial dependence in predictive vegetation models. Ecological Modelling 202:225-242.

Parente HN (2009) Avaliação da vegetação e do solo em áreas de caatinga sob pastejo caprino no Cariri da Paraíba. Thesis, Universidade Federal da Paraíba.

Rodal MJN, Sampaio EVSV, Figueiredo MA (1992) Manual sobre métodos de estudo florístico e fitossociológico: Ecossistema de Caatinga. Brasília-DF.

Santos JT (2012) Atributos físico-químicos do solo, dinâmica da serrapilheira e composição bromatologica de espécies da Caatinga sob pastejo caprino. Thesis, Universidade Federal da Paraiba.

Salviano AAC (1996) Variabilidade de atributos de solo e de Crotalaria juncea em solo degradado do município de Piracicaba-SP**.** Thesis, Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo.

Snedecor GW, Cochran WG (1967) Statistical methods. Ames, Iowa.

Silva AP (1988) Variabilidade espacial de atributos físicos do solo. Thesis, Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo.

Souza ZM, Marques Junior J, Pereira GT, Barbieri DM. Variabilidade espacial da textura de um Latossolo Vermelho eutroférrico sob cultivo de cana-de-açúcar. Engenharia Agrícola 24:309-319.

Vieira SR (2000) Geoestatística em estudos de variabilidade espacial do solo. In: Novais RF, Alvarez VVH, Schaefer CEGR (Ed.). Tópicos em ciência do solo. Viçosa, MG, pp 1-54.

Vieira MB (1998) Vetorização e análise de tendência de cartas de lineamentos geológicos**.** Dissertation, Universidade Federal de Minas Gerais.

Zhu Y, Shao M (2008) Variability and pattern of surface moisture on a small-scale hillslope in Liudaogou catchment on the northern Loess Plateau of China. Geoderma 147:185-191.