Theoretical Estimation of the Sampling Size of Geostatistics considering Gaussian Variogram Model

André Mendes[1†], Gerson Rodrigues dos Santos, Paulo Cesar Emiliano, Patrícia de Sousa Ilambwetsi, Amy Leigh Kaleita

[1] *Departamento de Estatística, Centro de Ciências Exatas, Universidade Federal de Viçosa.*
[2] *Iowa State University, USA.*

Resumo: *Na Geoestatística Clássica existe uma grande necessidade de pesquisas que criem e/ou investiguem métodos de amostragem de dados geoespaciais. Além da complexidade do assunto, alguns trabalhos apresentam soluções que utilizam mecanismos teóricos e práticos de diferentes áreas do conhecimento científico que atendem demandas específicas de pesquisadores da área. O objetivo deste artigo é utilizar a Teoria da Informação de Sinais Elétricos, principalmente considerando o Teorema da Taxa de Nyquist, para determinar um tamanho ótimo para amostras georreferenciadas que usam grade quadrática regular, no qual o modelo de dependência espacial é o gaussiano. O que se deseja alcançar teoricamente é uma densidade de amostragem necessária para a reconstrução de mapas populacionais de variáveis nas quais as condições de regularidade necessárias em geoestatística foram verificadas, a saber: estacionariedade de 1ª e 2ª ordem e/ou estacionariedade do variograma, ausência de outliers e tendências, e variograma isotrópico. Como resultado, pode-se afirmar que a distância máxima entre os pontos da grade regular quadrática é de aproximadamente 30% do alcance prático observado no variograma da primeira amostragem experimental.*

Palavras-chave: Geoestatística; Amostragem; Taxa Nyquist.

Abstract: *In Classical Geostatistics there is a great need for research that creates and/or explores methods of geospatial data sampling. In addition to this complex subject, some papers present solutions that use theoretical and practical mechanisms from different areas of scientific knowledge addressing specific demands of researches in the field. The purpose of this paper is to apply the Electrical Signal Information Theory, especially considering the Nyquist Rate Theorem, to determine an optimal size for georeferenced samples using a regular quadratic grid based on the spatial dependence Gaussian model. We expect to achieve a necessary sampling density to rebuild population maps of variables in which the following regularity conditions required in geostatistics were certified: 1st and 2nd order stationarity and/or stationarity of the variogram, absence of outliers and trends, and isotropic variogram. As a result, we can state that from the data set used, the greatest distance between points of the regular quadratic grid is nearly 30% of the practical range observed on the variogram of the first experimental sample.*

Keywords: Geostatistics; Sampling; Nyquist Rate.

[†] Correspondent author: amendesmat@yahoo.com.br.

1. Introduction

For any scientific data analysis, it is extremely important to highlight value information (stated as representativeness), especially when we are dealing with samples (OLIVEIRA et al., 2014). This performance is also required when information regarding time and space is incorporated in the process (stated as regionalization) (YAMAMOTO; LANDIM, 2013). The aspect of regionalization of variables in a process is the reason why spatial statistics conquers enthusiasts all over the world. It happens mainly because questions can be clarified as considers the georeferenced location of the target variable.

Cressie and Wikle (2011) stated that regionalized information has always been essential to human being. When the topic was related to survival and/or achievements, the civilizations applied this mechanism to improve the display of orientation systems.

Yamamoto and Landim (2013) declared that Classical Statistics is important for many studies, although they have to assume a set of assumptions that are difficult or even impossible to verify, such as the requirement for random samples and the knowledge of the probability distribution of the variable studied.

The estimation of sample size in Spatial Statistics is one of the problems cited by Webster and Oliver (1992), Souza et al. (2014) and Sartori and Zimback (2011).

Hoef (2002) and Clark (2009) present another problem of Classical Statistics with regard to the sampling size needed to obtain the same precision as the Spatial Statistic: approximately 9 times higher.

Pinto, Santos, and Oliveira (2014) suggested the subdivision of Spatial Statistics. Among them, the Geostatistics is highlighted as declared by Clark (1979), Armstrong (1998), Brooker (1991), Clark and Harper (2000), Druck et al. (2004), Olea (2009), Webster and Oliver (2007), Diggle, Menezes, and Su (2010), Santos et al. (2011), and Ferreira, Santos, and Rodrigues (2013).

Santos et al. (2011) mentioned the advantage of using Geostatistics modelling phenomena that do not exhibit the assumptions of Classical Statistics. Geostatistics uses the sampled neighbourhood sampled in order to reinforce the perception that the spatial dependence structure in a phenomenon to improve interpolations that are not a trend and have the least variance. Besides that, Vieira (2000) demonstrates that is still possible to recognize the uncertainty around predictions throughout the kriging variance.

Yamamoto and Landim (2015) registered the studies from distinct fields using Geostatistics as the main tool to analyse sampled data. Following this trend, plenty of papers concerning the size of geostatistics sampling are rising such as Brus and Heuvelink (2007), Modis and Papaodysseus (2006), Peigne et al. (2009), Vasat, Heuvelink and Boruvka (2010), Diggle, Menezes and Su (2010), and others.

Modis and Papaodysseus (2006) calculated (measured) the theoretical size to obtain an ideal density sampling based on the Information theory, formerly developed to Electrical Signals (WHITTAKER, 1915; SHANNON, 1949).

The paramount of this research is to apply Geostatistics in association with the Electrical Signal Information Theory, mainly considering the Nyquist Rate Theorem, in order to establish an ideal sampling size to those geostatistics/researchers. Those scientists use regular quadratic grid based on the spatial dependence Gaussian model. Then, our aim is to obtain a theoretical sampling density required to rebuild populational maps of the chosen variable in which all the regularity conditions required by Geostatistics were certified: 1st and 2nd order stationarity and/or stationarity of the variogram, absence of outliers and trends, and isotropic variogram.

The literature employed to our aim are: Modis and Papaodysseus (2006) introducing the theoretical methodology applied to spherical and exponential models; Yfantis et al. (1987) demonstrating the theoretical efficiency between different sampling grids; Vasat et al. (2010) indicating an alternative to reduce the sampling size in a multivariate process; Ferreira, Santos and Rodrigues (2013) explaining a systematic and interactive methodology of the geostatistics data analysis; and Santos et al. (2011) suggesting a comparative precision study with maps produced by distinct linear geostatistical interpolators.

2. Material and methods

Hereafter, we are displaying the methodology adopted in this study and also the description of the researched area. We are also describing the purpose of the method and introducing the data.

2.1 Description of the research area

We analysed an area comprising 5.7 km^2 in the municipality of Viçosa, belonging to the "Zona da Mata" region of Minas Gerais State, Brazil. The site is surrounded by latitudes 20º45'39'' S and 20º46'53'' S, and the longitudes 42º52'24'' W and 42º53'49'' W as seen in Figure 1.
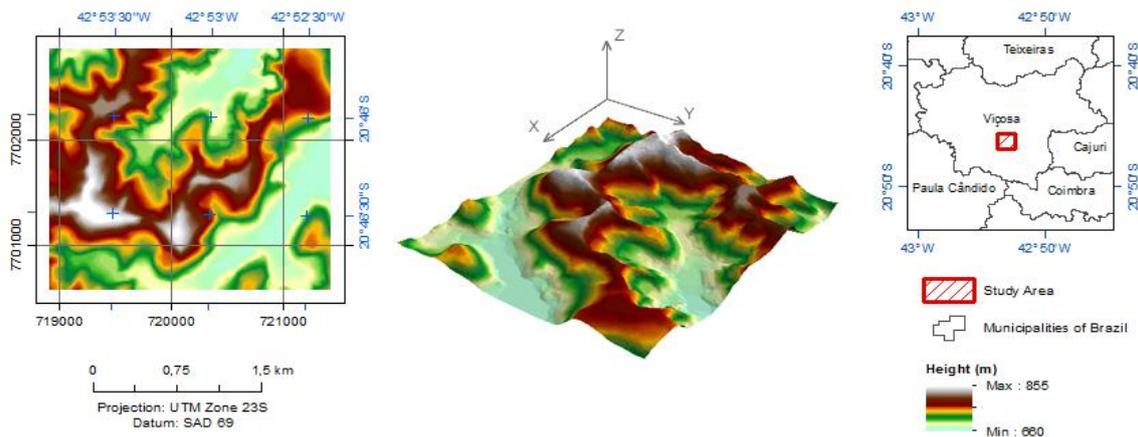

Figure 1 - Study site encompassing an area of 5.7 km^2 in the municipality of Viçosa, in Minas Gerais State, Brazil.

The altimetry data used are in accordance with the geodetic reference system Sirgas 2000 and the UTM (Universal Transverse Mercator coordinate system) projection system zone 23 S. Those data correspond to roughly 230 thousand spot heights known, which are interspersed around 5 meters both in X and Y directions. The minimum height recorded is 660 meters while the maximum is 885 meters (Figure 2).
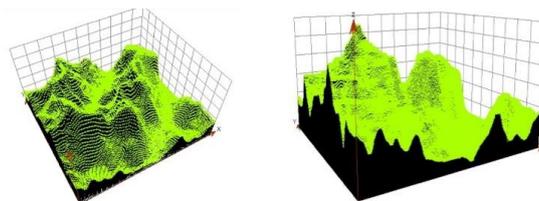

Figure 2 - Three-dimensional representation of the altimetry survey from a section in "Zona da Mata" in the municipality of Viçosa, Minas Gerais State, Brazil with roughly 230 thousand spot heights.

2.2 The methodology purposes

Modis and Papaodysseus (2006), based upon the Theorem of the Electrical Signal Information Theory called Nyquist Rate and also the Fourier transformed to the Geostatistics correlogram functions, proved that is possible to achieve a sampling size suitable to Geostatistics aiming a complete rebuild of the population studied.

However, those authors presented only one solution to the theoretical models of both spherical and exponential variograms. As a final suggestion (MODIS; PAPAODYSSEUS, 2006), a practical algorithm was proposed:

i)   Start with an appropriate sampling density;
ii)   Regulate to an equivalent covariance model;
iii)   Establish the superior practical upper limit of the spectrum from the model (Nyquist Rate);
iv)   Based on iii, determine an ideal sampling density;
v)   If the ideal sampling density is not achieved, the sampling process should be repeated by preference;
vi)   If the ideal density is economically unattainable, the available ways should be recalculated.

Indeed, the spherical and exponential models are frequently applied in many fields. the Gaussian model, defined by equation 1, has a great value to other regionalized variables.

$$\gamma(h) = \begin{cases} 0 & \text{if} |h| = 0 \\ c\left[1 - exp\left(-\frac{3|h|^2}{a^2}\right)\right] & \text{if} |h| > 0 \end{cases} \tag{1}$$

In the Gaussian model form equation 1, c represents the level, h is a distance vector between the points and a the range of spatial dependence.

According to Ferreira, Santos, and Rodrigues (2013), those variograms adjusted by the Gaussian model are characterized by a spatial dependence with low deviations among closer neighbours and higher ones to those who are more distant. All of them inside the variogram range. Those features are typical to variables like altimetry and bathymetry supporting the use of the first one in this survey.

In compliance with Modis and Papaodysseus (2006) and based on the Theorem of Nyquist Rate, the rebuilding of population maps is possible only if the sampling density adopting regular grids is greater or the same as a critical value. This value can be calculated in regard to the correlation function under some conditions such as stationarity, isotropy, strong spatial dependence, and finite range.

As mentioned before those authors introduce the theory and its results only for spherical and exponential models. Moreover, they demonstrated that for the variables in the mining field (homogeneous ore) the practice resembles the Geostatistics theory and the sampling density must be greater or the same as half of the effective reach of the experiment, all regulated in the variogram.

Following this idea, this study focuses specifically on developing the Nyquist Rate Theory to the Gaussian model, demonstrating this theory could be applied to altimetry, and finding the sampling density for this model.

As part of equation 1, Abramowitz and Stegun (1972) suggested the equation 2 using the Fourier Transform in the correlation function of the Gaussian model, which is inversely related to the variogram.

$$R(\omega) = exp\left(-\frac{3t^2}{a^2}\right)cos(\omega t) \qquad (2)$$

where ω is the sampling rate related to the signal frequency, and t is a sampling instant.

The same authors stated that the power spectrum of the random subjacent function is in equation 3. This equation is the model which displays the performance of the correlation function in the Gaussian model.

$$S(\omega) = \frac{a}{2}\sqrt{\frac{\pi}{3}}\,exp\left(-\frac{\omega^2 a^2}{12}\right) \qquad (3)$$

As $S(\omega)$ tends to zero due to stabilization, $exp\left(-\frac{\omega^2 a^2}{12}\right)$ ends on zero for ω infinite. Nevertheless, Journel and Huijbregts (1978) stated the Gaussian model should have its nullity on 0.05 since it is an asymptotic axis. Therefore, $\omega = \frac{6}{a}$.

The sampling size T of the sampling Theorem (Nyquist Rate) is given by equation 4.

$$T \leq \frac{\pi}{\omega} = \frac{\pi}{\frac{6}{a}} = \frac{\pi}{6}a \qquad (4)$$

which means almost half of the theoretical achievement as the rate claimed by Modis and Papaodysseus (2006) as shown in Figure 3.
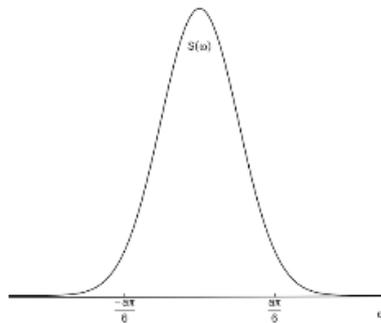

Figure 3 - Depiction of Fourier Transform to the Gaussian correlation model.

Nonetheless, Olea (1999) reported that some variografic models including the correlogram do not reach stabilization of the curve in the theoretical range a. Furthermore, the Gaussian model is one of them. Thereby, this author demonstrates the transformation from the theoretical to the practical achievement a_p throughout:

$$a_p = \sqrt{3}a \Rightarrow a = \frac{\sqrt{3}}{3}a_p \qquad (5)$$

Accordingly, the sampling size T in terms of practical achievement is given by equation 6.

$$T \leq \frac{\pi\sqrt{3}}{18}a_p \qquad (6)$$

The result of this equation indicates that the greatest distance between two points in a sampling regular quadratic grid must be nearly 30% of the practical achievement. This means that a first sample called experimental sampling should aim a 30% sampling density in the practical achievement if the regularity conditions to the Gaussian correlogram and consequently to the variogram are approved.

Examining a regular quadratic grid, we can notice the greatest distance between points is on the diagonals. Seeing that, the size of the maximum distance for sides should be converted before using the ratio $d = S\sqrt{2}$, where d meaning diagonal and S the square size.

2.3 Data description

The data set of a Brazilian region sited in Zona da Mata, municipality of Viçosa, Minas Gerais State, Brazil comprised 229,414 points. In order to verify the theory used in this study, the group aforementioned was reduced (following and checking the regularity conditions required for the method) 17 times reaching a sampling size of 49 points.

As an efficiency criterion, we considered kriging variance, the coefficients of Simple Linear Regression (SLR) among predicted and observed values of cross-validation, and data variance to the parameter assessment in variogram modelling (VIEIRA, 2000; MODIS and PAPAODYSSEUS, 2006; SANTOS et al., 2011; FERREIRA, SANTOS and RODRIGUES, 2013; YAMAMOTO and LANDIM, 2013).

Due to some software limitations regarding data set input, we adopted ArcGIS (ESRI, 2014) to all computational analyses. In order to reduce the sampling size, we applied the regular selection of altimetry data using the regular sampling tool of ArcGIS. Firstly, we defined spatial arrangement in X and Y directions to employ the regular selection. As a result, the program created an intermediate grid of points based on the established arrangement, and also assembled the closer point of the altimetry database to each point produced in the intermediate grid. Afterward, we selected those closer points gathering an "almost regular" altimetry data set. As long as this study has a big data set with a high density of points, this "almost regular" can be considered as a regular selection. This fact was also proved by the Proximity Analyzer tool using the command Near in the Toolbox (ESRI, 2014).

Moreover, the simple kriging output (SANTOS et al., 2011) of all altimetry data set created the population maps of sites studied in this survey according to Figure 4.
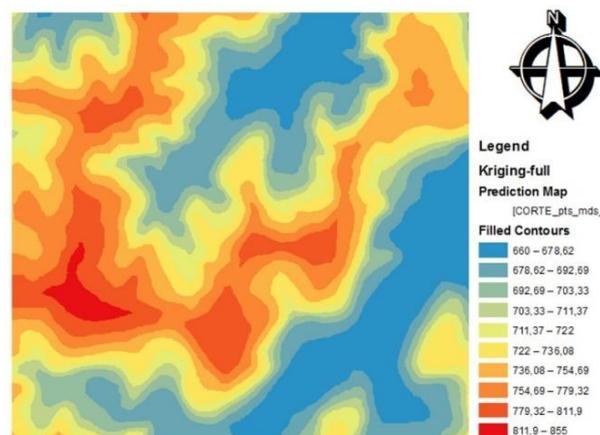

Figure 4 - Simple Kriging of altimetry data from a section of "Zona da Mata" in the municipality of Viçosa, Minas Gerais State, Brazil.

3. Results and discussion

Observing Table 1, we can notice that the averages and variances have not varied broadly. Only when the sampling variation was 99.96% and 99.98% smaller, we registered a larger range of deviation on those measurements reaching on this last sample the minimum limit of sampling size for Geostatistics as recommended by Yamamoto and Landim (2013).

The reductions were performed according to the regular spatial arrangement between points chosen in quadratic grids with sizes of 7 m, 10 m, 15 m, 20 m, 25 m, 30 m, 40 m, 50 m, 70 m, 90 cm, 130 m, 160 m, 190 m, 220 m, 260 m, 300 m, and 340 m.

The analytical data displayed in Table 1 indicate that decisions on sampling representation cannot be accomplished exclusively by size, average, and variance.

Table 1 - Sampling sizes, sample reduction, averages, variances, and spatial arrangement of quadratic regular grids of the altimetric survey from a section in "zona da mata" region near to Viçosa, in Minas Gerais state, Brazil.

SamplingSize	Reduction (%)	Average	Variance	Spatial Arrangement
229,414	0.00%	721.47	1,473.48	"population"
116,708	-49.13%	721.50	1,474.94	7 meters
57,228	-75.05%	721.46	1,474.94	10 meters
25,384	-88.94%	721.57	1,475.87	15 meters
14,364	-93.74%	721.37	1,473.79	20 meters
9,292	-95.95%	721.63	1,476.17	25 meters
6,308	-97.25%	721.65	1,477.33	30 meters
3,591	-98.43%	721.36	1,474.10	40 meters
2,300	-99.00%	721.49	1,474.48	50 meters
1,188	-99.48%	721.23	1,466.05	70 meters
700	-99.69%	721.40	1,483.56	90 meters
342	-99.85%	721.26	1,460.84	130 meters
224	-99.90%	721.16	1,498.39	160 meters
156	-99.93%	721.96	1,538.68	190 meters
110	-99.95%	721.99	1,514.61	220 meters
90	-99.96%	719.88	1,477.86	260 meters
64	-99.97%	721.78	1,569.59	300 meters
49	-99.98%	719.12	1,390.62	340 meters

Figures 5, 6, and 7 exhibit the three-dimensional picture of sampling regular grids exposing the representative loss of population, when the sample does not have a suitable size suggested by the criteria being adopted.
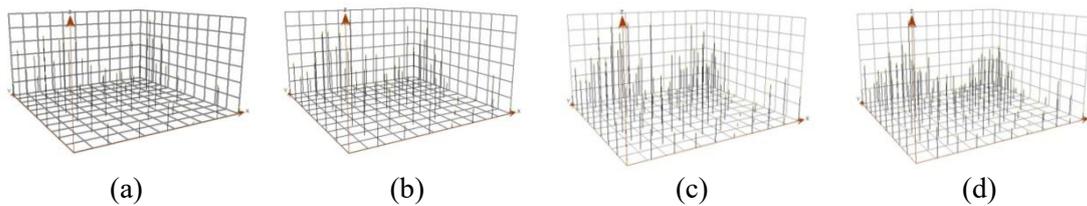
(a) (b) (c) (d)

Figure 5 - Three-dimensional sampling figure of the quadratic regular grid of the altimetric survey in a section of "Zona da Mata" in the municipality of Viçosa, Minas Gerais State, Brazil. The grids displayed (a) 48 points, (b) 64 points, (c) 90 points, and (d) 110 points.
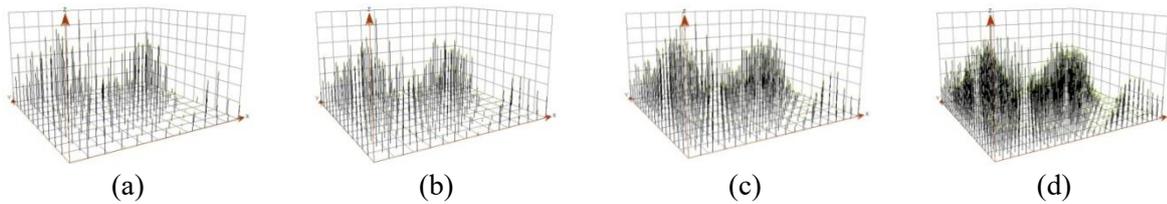

(a) (b) (c) (d)

Figure 6 - Three-dimensional sampling figure of the quadratic regular grid of the altimetric survey in a section of "Zona da Mata" in the municipality of Viçosa, Minas Gerais State, Brazil. The grids displayed (a) 156 points, (b) 224 points, (c) 342 points, and (d) 700 points.
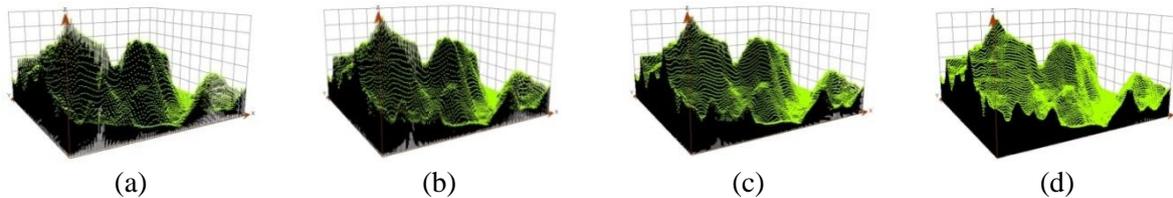

(a) (b) (c) (d)

Figure 7 - Three-dimensional sampling figure of the quadratic regular grid of the altimetric survey in a section of "Zona da Mata" in the municipality of Viçosa, Minas Gerais State, Brazil. The grids displayed (a) 9,292 points, (b) 14,364 points, (c) 25,384 points, and (d) 57,228 points.

According to expectation, on the report of Oliveira et al. (2009) as the sampling size increases the simple three-dimensional view of data displays the real performance of population, what is observed in Figures 6 and 7.

Oliveira et al. (2009) stated that researchers are always concerned about establishing the major indicators of a typical sample to Classical Statistics. Nevertheless, those measurements did not overtake this aim by themselves.

Yamamoto and Landim (2013) suggested that although many and distinct trials, this survey highlights an equal matter of the subject including Spatial Statistics. Therefore, until we find the "best" indicators, it is essential to evaluate scientific surveys on this topic looking forward feasible mechanisms for that determination. Modis and Papaodysseus (2006), Clark (2000), and Yamamoto and Landim (2013) presented articles that have tried it since 1975. Based on those papers, we noticed the common procedure is to apply the kriging as the efficiency criteria.

Authors such Vieira (2000), Santos et al. (2011), and Ferreira, Santos, and Rodrigues (2013) employed the SLR (Simple Linear Regression) as a criterion along with the kriging variance among kriging predicted and observed values after the cross-validation process.

As determined by Vieira (2000), and Ferreira, Santos, and Rodrigues (2013), the theoretical slope must be 1. Notwithstanding, we must notice how near this value is from the theoretical one. The RLS of this slope in function of kriging average variance (RMS) established the estimated model $\hat{Y} = -1,75X + 1,72$ with a determination coefficient of $R^2 = 0,9593$. That means about 96% of RMS variation can be explained by the slope deviation.

Modis and Papaodysseus (2006) stated that the spatial indicators adopted in scientific research of this field are a function among the kriging medium variances and the sampling density under the same conditions and different sizes conserving equivalent sample conditions of the examined variable (as the sample exhibited in this paper). Those authors further claim that the ideal sample size was determined by the stabilization of the curve fitted to the graph. Furthermore, the difference among average variances of neighbours did not improve precision performance throughout kriging.

The graph of this function to the examined area on Figure 8 displays a stabilization of the kriging average variance named as RMS (Root Mean Square) around the sampling size 156, 224, and 342 points.

Among them, we could choose between sample 224 or 342. However, following the theory presented in this article, the ideal sampling size is 342 points (highlighted with an arrow in Figure 8), with the side distance between the points around 130 meters (or a diagonal of 190 meters) corresponding to 21.6% (or 31.5% if we consider the diagonal) regarding the practical reaching of 603 meters.
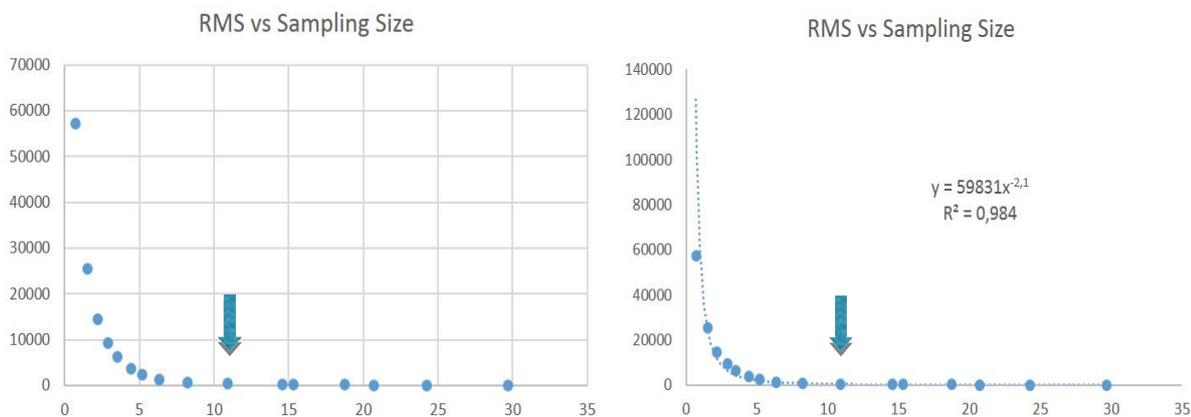

Figure 8 - Graphic demonstrating the relationship between kriging average variance and the sampling size of altimetry data from a section of "Zona da Mata", Minas Gerais State, Brazil. On the left, a dispersion diagram, and on the right, a diagram with a tendency line, the model and R^2.

In Figure 9, we are emphasizing variogram performance applied to distinct sampling sizes. Moreover, in Equation 1 the adapted the Gaussian model to experimental variograms.
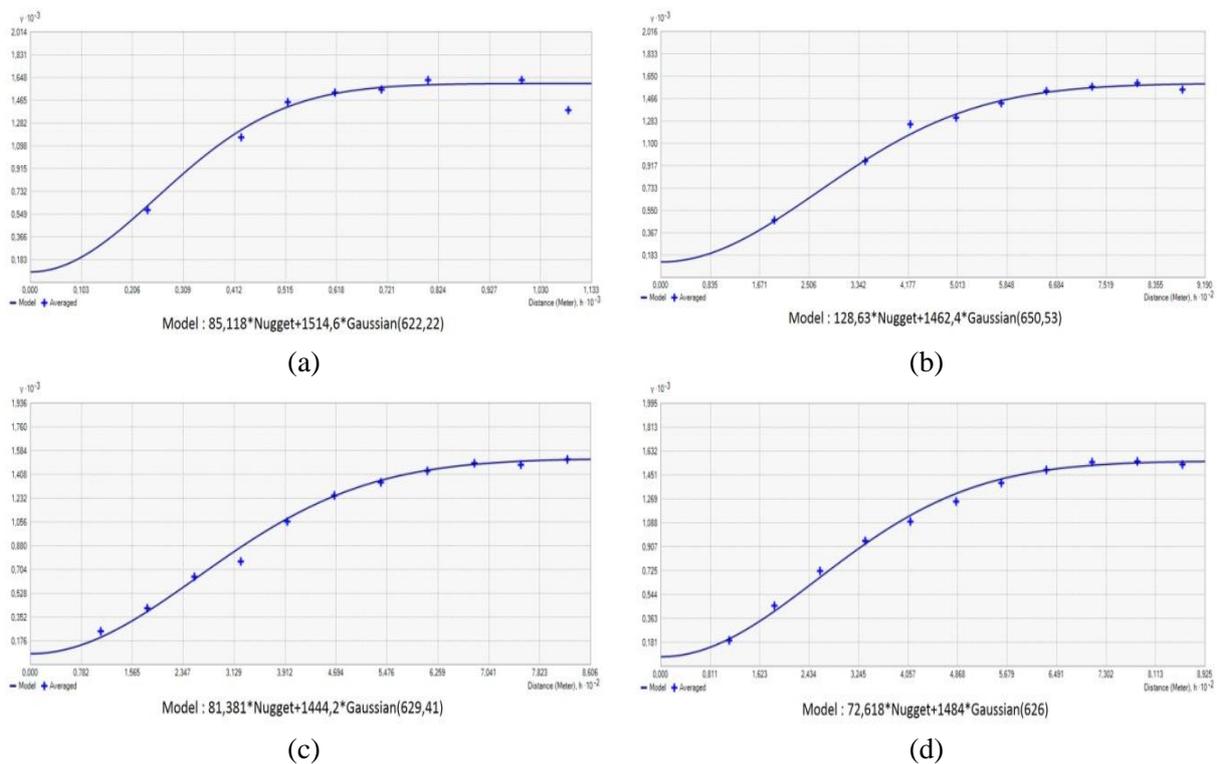
Model : 85,118*Nugget+1514,6*Gaussian(622,22)

(a)

Model : 128,63*Nugget+1462,4*Gaussian(650,53)

(b)

Model : 81,381*Nugget+1444,2*Gaussian(629,41)

(c)

Model : 72,618*Nugget+1484*Gaussian(626)

(d)

Figure 9 - Experimental Variograms ("+") and adapted Gaussian models (lines) to the spatial dependence of altimetry data from a section of "Zona da Mata", Minas Gerais State, Brazil. The sampling size are (a) 156 points, (b) 224 points, (c) 342 points, and (d) 700 points.

Clark (2000) said that many Geostatistic researchers believe the nugget effect can be produced by a low sampling density. Although variograms are displaying very similar values, the nugget effect increased significantly deviating since the densest grid up to the less dense grid from 42.51 m^2 to 460 m^2 (considering this unit is referring to the variance rather than the area).

The method proposed in this work highlights the reach as a unit of great importance, since the sampling density will vary theoretically in its function. However, here is the point of great concern of this type of study, because the range is not kept constant with the reduction of sampling size.

Modis and Papaodysseus (2006) dealt with homogenous ores and did not detect variation when estimating the variografic range. On the other hand, the range estimated here fluctuated since the most to the less dense grid from 603 to 650 meters.

Even though we identified oscillations on those ranges, the average lateral distance of the quadratic regular grid we suggest is 130 meters corresponding to 342 points.

The choice of this value is in accordance with the Gaussian model performed here, and all Geostatistic regularity conditions were also certified and proofed.

Santos et al. (2011) mentioned that as maps reproduce reality, people tend to accept them as truth. Therefore, maps production through data interpolation is a relevant step to Geostatistic analyses since they will be at least criticized for those who know the mapped region. Accordingly, Figures 10 and 11 display maps obtained based on interpolation by simple kriging as recommended by Santos et al. (2011).
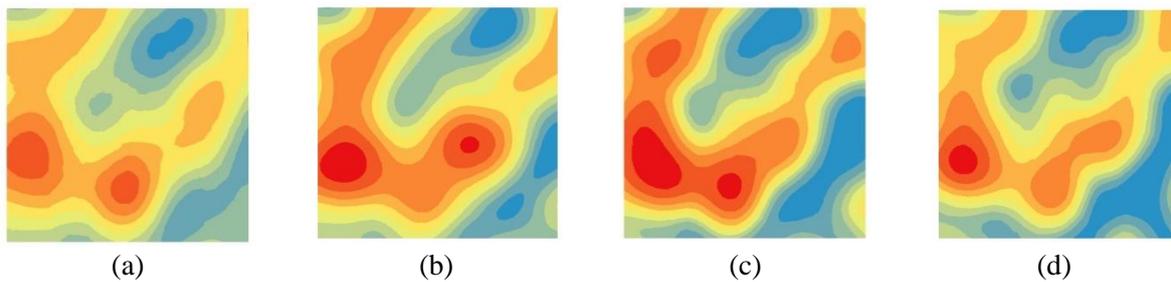
(a)	(b)	(c)	(d)

Figure 10 - Simple kriging of the altimetric survey in a section of "Zona da Mata", Minas Gerais State, Brazil. The sampling size are (a) 49 points, (b) 64 points, (c) 90 points, and (d) 110 points.
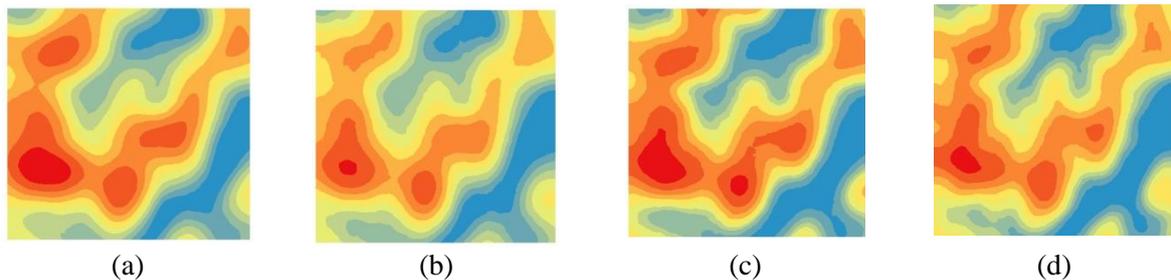

(a)	(b)	(c)	(d)

Figure 11- Simple kriging of the altimetric survey in a section of "Zona da Mata", Minas Gerais State, Brazil. The sampling size are (a) 156 points, (b) 224 points, (c) 342 points, and (d) 700 points.

As exhibited in Figures 10 and 11, the population map starts to be reasonably depicted by a sampling size of 342 points which survey accuracy achieved stabilization according to Figure 8.

Vieira (2000), Modis and Papaodysseus (2006), Santos et al. (2011), Ferreira, Santos, and Rodrigues (2013), and Yamamoto and Landim (2013) specified that in the variogram modelling, the estimated range should appraise data deviation, which in turn is assessed by the sample variation. Even with a simple reduction of sampling, the same happened with this data set.

Another significant stage of a Geostatistic analysis is the process of cross-validation. Among the essential steps in this process are the average and the residuals produced by observed and predicted values. As stated by Vieira (2000), Santos et al. (2011), Ferreira, Santos, and Rodrigues (2013), we expected the average of residuals obtained in this process is null, and the variance is unitary as shown by Mood, Graybill, and Boes (1974). For truth, the closeness of those values is the focus of the analysis.

4. Final remarks

The present work aimed to use the Nyquist Rate Theorem to determine an ideal size for georeferenced samples using a regular quadratic grid. The important theoretical part for the Gaussian model of spatial dependence fitted to the data was developed and, as a result, a minimum density was presented as a function of an estimated practical range, one of the parameters of the experimental variogram adjusted by this model. In practical terms, this density is about 30% of the estimated value for the parameter, under regularity conditions. The developed theory was applied to a large data set and the theoretical density was proven by practice, adopting as criteria of evaluation some procedures and measures already consolidated in the area of Classical Geostatistics.

5. Acknowledgments

6. References

ABRAMOWITZ, M.; STEGUN, I. A. **Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.** Washington, D.C.: U.S. Government Printing Office, 1972, 1046 p.

ARMSTRONG, M. **Basic Linear Geostatistics.** Springer, Berlin, 153 p. 1998.

BROOKER, P. I. **A geostatistical primer.** World Scientific, 1991. 95 p.

BRUS, D.J.; HEUVELINK, G.B.M., 2007. Optimization of sample patterns for universal kriging of environmental variables. **Geoderma**, n. 138, p. 86-95, 2007.

CLARK, I. **Practical geostatistics.** London: Applied Science Publishers, 1979. 130p.

CLARK, I. SAIMM Conference, Fourth World Conference on Sampling & Blending, 21-23 October 2009. **The Journal of The Southern African Institute of Mining and Metallurgy**, v. 110, p. 307-312, 2010.

CLARK, I.; HARPER, W. V. **Practical geostatistics.** Geostokos (Ecosse) Limited, 2000. 416 p.

CRESSIE N. A. C.; Wikle, C. K. **Statistics for spatio-temporal data.** Wiley: New Jersey, 2011, 588 p.

DIGGLE, P. J.; MENEZES, R.; SU, T. Geostatistical inference under preferential sampling. **Journal of the Royal Statistical Society: Series C (Applied Statistics)**, v. 59, n. 2, p. 191-232, 2010.

DRUCK, S.; CARVALHO, M. S.; CÂMARA, G.; MONTEIRO, A. M. V. **Análise Espacial de Dados Geográficos**. Brasília, EMBRAPA, 2004. 209 p.

ESRI. **ArcGIS 10.2.2 for Desktop.** ESRI: Redlands, USA, 2014.

FERREIRA, D. F. **Estatística básica**. 2 ed. Lavras: Editora UFLA, 2009. 664 p.

FERREIRA, I. O.; SANTOS, G. R.; RODRIGUES, D. D. Estudo sobre a utilização adequada da krigagem na representação computacional de superfícies batimétricas. **Revista Brasileira de Cartografia**. n. 65/5, p. 831-842, 2013.

JOURNEL, A. G., HUIJBREGTS, C. J. **Mining Geostatistics**. Academic Press, London, 1978, 600 p.

MODIS, K, PAPAODYSSEUS, K. Theoretical estimation of the critical sampling size for homogeneous ore bodies with small nugget effect. **Math. Geol.**, v. 38, n. 8, p. 489-501, 2006.

MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. **Introduction to the Theory of Statistics.** McGraw Hill. 1974, 480 p.

OLEA, R. A. **Geostatistics for engineers and earth scientists.** Kluwer Academic Publishers, London, 1999, 303 p.

OLEA, R. A. **A practical primer on geostatistics.** Open-File Report, 2009-1103. U.S. Geological Survey, 2009. 346 p.

OLIVEIRA, M. S. de; BEARZOTI, E.; VILLAS BOAS, F. L.; NOGUEIRA, D. A.; NICOLAU, L. A.; OLIVEIRA, H. S. S. de. **Introdução à Estatística**, 2 ed. Lavras: UFLA, 2014. 462 p.

PEIGNÉ, J. et al. Soil Sampling based on field spatial variability of soil microbials indicators. **European Journal of soil Variability**, v.45, p.488-495, 2009.

PINTO, E. S. O.; SANTOS, G. R.; OLIVEIRA, F. L. P. Análise Espaço-Temporal Aplicada às Ocorrências de Hipertensão e Diabetes nos Municípios do Estado de Minas Gerais. **Revista Brasileira de Biometria**, v. 32, p. 238-266, 2014.

SANTOS, G. R.; OLIVEIRA, M. S.; LOUZADA, J. M.; SANTOS, A. M. R. T. Krigagem simples versus krigagem universal: qual o preditor mais preciso? **Revista Energia na Agricultura**, Botucatu, v. 26, n. 2, p. 49-55, 2011.

SANTOS, P. C. et al. O emprego da geoestatística na determinação do tamanho "ótimo" de amostras aleatórias com vistas à obtenção de estimativas dos volumes dos fustes de espécies florestais em Paragominas, estado do Pará. **Acta Amazonica**, v. 41, n. 2, p.213-222, 2011.

SARTORI, A. A. C.; ZIMBACK, C. R. L. Recomposição florestal visando à conservação de recursos hídricos na bacia do Rio Pardo, São Paulo. **Revista Energia na Agricultura**, v. 26, n.4, p. 43-53, 2011.

SHANNON, C. E. **The Mathematical Theory of Communication**. Univof Illinois Press, 1949.

SOUZA, Z. M.; SOUZA, G. S.; JÚNIOR, J. M.; PEREIRA, G. T. Número de amostras na análise geoestatística e na krigagem de mapas de atributos do solo. **Ciência Rural**, v. 44, n. 2, p. 261-268, 2014.

VÁSAT. R.; HEUVELINK, G. B. M.; BORUVKA, L. Sampling design optimization for multivariate soil mapping. **Geoderma**, v. 155, n. 3-4, pp. 147–153, 2010.

VER HOEF, J. Sampling and geostatistics for spatial data. **Écoscience**, v. 9, p. 152-161, 2002.

VIEIRA, S. R. Geoestatística em estudos de variabilidade espacial do solo. In: NOVAIS, R.F.; ALVAREZ, V.H.; SCHAEFER, G.R. (Ed.) **Tópicos em ciência do solo**. Viçosa: SBCS, 2000. v.1, p.1-54.

WEBSTER, R.; OLIVER, M. A. **Geostatistics for environmental scientists.** 2. ed. Wiley, 2007. 315p.

WEBSTER, R.; OLIVER, M. A. Sample adequate to estimate variograms of soil properties. **Journal of Soil Science**, v. 43, p. 177-192, 1992.

YAMAMOTO, J. K.; LANDIM, P. M. B. **Conceitos e Aplicações.** São Paulo: Oficina de Textos, 2013. 216 p.

YFANTIS, E. A.; FLATMAN, G. T.; BEHAR, J. V. Efficiency of kriging estimation for square, triangular and hezagonal grids. **Mathematical Geology**, v. 19, p. 183-205, 1987.

WHITTAKER, J. M. On the Functions Wich are Represented by the Expansions of the Interpolation Theory. **Proc. Royal Soc. Edinburgh**, Sec. A, vol. 35, pp. 181-194, 1915.