Revista GEOCIÊNCIAS

unesp

ISSN: 1980-900X (online)

ESTIMATING THE SAMPLING ERROR FROM THE COVARIOGRAM OF SPATIALLY CORRELATED DATA

ESTIMANDO ERRO AMOSTRAL A PARTIR DE COVARIOGRAMAS DE DADOS ESPACIALMENTE CORRELACIONADOS

Victor MIGUEL-SILVA[1], Daniele Costa de MESQUITA[2]

[1]Gerência de Padronização e Integração (COI Planejamento), VALE - Av. Dr. Marco Paulo Simon Jardim, 3580 - Vila da Serra, Nova Lima – MG. E-mail: victormiguelsilva@hotmail.com

[2]Programa de Pós-graduação em Geografia - Tratamento da Informação Espacial - PUC Minas. E-mail: dani.mesquit@gmail.com

ABSTRACT - The total sampling-error attached to a set of samples has a central role in the selection of the statistical method to extract information from this noisy data. However, commonly direct measurements of the sampling error are not available and then, the magnitude of the error is unknown. In this article, we present a mathematically sound solution for estimating the sampling error directly from spatially correlated observations. The method is based on the difference between the global variance and the inferred y-axis intercept of the covariogram computed from the same data. We developed the mathematical proofs of the method, and its performance is analyzed by applying it to five variables from a stream-sediments dataset of a multi-element geochemical survey. The estimated total sampling error is satisfactory close to the value experimentally measured by field replicates.
Keywords: Linear Regression. Covariogram. Sampling error.

RESUMO - O erro amostral total associado com um conjunto de amostras tem um papel central na seleção do método estatístico a ser utilizado para extrair informação desses dados com ruído. No entanto, é comum que a magnitude desses erros amostrais seja desconhecida porque medições diretas desse erro não estão disponíveis. No presente artigo apresentamos um método sólido do ponto de vista matemático para estimar o erro amostral diretamente de observações especialmente correlacionadas. O método se baseia na diferença entre a variância global e a interseção do eixo-y do covariograma calculado para os mesmos dados. As provas matemáticas são desenvolvidas e sua performance é analisada ao se aplicar o método para cinco variáveis de amostras de sedimento corrente de uma campanha geoquímica multivariada. A proximidade entre o erro inferido e o valor experimentalmente medido por replicatas de campo foi satisfatório.
Palavras-chave: Regressão Linear. Covariograma. Erro amostral.

INTRODUCTION

Geochemical surveys are an important part of geoscientific investigations in both mineral exploration and environmental monitoring. The quality of the decision-making is proportional to the quality of the available data. Thus, geochemical surveys have been adopting quality assurance and control programs (QA/QC) in over the last decades to measure and minimize the total-sampling errors. Understanding of the resulting total-sampling error associated with available datasets is fundamental to correctly manage them. In real-world datasets, however, commonly quality control data is not available and then, usually the sampling error is incorrectly assumed as null (Gy, 1982). Commonly consequences of assuming the data as free of errors are overlooked, such as slope bias in linear regression (Deming, 1943) or incorrect estimates performed by Kriging (Journel & Huijbreigts, 1979; Isaaks & Srivastava, 1989).

This article presents a mathematically sound approach for estimating the total-sampling error associated to spatially correlated data. It is based on the property that the underlying error-free process variance sums up with the total sampling-error variance, while the sampling

errors do not affect the covariogram for any distance **h** > 0. Therefore, we infer the total sampling-error variance as the difference of the covariogram y-axis intercept minus the measured global variogram.

We organize this paper as follows: in Methods the sampling error model and the variance and covariance properties which the proposed approach relies on. The covariogram theory and its application discussed. We analyze the method performance by applying it to a dataset composed of 752 stream-sediment from the Bauru basin (São Paulo State – Brazil). The proposed-approached estimates are compared to error experimentally measured by field replicates. The comparison results show the proposed approach as plausible when quality-control data is not available. The Discussion section illustrates the risks of assuming the available data as free of error without further analysis.

METHODS

Estimating the sampling error

The sampling errors ε_i are intrinsically associated with all (Gy, 1982). We cannot completely cancel the error ε_i, but it can be minimized through proper quality assurance and quality control program (QA/QC). The relationship between observations $z_i(\mathbf{u})$ of a given variable i, i=1,..,I and its underlying true value is $T_i(\mathbf{u})$ is given by $z_i(\mathbf{u}) = T_i(\mathbf{u}) + \varepsilon_i$, where ε_i is assumed as unbiased $E\{\varepsilon_i\} = 0$, uncorrelated with the geological process value $Cov\{\varepsilon_i, t_i(\mathbf{u})\} = 0$, and has a variance $Var\{\varepsilon_i\} = \sigma^2_{i,error}$.

Statistical Background - The sampling srror influence on variance and covariance

The proposed approach relies on the properties that the global variance $Var\{z_i(\mathbf{u})\}$ is derived from the additivity between $Var\{T_i(\mathbf{u})\}$ and $\sigma^2_{i,error}$ (Equation 1).

$$Var\{T_i(\mathbf{u}) + \varepsilon_i\} = Var\{T_i(\mathbf{u})\} + \sigma^2_{i,error} = Var\{z_i(\mathbf{u})\} \quad (1)$$

The covariance between two points separated by a distance **h** > 0 is not affected by ε_i when $Cov\{\varepsilon_i, t_i(\mathbf{u})\} = 0$. It is due to the following bi-linearity property (Equation 2).

$$Cov\{z_i(\mathbf{u}), z_i(\mathbf{u}+\mathbf{h})\} = Cov\{T_i(\mathbf{u})+\varepsilon_i, T_i(\mathbf{u}+\mathbf{h}) + \varepsilon_i\} = Cov\{T_i(\mathbf{u}), T_i(\mathbf{u}+\mathbf{h})\} + Cov\{T_i(\mathbf{u}), \varepsilon_i\} + Cov\{\varepsilon_i, T_i(\mathbf{u}+\mathbf{h})\} + Cov\{\varepsilon_i, \varepsilon_i\} = Cov\{z_i(\mathbf{u}), z_i(\mathbf{u}+\mathbf{h})\} = Cov\{z_i(\mathbf{u}), z_i(\mathbf{u}+\mathbf{h})\} = Cov\{T_i(\mathbf{u}), T_i(\mathbf{u}+\mathbf{h})\} \quad (2)$$

Therefore, $\lim_{\boldsymbol{h}->0} Cov\{\mathrm{zi}(\mathbf{u}), \mathrm{zi}(\mathbf{u}+\mathbf{h})\}$, tends to the error-free variance when their separation distance **h** tends to zero. When **h** = 0, we have $Cov\{z_i(\mathbf{u}), z_i(\mathbf{u}+\mathbf{h})\} = Var\{z_i(\mathbf{u})\}$.

In words, Equation (1) shows that the deviation between the measured and the error-free variance is due to the total sampling-error variance, while Equation 2 shows that the total sampling-error does not affect the measured covariance.

Therefore, the covariance between two observations tends to their error-free variance value when their separation distance **h** tends to zero. It is valid only when the points are not exactly coincident (**h** = 0), because the covariance of value and itself is its variance and thus, affected by the total sampling-error variance. Considering the presented properties, we get:

$$Var\{\mathrm{Ti}(\mathbf{u})\} \cong \lim_{\boldsymbol{h}->0} Cov\{\mathrm{Ti}(\mathbf{u}), \mathrm{Ti}(\mathbf{u}+\mathbf{h})\}$$

Rearranging the equation above and considering Equation (1) and (2) equivalences:

$$Var\{\mathrm{Zi}(\mathbf{u})\} - \sigma^2\mathrm{i,error} \cong \lim_{\boldsymbol{h}->0} Cov\{\mathrm{Zi}(\mathbf{u}), \mathrm{Zi}(\mathbf{u}+\mathbf{h})\} \quad \sigma^2\mathrm{i,error} \cong Var\{\mathrm{Ti}(\mathbf{u})\} - \lim_{\boldsymbol{h}->0} Cov\{\mathrm{Zi}(\mathbf{u}), \mathrm{Zi}(\mathbf{u}+\mathbf{h})\} \quad (3)$$

Equation 3 is a relevant estimator of the total sampling error when quality-control data is not available. We present in the next section how to use an experimental covariogram to estimate the covariance when **h** tends to zero, hereafter referred to as $\lim_{\boldsymbol{h}->0} Cov\{\mathbf{h}\}$.

Modelling the Experimental Covariogram

The measure of spatial dependence, such as the variogram, covariogram, correlogram, among others, is the key to understand how the spatial correlation of the variable of interest behaves. For details, Isaaks & Srivastava (1989) describe methods for evaluating the spatial continuity of data. Among these, the covariogram measures the average similarity between two attribute values approximately separated by vector **h**:

$$Cov\{\mathbf{h}\} = \frac{1}{N(\mathbf{h})} \sum_{N(\mathbf{h})} \mathrm{Zi}(\mathbf{u})\mathrm{Zi}(\mathbf{u}+\mathbf{h}) - m^2$$

Where N(**h**) is the number of pairs at a distance **h** from each other and *m* is the mean of the values Zi(**u**). Figure 1 illustrates this naming convention

relative to the separation vector **h** in the isotropic case. In practice, we computed **h** with some lag tolerance because the number N(**h**) of pairs separated exactly at a distance **h** may be very small in a non-regularly spaced sampling grid. A general rule of thumb is to use the lag tolerance as half of the lag.

The proposed approach requires defining the nugget effect from the experimental covariogram. For sake of simplicity, we recommend using an isotropic variogram as the nugget effect is an isotropic property. Moreover, the isotropic model maximizes the number of pairs and increases the model stability.

The first 2 or 4 experimental points of the experimental covariogram are considered to define the nugget-effect fitting may be performed by visual analysis of, being these the most instructive to extrapolating the behaviour to the y-axis intercept. The value from $\lim_{\boldsymbol{h}->0} Cov\{\boldsymbol{h}\}$ makes it possible to estimate $\sigma^2_{i,error}$ through equation 3.

Two interesting properties of the proposed approach are its robustness to outliers and to the presence of mixed populations related to different geological backgrounds:

(i) The first occurs because outliers affect both variance and covariances and the gap between them stills roughly constant.

(ii) The second property is that the influence of two or more mixed subpopulations tends to be small. It is because the nugget effect is commonly estimated using the initial lags and these are, in general, slightly affected by this mixture. The resulting nugget effect is an average statistic weighted by the number of pairs of each subpopulation in the initial lags.

Next, we apply the proposed approach to a real-world dataset and compare its estimated error to the total error measured from field-replicate samples.

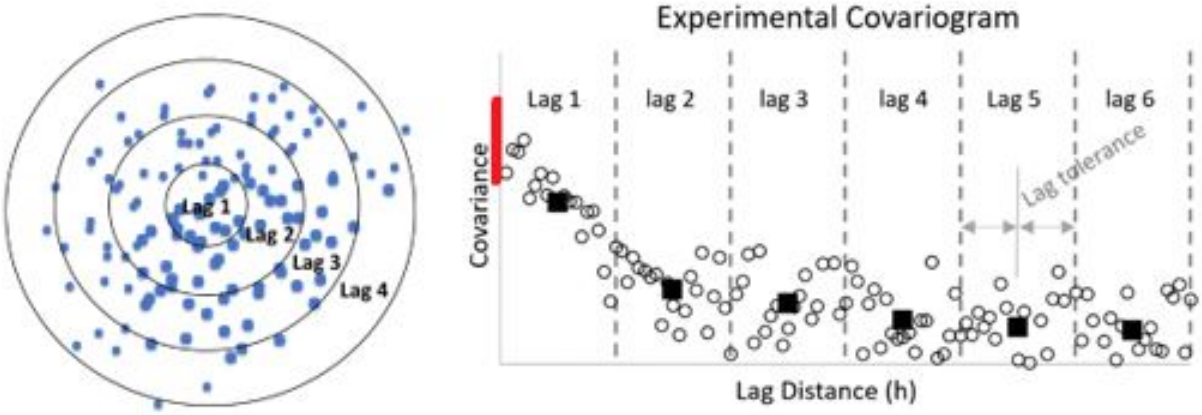

Figure 1 - A) the spatial distribution of a set of samples inside each lag (black circles), B) the covariance between each pair of samples within lags (empty circles) and the average covariance of each lag in an experimental covariogram (black squares). The red line indicates the range of probable positions where the experimental data intercepts the y-axis.

DATA PRESENTATION

The proposed approach is applied to a set of stream-sediment samples from a multi-element geochemical survey entitled "low-density geochemical mapping in Brazil". It was carried out in the São Paulo State by the Geological Service of Brazil (CPRM) between 2010-2013. Each sample was collected to represent a basin area between 100 and 200 km². Field replicates were collected with a rate of 10%. More details available in Mapa (2015).

A brief geological overview

The State of São Paulo is geologically inserted in the South American platform, being composed of rocks from Archean to the Holocene (Bistrichi et al., 1981). The State's regional geology is divided into platform basement (AAC), which is mainly composed of crystalline pre-Cambrian metamorphic and igneous rocks that crop out in the east (Figure 2). Above AAC is the Paleo-Mesozoic Parana basin, composed of sedimentary, volcanic and subvolcanic rocks. The present paper analyzes the stream-sediments samples were collected on the hydrographic network of the Bauru and Caiuá Groups (BBA) of the Bauru Basin (Milani et al., 2007).

The Bauru Group crop out on six Brazilian states (São Paulo, Paraná, Mato Grosso, Mato Grosso do Sul, Minas Gerais and Goiás), and parts of Paraguay (Caiuá Formation). The Bauru Group is the Lower-Upper Cretaceous sedimentary sequence of the Paraná Basin. It predominantly comprises eolian sandstones in its Lower Creta-

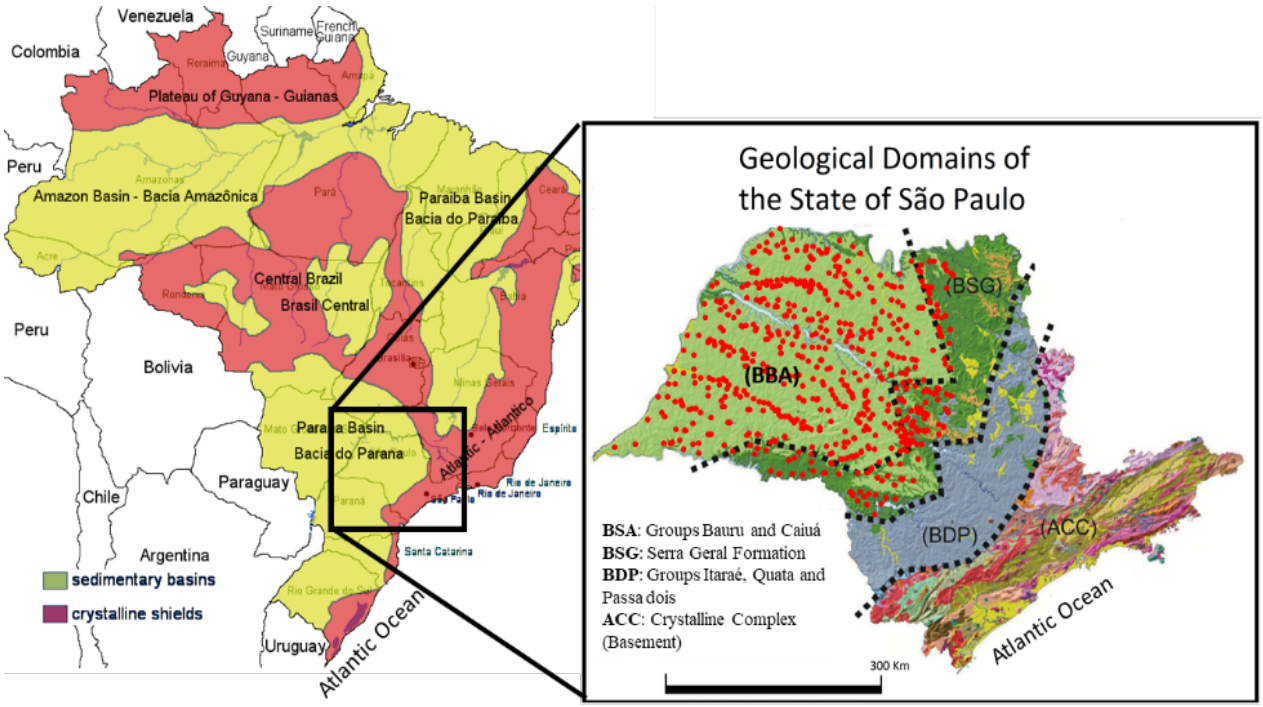

Figure 2 – Schematic map of the Brazil main sedimentary basin and crystalline shields. The left-hand side figure zooms in the State of São Paulo area and its main geological domains, the red circles show places which stream-sediment samples were taken (Altered from Mapa, 2015).

ceous sections and alluvial to fluvial conglomerates, sandstones, siltstones, and mudstones with subordinate lacustrine mudstones in its Upper Cretaceous sections. This package of rocks was deposited in a subsiding sedimentary basin, which developed in the central-southern South American Platform as a result of thermo-mechanical subsidence following the opening of the Atlantic Ocean between South America and Africa (Fernandes & Ribeiro, 2015).

The geochemical database – origin and exploratory analysis

As presented by Mapa (2015), the dataset is composed of 1422 stream-sediment samples. The team collected the samples in active drainage channels, preferentially at linear stretches or spots with moderate turbidity with deposition of fine particles. For better homogenization, each sample was composited of subsamples collected at different spots along approximately 100 meters upstream of the drainage access. Each sample was collected with a plastic cup and the material retained on a 1-mm mesh discarded. The remaining fine material was packed and identified. The sampled material was prepared and analyzed by Brazilian commercial laboratory SGS-GEOSOL. These samples were oven-dried at 50°C, sieved to <80#, sub-sampled and sieved again to <150#. The resulting pulp was digested by aqua regia and then analyzed by ICP-MS and ICP-AES (Inductively Coupled Plasma - Mass Spectrometry and Atomic Emission Spectrometry, respectively). Among the 32 available elements, the proposed approach is applied to five elements (Al, Co, Cu, Mn, and Zn). For sake of brevity, the study was carried out considering only five variables. However, the proposed approach is possible to be applied to any of the 32 variables with spatial correlation. Table 1 summarizes the statistics of the 752 samples from the Bauru basin area (Figure 2).

Table 1 – Summary table of stream-sediment samples.

	Al	Co	Cu	Mn	Zn
N. Samples	752	752	752	752	752
Minimum	3	0.2	0.6	0.16	0.5
1º quartile	16	1.9	3.8	1.84	4.0
Median	32	3.9	8.2	3.29	9.0
Mean	62.7	7.97	22.19	5.18	22.48
3º quartile	78	10.07	22.77	6.32	24.0
Maximum	581	93.1	277	100	260
Variance	6006	104	1267	38.7	1081
CV	2.4	2.61	4.34	1.88	3.65

Figure 3 shows the boxplot of the six variables under analysis. We may observe positively skewed distributions and a large distance between mean and median, where mean larger than the median is explained by the skewness and the presence of outliers.

Next, we present the available field replicates and their statistics are compared to table 1 values.

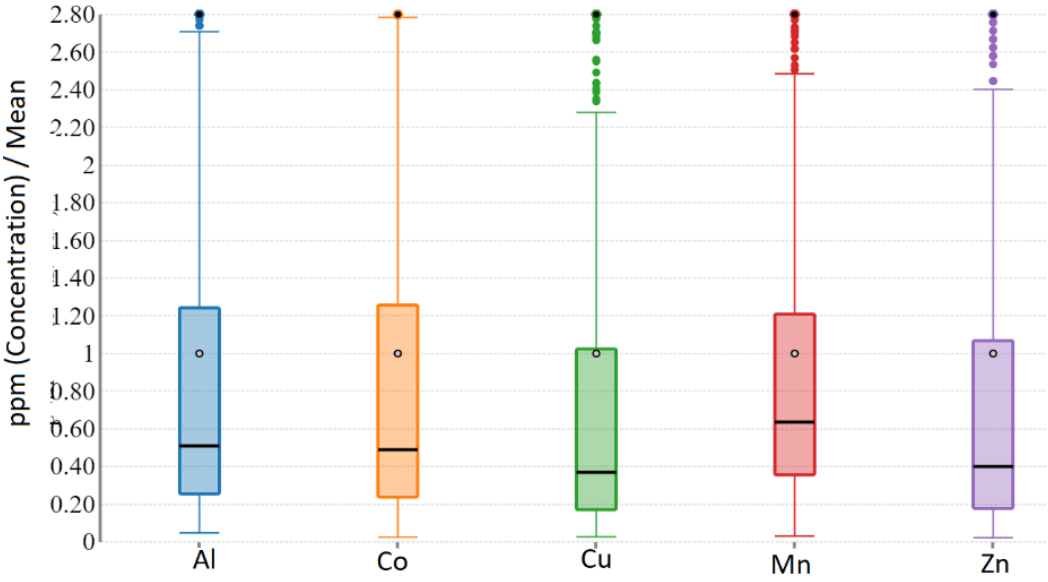

Figure 3 – Boxplot of the variables under study measured by stream-sediment samples.

The geochemical database – field-replicate samples

Replicates samples are defined by IUPAC (1997) as "*Multiple (or two) samples taken under comparable conditions. This selection may be accomplished by taking units adjacent in time or space. Although the replicate samples are expected to be identical, often the only thing replicated is the act of taking the physical sample*". The Field Replicates (FR) used in this study were independently collected as close as possible of each other in space and time. Both samples are prepared and analyzed using the same procedures. The variance between FR allows to measure the errors that arise from sample acquisition in the field, which may contribute with over 80% of the total sampling error. The uncertainty may be measured from field replicates using different equations and metrics (Abzalov, 2011).

In this article we compute the variances of the measurement errors $\sigma^2_{i,error}$ from the difference between the variance of each observation set Z_i and their covariance:

$$\sigma^2_{i,error} = 0.5\{\text{Var}\, Z_i(x) + \text{Var}\, Z_2(x)\} - \text{Cov}\,\{Z_1(x), Z_2(x)\} \quad (4)$$

In words, we compute the total sampling error from field replicates as the difference between their average variance of each pair minus their covariance. Table 2 shows the statistics of the matched pair of FR from the Bauru Basin area.

Table 2 – Summary table of field replicates samples.

	Al	Co	Cu	Mn	Zn
N. Pairs	60	62	62	62	58
Minimum	5.0	0.5	0.7	0.5	1.0
1º Quartile	15.50	2.10	4.00	2.13	3.00
Median	27	4.3	8.9	3.3	9.0
Mean	54.3	7.7	24.9	5.3	24.2
3º Quartile	70	10	21	7	17
Maximum	241	44.4	277	23	260
Variance Orig.	3147	71	1915	21	1787
Variance Repl.	3000	51	1808	18	1504
Covariance	2561	52.1	1753.4	16.3	1579
Total Error	512	8.96	108.2	3.2	66.4
Total Error (%)	16%	12.6%	5.6%	15.5%	3.7%

There are 62 pairs of field replicates for each available variable. The exception is the Al and Zn which some values were reported as “less than the limit of detection”. This is usually reported when an analytical procedure detects an element, but the value is too low to be accurately quantified.

The distribution and statistics of original (Table 1) and control samples (Table 2) of the variables under study are similar and that FR samples are representative of the original data. Figure 3 shows the scatter-plot between each matched pair of values and their statistics.

Next session we move forward to illustrate the proposed-approach, where the relationship between global data variance and the nugget effect of the covariogram are used to estimate the sampling error associated with a given spatially correlated variables.

Estimating the sampling error from the variance – covariance relationship

The experimental covariogram was fitted to the observations of each element under study using a lag separation distance of 14 Km, a lag tolerance and bandwidth of 7 Km, and an angular tolerance of 22.5º (Figure 4). The nugget effect is the only parameter of the covariogram model that is necessary to be defined. It was estimated by extrapolating the trend of the first 2-4 experimental points to the y-axis intercept. The linear extrapolation (red line), the range of values where the y-axis intercept probably occurs (orange bar) and the actual covariance measured from field replicates (value highlighted by the green box).

Next, we compare the proposed approach estimates and experimental total error values results. These values are compared in Table 3 with experimental values measured by FR samples using equation 4. Figure 5 presents graphically the data summarized in table 3.

The deviation between the experimental and the estimated error shows that the proposed approach is a reliable solution when direct measurements of the sampling error are not available.

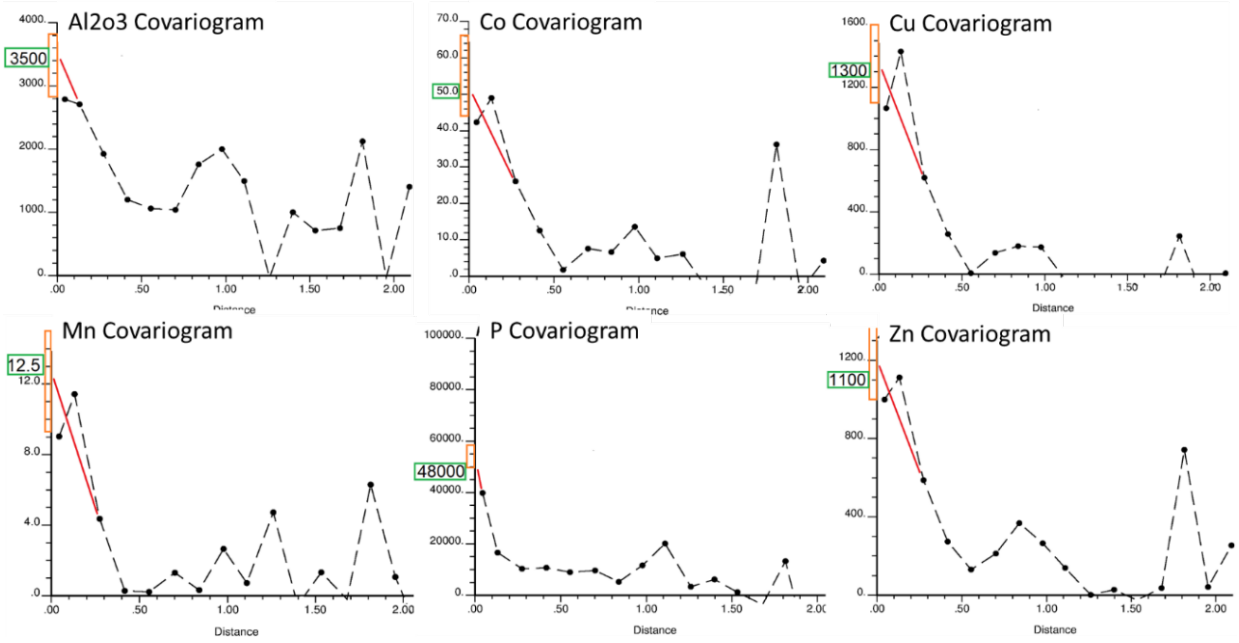

Figure 4 – Experimental covariograms. The linear extrapolation (Red line), the range of values where is plausible to define the y-axis intercept (orange bar) and the actual covariance measured from field replicates (green-box value). Distances measured in fractions of 100 Km.

Table 3 – Comparison between the total sampling relative to the total data variance, estimated from the proposed approach and experimentally from field-replicate samples.

	Proposed Approach	**RF Samples**
Al	10.5%	6.3%
Co	13.0%	14.6%
Cu	2.0%	5.8%
Mn	16.9%	16.5%
P	11.5%	12.3%
Zn	1.8%	8.7%

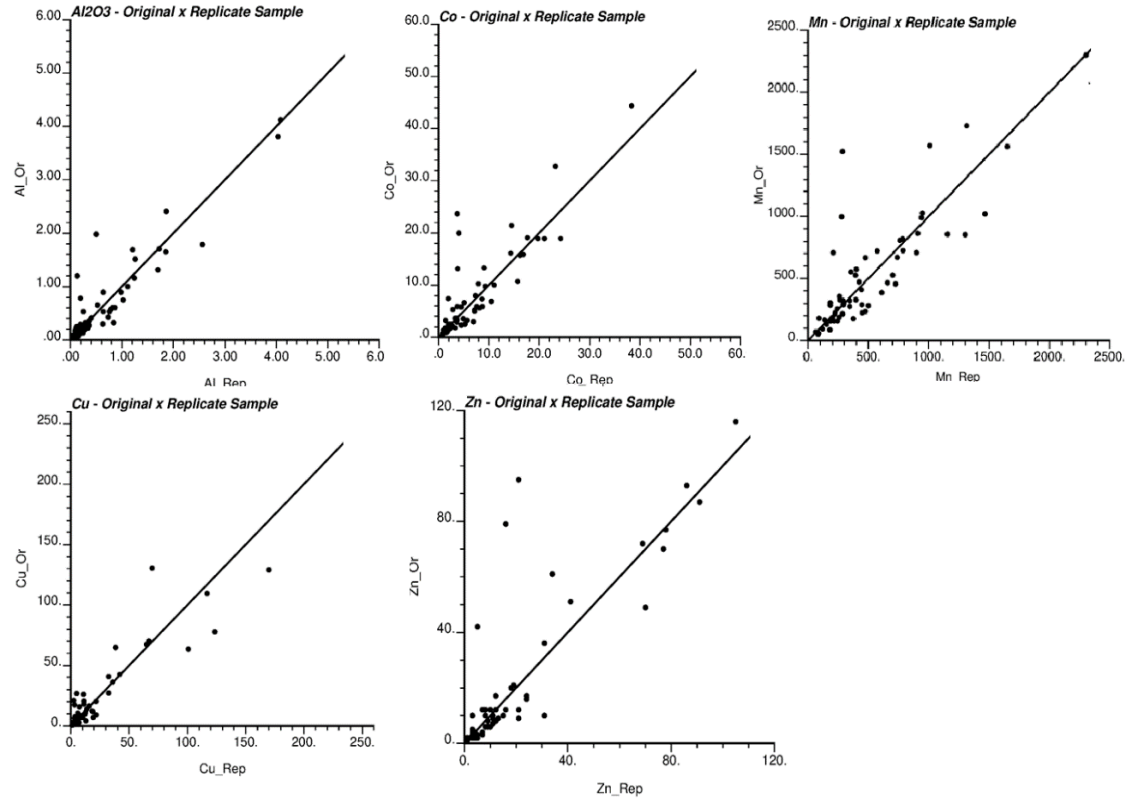

Figure 5 – Scatter plot between matched pairs of field replicates. In black, the y=x line. The red squares highlight the values classified as outliers.

RESULTS

We used the proposed approach to estimate the total sampling errors for each variable, considering the data variance (Table 1) and their inferred covariance (Figure 6) using the relationship $\sigma^2_{i,error} = Var\{z_i(\mathbf{u})\} - C(\mathbf{h})$ where **h** tends to zero.

It is relevant to point out that the number of available FR samples is small (64) and that no outliers were removed, or robust methods used for original and FR samples, avoiding the chance of biased procedures or decisions on data treatment that can affect results.

Next, we discuss how the measurement error impacts the results of many statistical methods widely used in geochemical survey data treatment and analysis

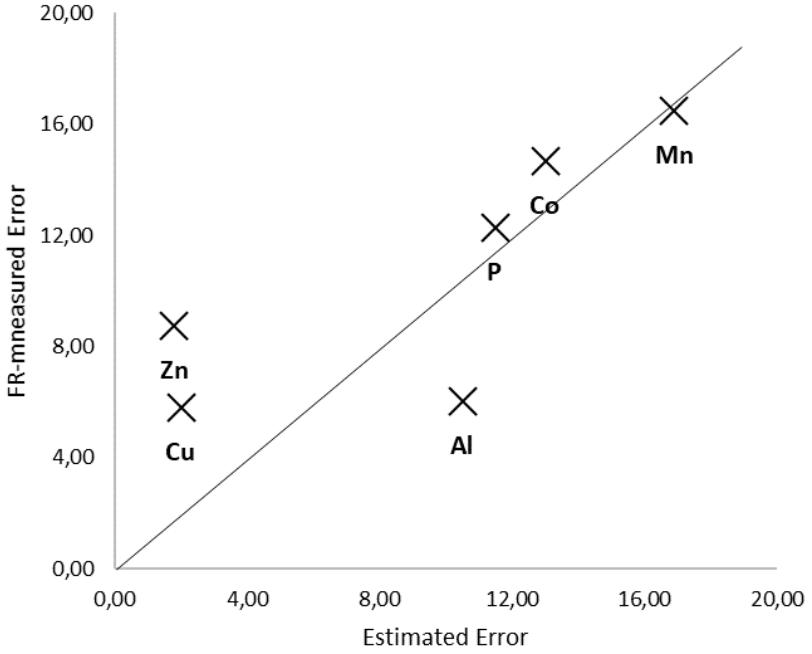

Figure 6 – Relationship between error estimated by the proposed approach (X-Axis) and the error experimentally measured from Field-replicate samples (Y-axis).

DISCUSSIONS

The correct estimate of the total-error variance of each variable (or set of variables) is of paramount importance in the methods and statistical tools selection. In this article we presented one mathematically sound method to estimate sampling errors when quality-control data are not available. In the presented case study, the novel approach results were satisfactorily close of the errors directly measured from field replicates. Now we move forward and discuss the importance of considering the presence of the sampling error in methods widely used in geosciences:

Kriging (Matheron, 1963): The variogram nugget-effect is composed of the total-sampling error and micro-scale components (Cressie, 1993, p. 59; Journel & Huijbreigts, 1978). The proportion between these two components determines the behavior of the kriging predictions at nodes spatially coincident and in regions with observations. However, the presence of sampling error is ignored, and the nugget effect is assumed to be fully composed of micro-scale variation. This decision constrains the kriging estimator to honor noisy observations.

Figure 7 shows the underlying process (black line) and some observations associated with "sampling errors" (the dots).

The solid colored lines are the kriging predictions and the dashed lines their kriging variance. The blue-green line was estimated considering a null nugget effect, while the red and blue lines were estimated considering the nugget effect fully composed of micro-scale or sampling error, respectively.

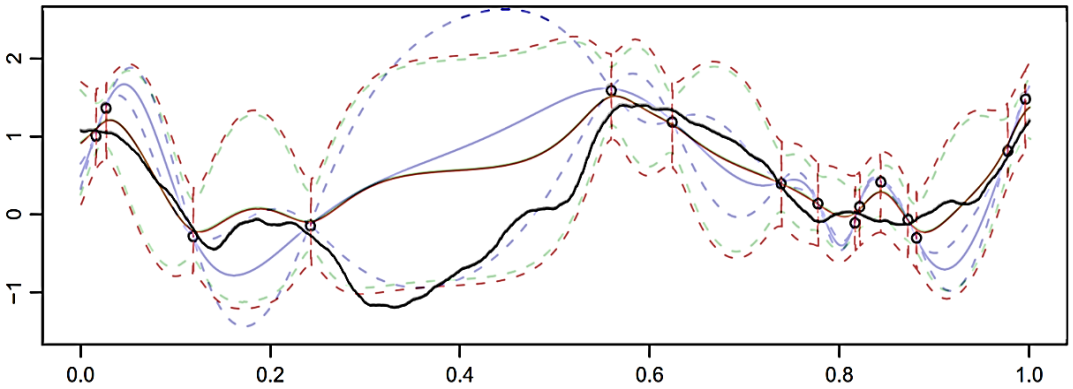

Figure 7 - Kriging predictions under three scenarios. In each scenario, the black line is the true underlying smooth process. Adapted from Paciorek (2008).

The proposed approach allows the practitioner to estimate the nugget-effect composition from the total-error component. Next, we discuss methods how considering or not the sampling error leads to different regression lines.

Linear-Regression models: A common step in the statistical data analysis of geological data is to find or show a relation between two or more variables. A method widely used is the least-squares fitting, which relies on assumption that those regressors (y-axis data) are free of errors. This method, however, may not be adequate for fitting straight lines when both variables are subject to error. Classic studies about linear regression models with measurement error may be found in Deming (1943), York (1966).

In figure 8 we compare the standard regression method to the Deming regression, which accounts for the ratio of the variance errors attached to observations on the x- and the y-axis. Under the assumption of linear association between these variables, we apply both methods for defining the relationship between Al and Co, Cu, Mn or Zn values.

Figure 8 shows that the decision of considering or not the error attached to each variable is relevant and may lead to different results.

The point of this discussion it not about the type of relationship between the variables, but that the decision of considering or not the error associated to variables must be carefully analyzed for each case and it is valid to any type of regression and equation fitting. We may extend the same discussions about the total sampling error to Principal Component Analysis, hypothesis testing, among other statistical methods used in geoscientific problems.

We remind the reader that the good results were achieved using the covariogram in its original formulation and raw data without any treatment, outlier removal, or parameters individually adjusted for each variable. However, we highly recommend in other real-world problems

to test statistical methods adequate for specific geological or statistical conditions, such as robust estimators or covariograms adequate for strongly skewed statistical distributions when necessary.

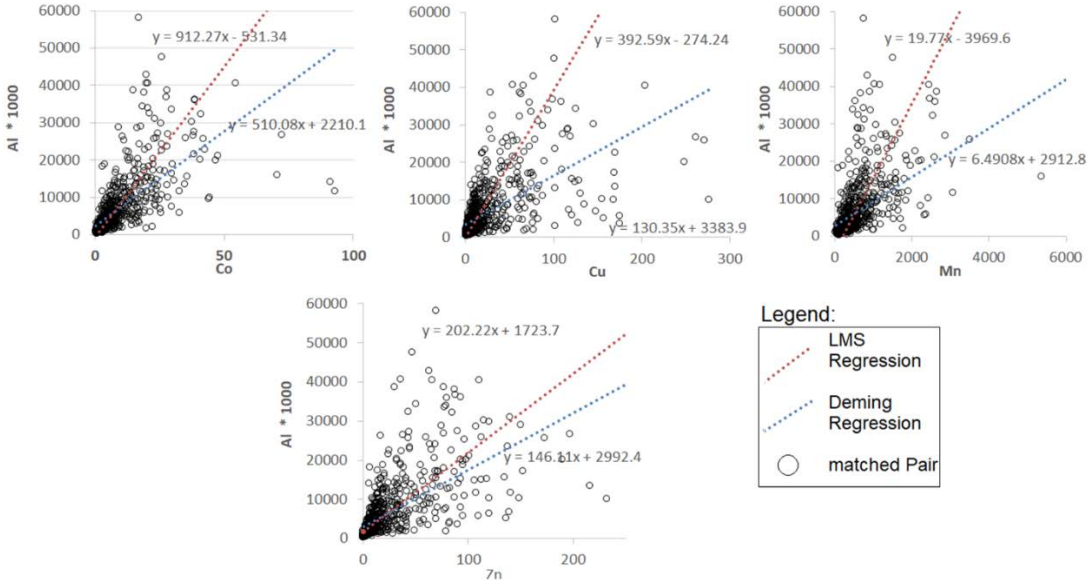

Figure 8 - Scatterplot between Al concentration against Co, Cu, Mn and Zn values. The red and blue dotted lines represent the regression equation fitted by Deming and Least-minimum squared methods, respectively.

CONCLUSIONS

In this study, we presented a novel approach to estimate the total sampling error associate with a dataset. It is a helpful alternative if experimental measurements are not available. The proposed approach is theoretically sound for manage spatially correlated observations. The mathematical proofs were corroborated by the presented case study, where the total-sampling error for five variables from a stream-sediment dataset estimates by the proposed approach and measured by field-replicates samples showed similar results.

ACKNOWLEDGEMENTS

The authors acknowledge Felipe Brito Mapa of the Serviço Geológico do Brasil (Brazil Geological Survey) SGB/CPRM for providing his dissertation raw data.

REFERENCES

ABZALOV, M. **Sampling errors and control of assay data quality in exploration and mining geology**. IntechOpen, 2011.

BISTRICHI, C.A.; CARNEIRO, C.D.R.; DANTAS, A.S.L.; PONÇANO, W.L.; CAMPANHA, G.A.C.; NAGATA, N.; ALMEIDA, M.A., STEIN, D.P.; MELO, M.S.; CREMONINI, O.A.; HASUI, Y.; ALMEIDA, F.F.M. 1981. **Mapa Geológico do Estado de São Paulo.** São Paulo, IPT (Monografias, 6, anexo).

CRESSIE, N. 1993, **Statistics for spatial data.** Wiley, New York, 928 p.

DEMING, W. E. 1943. **Statistical adjustment of data.** Wiley, New York, 269 p.

FERNANDES, L.A. & RIBEIRO C.M.M. 2015. Evolution and palaeoenvironment of the Bauru Basin (Upper Cretaceous, Brazil). **Journal of South American Earth Sciences,** v. 61, p. 71-90.

GY, P. 1982. **Sampling of particulate materials: theory and practice**, 2nd ed. Amsterdam: Elvesier. 431 pp.

ISAAKS, E.H. & SRIVASTAVA, R.M. **An introduction to applied geostatistics**: Oxford University Press, New York, 561 p., 1989.

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

IUPAC. **Compendium of Chemical Terminology**, 2nd ed. (the "Gold Book"). Compiled by A. D. McNaught and A. Wilkinson. Blackwell Scientific Publications, Oxford (1997). Online version (2019-) created by S. J. Chalk. 2019.

MAPA, F.B. **Geoquímica multielementar de sedimentos de corrente no Estado de São Paulo: abordagem através da análise estatística multivariada.** São Paulo. 2016. Master's Dissertation, Instituto de Geociências, University of São Paulo,

MATHERON, G. **Principles of geostatistics.** Econ. Geol., v. 58, p. 1246–1266, 1963.

MILANI, E.J.; FRANÇA, A.B. & MEDEIROS, R.A. **Rochas geradoras e rochas-reservatório da Bacia do Paraná, faixa oriental de afloramentos**, Estado do Paraná. Boletim de Geociências da Petrobrás, Rio de Janeiro, v. 15, n. 1, p.135-162, 2007.

PACIOREK, C. Technical vignette 3: Kriging, interpolation. and uncertainty. Tech. rep., **Harvard School of Public Health**, 2008.

YORK, D. **Least-Squares Fitting of a Straight Line**. Canadian Journal of Physics. v. 44, p.1079-1086. 2001.

Submetido em 24 de junho de 2021
Aceito para publicação em 26 de outubro de 2021