UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE GEOCIÊNCIAS

Hybrid objective function applied to optimize infill sampling location

Gustavo Zanco Ramos

Orientador: Prof. Dr. Marcelo Monteiro da Rocha

TESE DE DOUTORADO
Programa de Pós-Graduação em Geociências: Recursos Minerais e Hidrogeologia

SÃO PAULO
2022

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE GEOCIÊNCIAS

Hybrid objective function applied to optimize infill sampling location

Gustavo Zanco Ramos

Tese apresentada ao Programa Geociências (Recursos Minerais e Hidrogeologia) para obtenção do título de Doutor em Ciências

Área de concentração: Recursos Minerais e Meio Ambiente

Orientador: Prof. Dr. Marcelo Monteiro da Rocha

SÃO PAULO

2022

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE GEOCIÊNCIAS

Hybrid objective function applied to optimize infill sampling location

Gustavo Zanco Ramos

Orientador: Prof. Dr. Marcelo Monteiro da Rocha

Tese de Doutorado

Nº 649

COMISSÃO JULGADORA

Dr. Marcelo Monteiro da Rocha

Dr. Giorgio Francesco Cesare de Tomi

Dra. Rita Parisi Conde Pozzi

Dr. José Alberto Quintanilha

Dr. Cleyton de Carvalho Carneiro

SÃO PAULO

2022

Serviço de Biblioteca e Documentação do IGc/USP

Ficha catalográfica gerada automaticamente com dados fornecidos pelo(a) autor(a) via programa desenvolvido pela Seção Técnica de Informática do ICMC/USP

Bibliotecários responsáveis pela estrutura de catalogação da publicação:
Sonia Regina Yole Guerra - CRB-8/4208 | Anderson de Santana - CRB-8/6658

Zanco Ramos, Gustavo
Hybrid objective function applied to optimize infill sampling location / Gustavo Zanco Ramos; orientador Marcelo Monteiro da Rocha. -- São Paulo, 2022.
109 p.
Tese (Doutorado - Programa de Pós-Graduação em Recursos Minerais e Hidrogeologia) -- Instituto de Geociências, Universidade de São Paulo, 2022.

1. Infill. 2. Optimization. 3. Drillhole. 4. Uncertainty. 5. Modeling. I. Monteiro da Rocha, Marcelo, orient. II. Título.

AGRADECIMENTOS

Ao Meu orientador, Prof. Dr. Marcelo Monteiro da Rocha, pela paciência e apoio na confecção deste trabalho.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo financiamento desta pesquisa.

À DATAMINE pela licença disponibilizada para uso do software Studio RM.

À GEOVARIANCES pela licença disponibilizada para uso do software ISATIS.

Ao Laboratório de Informática Geológica do Instituto de Geociência da USP pela disponibilidade de equipamento.

ABSTRACT

Ramos, G. Z., 2022, Hybrid objective function applied to optimize infill sampling location [Doctorate's Thesis], São Paulo, Instituto de Geociências, Universidade de São Paulo, 101 p.

Different moments of the exploration of mineralized bodies demand that sampling infill be made, those new samples have the objective of furthering knowledge about mineralized rock grade distribution. Usually, drillholes collars are located by geologists with experience and knowledge about the domain under analysis. Other methodologies can be applied to help the decision of where to locate the drillholes, for example, optimization of the infill drillhole location. Optimization is a method to assess the best parametrization to solve a problem, in the case of the infill location the problem depends on what the new samples are made for. Some research utilizes the kriging variance to guide the location of the new samples but has a limitation in assessing the sample distribution uncertainty. Another method that can be applied to locate the infill samples is simulation variance, which is dependent on the sample value. The application of a compost objective function to optimize the infill location is tested. This compost function considers both models kriged and simulated to search for the optimal infill drillhole configuration, therefore, considering both the sample spatial distribution and uncertainty. This method is compared with the objective function that uses either the kriged or simulated data directly to assess the competence of the compost one. Another test considers the influence of the values associated with the samples while searching for the optimum location of drillholes. Those tests have proven that the use of the simulation alone fared better in locating the infill samples in synthetic data than the compost or the kriging-dependent objective function. Both objective functions that utilize direct models, either kriged or simulated, fared better in different distributions. Considering the values associated with the samples, the median fares better than the other 3 values, mean, P10, and P90 of the simulated block distribution. Regarding the methodology of the search is important to notice that optimizing the direction of the drillhole tends to have a better response regarding the objective function but more tests should be made. The optimized infill location tends to further the representativity of the original sampling after the drillholes are done, therefore it can help assess portions of the domain with higher uncertainty that should be considered when the infill location decision is being made.

Terms: **Infill, Optimization, Kriging, Simulation, Objective function, Uncertainty, Modeling.**

FIGURES INDEX

TABLES INDEX

INDEX

1 INTRODUCTION

Infill sampling is necessary for all phases of the mining enterprise. Higher uncertainty of the mining profitability, while in the initial research, demands for the best and more reliable data to decrease it. The location of infill samples is usually dependent on the research demand, being either the region with higher grade and/or higher uncertainty. The decision as to where to locate infilled samples is made by a geologist or mining engineer that knows the data and genetic history of the mineralization. This approach is subjective but effective, despite that an objective approach is necessary to guide the location that considers the needs of the research and the original information disposed of in the research. One approach to contemplate possible infill location sites are made utilizing computational optimization. Different methods of how to optimize the infill were proposed, with many utilizing the estimation uncertainty to guide the model to locate new samples. Besides, the mining project phase is also considered, in the initial stages the focus is on the uncertainty of the data, and while exploiting the mine operational aspects are focused on increasing efficiency.

Several studies were made considering the kriging variance as a guide to locate infill sampling. Examples of those studies are: Szidarovszky (1983); Gershon (1987); Groenigen *et alli* (1999); Delmelle & Goovaerts (2009); Wilde (2009); Soltani *et alli* (2011); Mohammadi *et alli* (2012); Silva & Boisvert (2013); Soltani & Hezarkhani (2013); Dutaut & Marcotte (2020). Those works differ from each other by the optimization algorithm utilized or the application of the kriging variance as the objective function, that can be considered alone, be averaged in the domain, weighted, with the estimated value, and others. However, there is a limitation if the kriging variance is utilized to assess uncertainty, as its value is homoscedastic, considering only the distance and spatial configuration of the samples that were used to estimate the node. This limitation does not mean the application of the kriging variance is useless, once it can point out regions of the domain that are sub-sampled, what must be considered while locating infill drillholes. This limitation arises from the need to consider different approaches while using kriging variance, such as using weights, adding other values to the objective function, or using a completely different model to derive the objective function.

Another approach is infill optimization based on different methodologies, as examples: Boucher, Dimitrakopoulos, and Vargas-Guzman (2004); Boucher, Dimitrakopoulos & Vargas-Guzman (2005); Al-Mudhafer (2013); Martínez-Vargas (2017);

Dirkx & Dimitrakopoulos (2018). Their research is based on simulated models of the data, or the objective function considers other aspects, such as profitability, misclassification, and other relevant factors that should be considered when infilling samples. The important factor to decide the approach, the algorithm, or the objective function, is the purpose of the infill location search.

Two papers compose the bulk of this thesis, one using a compost approach, combining estimated and simulated models to derive an uncertainty factor of the original data that represents local and global uncertainty in a single function. Utilizing kriging variance and simulated model uncertainty, this function represents subsampled portions of the domain while indicating regions with higher local variability. Another distinction is the time frame of the application, in this thesis, the focus is given to the initial stage of the mineral prospect when the uncertainty of the enterprise is higher. To obtain the best information as possible, with high reliability, is extremely important to indicate and further the knowledge regarding the mineralized body spatial distribution, so the focus is to warranty that infill will provide higher representativity as possible. In that way, this work will present the compost objective function and compare it to the approach of the single objective function to assess the competence of each in locating the infill regarding the population data. Another test considers the effect of the value associated with the samples while optimizing the infill location. This method tests 4 possible values that can be utilized as the new sampled value to assess the effects while optimizing. The last test considers the optimization that varies the drillhole direction while searching for the optimum infill location. All analyses were held in synthetic data created by Takafuji (2015) and Takafuji *et al.* (2017) and a real data set on a mine site. The synthetic data is preferred to assess the competence of the methodology proposed as the population distribution is available allowing comparisons with the optimized infilled samples. While the real data demonstrates the application in more complex domains than the synthetic data.

2 HYPOTHESIS

The principal hypothesis in consideration is that sample grade uncertainty and their spatial position improves the infill location and considering these aspects combined as an objective function to be optimized would lead infill sampling to perform better in terms of sample representativity of the population. A secondary hypothesis is that grade values associated with the proposed infill samples would directly impact the optimized new boreholes positions.

3 OBJECTIVES

The main objective of this thesis is to evaluate the competence of the optimization in locating infill and answer the question: does an objective function combining estimated and simulated models fare better in representing the population data when locating infill, or does the singular model objective function is better to further representativity? The second objective is to assess the effects of the value associated with the possible infill samples while optimizing, and what is the best possible value that should be considered.

4 LITERATURE REVIEW

This review focuses on the details of the geostatistical estimation and simulation without over-detailing the models but pointing out the applicability of each. For more details, the author recommends reading the relevant citations presented in the text. A small explanation of the Kolmogorov-Smirnov test will also be presented.

4.1 Geostatistics

The term geostatistics was firstly coined by Matheron in 1963, at that time geostatistics studied the spatial relationship of the data related to mining. Matheron (1963) refers to Krige and Sichel as the precursors of the spatial variability analysis, even if they believed their proposals were classical statistics, their work guided the development of geostatistics as it is known. To differentiate the random variable from the data used by geostatistics, the term regionalized variable was also coined by Matheron (1963).

The regionalized variable values are attached to the spatial position of the analysis (Matheron 1963). There are 4 characteristics related to regionalized variables, being so (Matheron 1963):

Location: spatial coordinates of the data samples.

Support: related to the location, is the physical volume of the sampled data.

Continuity: the variability between neighbors tends to increase as distances increase.

Anisotropies: different directions could have a higher or a lower variability.

To analyze the characteristics of the regionalized variable, and spatial variance of data distribution, geostatistics utilizes the variogram (Matheron 1963). Defined by Matheron (1963) variogram is a graphic that represents the continuity of a mineralization. The variogram abscissa axis represents the distance h of the analysis, and the ordinate axis is the mean spatial variance of data with h vector between them (Matheron 1963). Matheron (1963) defines the variogram as a triple integer of the square difference of data separated by vector h, representing the mining 3D grid. However, the limit in applying the variogram proposed by Matheron is the fact that data is not taken continuously at the mineralized domain. The experimental variogram is utilized to calculate the spatial variability between sampled data, and is represented in the following equation (I) (Isaaks & Srivastava 1989; Chilès & Delfiner 1999):

$$2\gamma(h) = \frac{1}{n}\sum_{i=1}^{n}[Z(X_i + h) - Z(X_i)]^2 \qquad \text{(I)}$$

where $\gamma(h)$ is spatial variance; $Z(X_i)$ is the regionalized variable sampled in the point $(X_i) = (X_i, Y_i, Z_i)$; $Z(X_i + h)$ is the sample data at a distance vector h from $Z(X_i)$; n is the number of sample pairs with vector h separations between themselves.

If the domain is stationary the following relation (equation (II)) between the variogram and spatial covariance is valid (Armstrong 1998):

$$\gamma(h) = C(0) - C(h) \qquad \text{(II)}$$

where $\gamma(h)$ is the variogram $C(0)$ is the null distance covariance (statistical variance) and spatial covariance given a distance h is $C(h)$. This correlation is graphically represented in Figure 01.

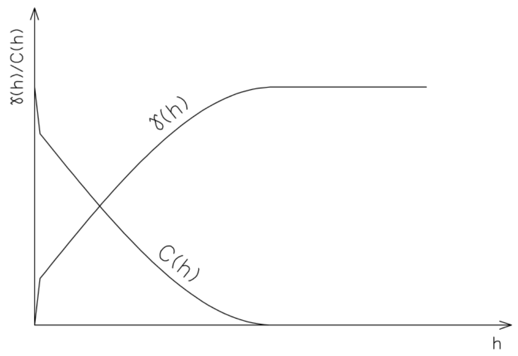


Figure 01 – Relation between variogram and covariogram. (Taken from Yamamoto 2001).

To quantify anisotropies the variogram must be calculated in different directions, varying the directions of the vector *h* the occurrence of anisotropies can be analyzed (Matheron 1963; Armstrong 1998). Different methods to calculate multidirectional variogram were proposed, the directions should be related to the geological structure of the data (Armstrong 1998; Deutsch 2002). In the case of no noticeable preferential direction, it is recommended to start North and take one direction every 45 degrees until South (Armstrong 1998; Olea 1999). The vertical Direction should be considered in the variogram if the data is tridimensional (Deutsch 2002). The value of *h* can be derived, from each direction, as the mean value of the first neighbor at each direction of interest (Olea 1999). Other parameters can be defined to better calculate and represent the spatial variability of the domain, for details refer to Olea (1999) and Deutsch (2002).

After the initial analysis of the anisotropy, the next step is to define a variogram model to represent the spatial variability continuously (Deutsch 2002). The model must be valid and adjust well to the experimental variogram while being admissible by a group of factors and tests. For more details on variogram models refer to Armstrong (1998); Olea (1999) and Deutsch (2002).

4.2Kriging

Kriging refers to a series of estimation methods that minimizes the mean square errors of estimate (Deutsch 2002). The term BLUE, best linear unbiased estimator, is used to describe kriging, as the method that minimizes the variance error between the real and estimated data (Armstrong 1998; Deutsch 2002). Generally, the estimative is made as a weighted mean with minimal variance (Armstrong 1998). Matheron (1963) defined kriging as a grade estimation of a panel computing weighted means of proximal samples, the name honors Professor Daniel G. Krige and his pioneer studies in quantifying spatial variability. The kriging estimator equation (III) presented by Matheron (1963) is:

$$Z^*(X_0) = \sum_{i=1}^{n} \lambda_i Z(X_i) \quad \text{(III)}$$

where $Z^*(X_0)$ is the estimated data at position X_0 given by the summation of the weights λ_i associated with each sample $Z(X_i)$.

Secondly, the weights should be such that the variance between real and estimated data should be minimal (Matheron 1963).

Matheron (1963) addresses as advantages of kriging that the method returns the best estimative values, with the smallest variance; and can help assess the mine's future production.

To Armstrong (1998) factors that guarantee kriging the estimation accuracy are: 1) number and quality of samples at each point; 2) position of samples in the domain; 3) distance between samples and estimated point, with more acuity around sampled sites; and 4) spatial continuity of the variable of interest.

In this thesis, two kriging methods will be considered, simple kriging and ordinary kriging.

Simple kriging considers that the mean is known (Armstrong 1998) and to define its weights the estimative error is needed (Olea 1999), therefore, the error expected value should be equal to zero. Simple kriging equation (IV) is presented as (Olea 1999 and Deutsch 2002):

$$Z_{SK}^*(X_0) = m_0 + \sum_{i=1}^{n} \lambda_i (Z(X_i) - m_i) \quad \text{(IV)}$$

where $Z_{SK}^*(X_0)$ the simple kriging estimator is given by the mean at the unsampled location m_0 plus the sum of the weighted difference between samples and the local mean m_i.

The weights are defined by a linear equation system derived from the estimated error variance (Armstrong 1998), where the number of equations is dependent on the number of samples used to estimate (Olea 1999). The linear equation deduction can be seen in detail by Amrstrong (1998) and Olea (1999). From those deductions the system is given by the following equation(V), in matrixial presentation (Olea 1999):

$$\begin{bmatrix} COV(X_1,X_1) & COV(X_2,X_1) & \cdots & COV(X_n,X_1) \\ COV(X_1,X_2) & COV(X_2,X_2) & \cdots & COV(X_n,X_2) \\ \vdots & \vdots & \vdots & \vdots \\ COV(X_1,X_n) & COV(X_2,X_n) & \cdots & COV(X_n,X_n) \end{bmatrix} * \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{bmatrix} = \begin{bmatrix} COV(X_0,X_1) \\ COV(X_0,X_2) \\ \vdots \\ COV(X_0,X_n) \end{bmatrix} \quad \text{(V)}$$

where $COV(X_1, X_1)$ is the spatial covariance between the sample X_1 and itself, $COV(X_2, X_1)$.is the spatial covariance between sample X_2 and sample X_1 and so on.

The simple kriging variance is presented as shown in the following equation (VI) (Deutsch & Journel 1998):

$$\sigma_{SK}^2(X_0) = COV(0) - \sum_{i=1}^{n} \lambda_i COV[X_0, X_i] \quad \text{(VI)}$$

Ordinary kriging considers the mean unknown but locally stationary, being a weighted mean that can be calculated as shown in equation (VII) (Armstrong 1998):

$$Z_{OK}^*(X_0) = \sum_{i=1}^{n} \lambda_i Z(X_i) \quad \text{(VII)}$$

Considering the mean of estimate error and the fact that the method is unbiased the equation, known as the unbiased condition of the ordinary kriging, is shown in equation (VIII) (Armstrong 1998):

$$\sum_{i=1}^{n} \lambda_i = 1 \quad \text{(VIII)}$$

Different from the simple kriging system, where there are n variables and n equations – each one related to a sample chosen to estimate – the ordinary kriging system has $n + 1$

equations and $n+1$ variables. The additional equation is related to the constraint imposed by the unbiasedness condition. Therefore, the ordinary kriging system is given by equation (IX):

$$\begin{bmatrix} COV(X_1,X_1) & COV(X_2,X_1) & \cdots & COV(X_n,X_1) & 1 \\ COV(X_1,X_2) & COV(X_2,X_2) & \cdots & COV(X_n,X_2) & 1 \\ \vdots & \vdots & \vdots & \vdots & 1 \\ COV(X_1,X_n) & COV(X_2,X_n) & \cdots & COV(X_n,X_n) & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix} * \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ -\mu \end{bmatrix} = \begin{bmatrix} COV(X_0,X_1) \\ COV(X_0,X_2) \\ \vdots \\ COV(X_0,X_k) \\ 1 \end{bmatrix} \quad \text{(IX)}$$

Ordinary kriging variance is computed as presented in equation (X):

$$\sigma_{KO}^2(X_0) = \sum_{i=1}^{n} \lambda_i \, \gamma(X_i, X_0) + \mu \quad \text{(X)}$$

Some kriging characteristics are important to be considered in this thesis, the first kriging is a conditional method, in other words, if an estimated point coincides with a sample, the estimative is equal to the sample value as can be shown in Chilès & Delfiner (1999). The second characteristic is the smoothing of the estimative, the dispersion of estimate distribution is smaller than the original/sampled data one, this is given by the fact that the estimated variance differs from the sample variance the exact value of the kriging variance (Chilès & Delfiner 1999).

Another important point is the homoscedastic characteristic of the kriging variance, meaning that the value is not dependent on the sampled values considered while estimating, which disables kriging variance as an uncertainty value (Armstrong 1994). The kriging variance value is dependent on the spatial configuration of the samples used while estimating, therefore, two identical configurations, with different sample values, will have the same kriging variance, given the same variogram model (Armstrong 1994).

4.3 Simulation

Olea (1999) points to limitations of kriging that are related to the smoothing effect:

- The estimation variogram differs from the sample variogram.
- The estimation histogram differs from the sampled data histogram.
- Kriging tends to underestimate high values and overestimate small values.

Stochastic simulations were proposed to deal with the smoothness of the estimative (Deutsch & Journel 1998; Olea 1999). The stochastic simulation in geostatistics refers to models that compute different and equiprobable values given a variogram model (Deutsch & Journel 1998). Kriging models represent a global trend while the simulated models are sensitive to local variability, the equiprobable models can be used to derive uncertainty of the distribution, therefore, the global characteristics and statistics are maintained to the detriment of the local accuracy (Deutsch & Journel 1998). Simulation computes l equiprobable realizations, called conditional if it is conditioned by the data (Olea 1999). The choice of which model is given by the user's interest, if he wants to minimize local error, he should opt to use kriging, while if the aim is to keet spatial continuity, simulation must be chosen (Olea 1999).

The focus of this work will be Sequential Gaussian Simulation (SGS) which is a conditional method. The algorithm of the SGS as presented in Deutsch & Journel (1998), Olea (1999) and Deutsch (2002) are:

1. Transform the sample distribution into a normal distribution $N[0,1]$, if the data is not normal, Gaussianity is a requirement of geostatistical simulation methods.
2. Adjust a variogram model to the normal distribution experimental variogram.
3. Define a grid or block model to be simulated.
4. Define the number of realizations.
5. Define a seed that shall produce a random path of nodes to be simulated at each realization.
6. Estimate the node by simple kriging and compute its simple kriging variance considering the local samples and previous simulated nodes in the neighborhood.
7. Consider $N\left(Z_{KS}^{*}(X_0), \sigma_{KS}^{*}(X_0)\right)$ as the normal distribution of the possible values at the given node.
8. Obtain the simulated value at the node drawing it at random pull from the normal distribution $N\left(Z_{KS}^{*}(X_0), \sigma_{KS}^{*}(X_0)\right)$ based on the equation (XI):

$$Z_{SGS}^{l}(X_0) = Z_{KS}^{*}(X_0) + R(X_0) \quad \text{(XI)}$$

where $Z_{SGS}^{l}(X_0)$ is the simulated node; $Z_{KS}^{*}(X_0)$ is the simple kriging estimator; $R(X_0)$ is drawn by classical Monte Carlo simulation that considers the variance value $\sigma_{KS}^{*}(X_0)$.

9. Add the $Z_{SGS}^{l}(X_0)$ value data to be considered in the next simulations on the same realization.
10. Repeat steps 6 to 9 until the last node is simulated.
11. If the data was transformed in normal back transform, it to the original distribution.

The use of a Gaussian distribution guarantees that the simulated distribution is representative of the sample distribution, which isn't when other distributions are considered (Deutsch 2002). Working in the Gaussian domain guarantees that the result is Gaussian and, therefore, that the mean, variance, and variogram model are reproduced (Deutsch 2002). However, the reproduction of the statistics is not perfect, even with the Gaussianity of the data, this is an effect associated with the uncertainty of the sampled values (Deutsch & Journel 1998). The ergodic fluctuations in results may occur, with the data being ergodic if the analyzed parameter tends to the real (sampled) with more realizations being made (Deustch & Journel 1998). With more conditional data the ergodic fluctuation should be smaller (Deutsch & Journel 1998). Therefore, if the data is stationary, ergodic, and different realizations are made, it is expected that the statistics of each realization reproduce the real model of the data (Deutsch & Journel 1998).

4.4 Optimization

Optimization refers to a decision-make an act or resolving complex problems by modifying values of the variables related to the objective while computing the quality of the results attained (Luenberger & Ye 2008). The objective is maximized or minimized while considering the restrictions imposed on the variables that the problem is subjected to (Luenberger & Ye 2008). The complexity, relations between variables, and limitations could turn the solution inviable, or even impossible to attain a definitive answer, therefore, usually the optimization approximates which results indicates the closer solution of the ideal desired (Luenberger & Ye 2008). The mathematical model of the objective shall be accurate to best represent the reality of the problem and the solution must be attained in a reasonable time (Luenberger & Ye 2008).

Linear programming, the original optimization methodology to solve linear problems, where developed by Dantzig in 1947 with the simplex algorithm (Dantzig 1981). Even when considering that as previous work Kantorovich, in 1939, deals with the planning and distribution of materials and workforce in the USSR, his work only gained notoriety with the development of mathematical programming (Dantzig 1981). The contributions of Dantizig, as presented by himself in 1981, are: practical planning can be formulated as mathematical systems of linear inequalities; selection of optimal planning with an objective; and the development of the simplex.

Optimization uses basic equations that represent the problem, this can be generically presented as the following equation (XII) (Schäffer 2012, Snyman & Wilke 2018):

$$Minimize\ f(x): x = [x_1, x_2, \dots, x_n] \ \in \ \mathbb{R}^n$$
$$Subjected\ to\ g(x) \geq L \tag{XII}$$

The function $f(x)$ represents the problem, it is called the objective function and must honor the relation imposed by $g(x)$ in the search and its relationship with the value of interest. Optimizations that consider limitations are called restrictions, those limitations are given by functions of inequality or equality (Snyman & Wilke 2018). Representing the problem using the objective function is extremely important once its development and modeling will indicate the formulation and methodologies necessary to attain the optimal result desired (Snyman & Wilke 2018).

Natural phenomena rarely can be represented by linear equations; therefore, the use of non-linear programming is preferred (Pintér 2009). One of the characteristics of non-linear functions is the occurrence of local optima, which in turn points to the interest in utilizing global optimizations (Pintér 2009). The limitation in those cases comes from the computational and time limitations that arise from the solution of non-linear problems, in the case of continuous data the number of iterations tends to infinity, which indicates the necessity of applying stochastic methods (Romeijn 2009).

Global stochastic optimizations must be applied when there is no evidence of the problem having a real solution, which may be related to dimensional reasons, the search space of the objective function, or the fact that optimal value cannot be attained in a feasible computational time (Schoen 1991). The addition of a stochastic parameter in the algorithm can be applied to solve those difficulties when optimizing (Schoen 1991). There are two distinctions applicable to stochastic optimization algorithms (Schoen 1991): 1) stochastic model, the objective function is given because of a stochastic process; and 2) in stochastic algorithms, the values of the objective function are defined stochastically. Different algorithms deal with both types of stochastic approaches through probabilistic and heuristic search (Schoen 1991). The application of global optimization recommends the search division in three parts (Schoen 2009): global search, which contemplates the whole domain; local search, limiting the region to where a given local optimum is acceptable; and stop criteria.

As heuristics is related to stochastic and global optimizations it is important to detail it. Heuristics refers to a methodology to define the best and most efficient result based on the different alternatives related to the desired objective (Pearl 1984). Two exigencies limit the heuristic search (Pearl 1984): the criteria that define the best value being simple; and the possibility to differentiate between good and bad values while searching. Therefore, heuristics do not guarantee that the best overall possible value overall is attained, once not all possible values are contemplated in the search. However, the result is sufficient in most of the applications (Pearl 1984). Applying heuristic optimization has the main goal indicate an acceptable good option among a big number of configurations even if not attaining the best result (Pearl 1984). When dealing with big domains, or with the number of possibilities tending to be infinite, heuristics reduce the number of evaluations of the objective function and attain solutions in a feasible computational time (Pearl 1984). According to Bianchi *et* al. (2008), there are two groups of heuristic algorithms: the constructive algorithms, which add solution components to search the solution; and the localized search algorithms, which use an initial solution that is developed by modifying its values.

Different modifications can be applied in heuristic optimization to evade local optimum, those algorithms are usually called metaheuristics (Voss 2009). Metaheuristics refer to methods of search applied to an objective function parameters domain to diversify the result and intensify the search to attain the result (Blum & Roli 2003). Diversifying the search refers to the exploration of the solution domain and intensifying is related to the comparisons between results to guide the search (Blum & Roli 2003). The search made by metaheuristics algorithms depends on the interest while searching, so it can be made while evading local optima with the algorithm searching for best results, as the simulated annealing (SA) algorithm does (Blum & Roli 2003). Other algorithms use learning where the components of the search "learn" to differentiate regions of the domain with higher interest, where the results tend to be better (Blum & Roli 2003).

This thesis will focus on the simulated annealing algorithm that was applied in the papers. Simulated annealing applies the effects of the annealing and cooling in solids to solve problems with high complexity or infinity number of solutions (Romeijn 2009; Gall 2014). The SA method is based on the work developed by Metropolis *et al.* (1953) that uses a computational methodology to describe the properties of substances by considering the interaction of the molecules that compose it while applying equations of state and simulating through a modified Monte Carlo method to obtain the spatial configuration of those molecules. Application of the methodology proposed by Metropolis *et al.* (1953) as a heuristic optimization is presented by Kirkpatrick *et al.* (1983), which shows the relation between the idea of metal smelting and the slow cooling of this system to search for the best position (solution) of each molecule that "evade" imperfections to form the crystal. The principal idea of Kirkpatrick *et al.* (1983) is to present the relation between the arrangement of the particles by the energy present in the system with the search for the most stable configuration, as the acceptance of only the best relates to a fast-cooling and more unstable system. Therefore, the algorithm usually accepts worse configurations to guarantee that the best solutions can be attained (Kirkpatrick *et al.* 1983; Romeijn 2009). The probability of acceptance of worse results is important in the SA method as it guarantees that the search for the optimal continues even when a local optimum is obtained, which can be evaded to continue the search for a possible best optimum (Romeijn 2009).

The basic SA algorithm varies at random the position of the particle and computes the result of the new configuration applied in the objective function, when a better value is attained the new position of the particle is accepted; if a worse result is obtained the acceptance of the new position is given by a probability test given by the method, which decides if it will or not be accepted (Kirkpatrick *et al.* 1983). The proceeding of acceptance of worse configurations is related to the temperature of the system, meaning, the domain is heated up to a maximum temperature at the initial point and is cooled progressively during the search, therefore, there is a higher probability of acceptance of worse results when the system temperature is higher; this procedure is called cooling schedule (Kirkpatrick *et al.* 1983; Romeijn 2009). The probability of acceptance of worse results is related to a Boltzmann distribution given by the relation with the system entropy, which is dependent on the temperature; when a worse value is obtained a random value is drawn from a uniform distribution (0,1), and in case of the drawn value being smaller than the probability calculated the worse value is maintained and the optimum is actualized (Pardalos & Mavridou 2009).

5 MATERIALS AND METHODS

5.1 Materials

The materials utilized in this thesis were: the synthetic data used to assess the optimization methods and the original sampling made over that domain; the real data of the Capanema mine, and the computational programs used to develop the geostatistical analysis, and sample the synthetic domain.

5.1.1 Synthetic domain and sampling

The optimization algorithms were applied in a synthetic database developed by Takafuji (2015) and Takafuji *et al.* (2017). The data set mimics metamorphized sandstone and phyllite that were folded and affected by a reverse fault, where the copper mineralization occurs in a quartzite (Takafuji 2015 and Takafuji *et al.* 2017). The whole synthetic domain has 600 meters on the North axis, 300 meters on the East axis, and 300 meters in depth.

A sampling composed of 32 drillholes was drawn at random with most of the drillholes being perpendicular to the ore body. The statistics of the sampling are presented in Table 01. In Figures 02 and 03 the base map of the drillholes samples and the histogram of the sampling are, respectively, presented.

Table 01 – Statistic of the copper sampling made over the synthetic domain.

Basic statistic	
Mean	0.602
Variance	0.034
CV	0.308
Minimum	0.501
Maximum	1.656
Median	0.524

(Taken from Ramos 2016)

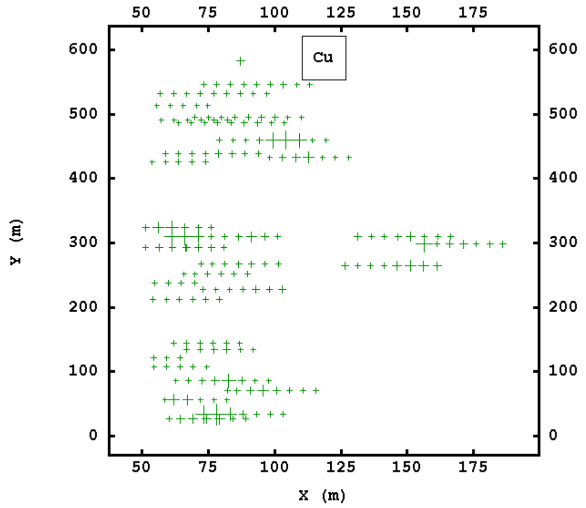


Figure 02 - Base map of the sampling made. The samples from each drillhole are presented as crosses, with the size of each cross being dependent on the copper percentage attained. (Source: Author).

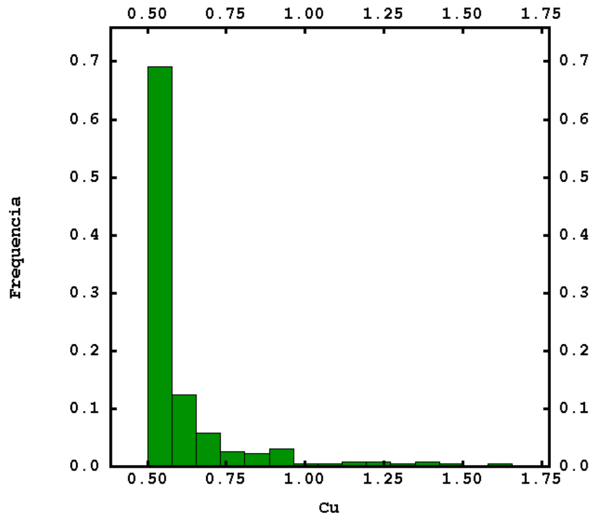


Figure 03 – Original sampling histogram (Source: Author).

5.1.2 Capanema Mine data set

The Capanema Mine is in the Quadrilatero Ferrífero area on Minas Gerais state, Brazil. The iron ore mineralization occurs in a BIF that is subjected to a synclinal fold. Details of the genetical and geological mineralization body of Capanema Mine can be seen in Massahud & Viveiros (1983), Fonseca (1990), and Rocha (1999). The infill tests were made in the drillhole sampling while the competence of the new data location was taken regarding the rockdrill data set of the mine while active as the population. The drillhole data is composed of 69 collars that were regularized by benches of 13 meters in height. Capanema mine is exploited on an open pit with approximate dimensions of 2000 (NW axis) by 400 (NE axis) meters and 200 meters in depth. In Figure 04 the base map of the samples is presented, most of the drillholes are vertical with one exception. Table 02 presents the statistics of the Capanema Mine drillhole and rockdrill sampling.

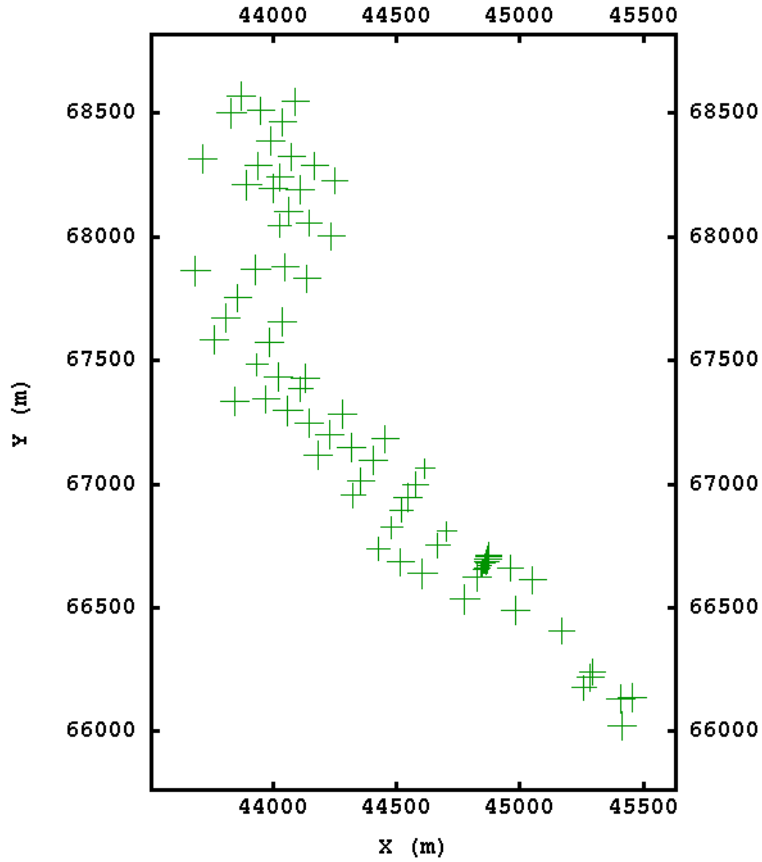


Figure 04 - Capanema Mine drillhole sampling base map. The size of the crosses indicates iron grades at the location (Source: Author).

Table 02 – Statistic of Capanema Mine drillhole and rockdrill sampling.

Basic statistic		
	Drillhole	Rockdrill
Mean	55.77	59.95
Variance	58.22	28.12
CV	0.1368	0.0884
Minimum	27.15	3.08
Maximum	66.8	67.67
Median	57.66	61.56

5.2 Methods

The methods will be presented concerning the applied objective functions and optimization algorithm. The methodology and details of the geostatistical analysis were not presented, and such is not the focus of the thesis.

5.2.1 Objective Functions

Five objective functions were applied in this thesis, two presented in Ramos (2016) and three developed in the present work. The objective functions are: Simulated Block Variance; Simulated Block Coefficient of Variation; Kriging Variance Sum; Simulated and Kriged Block Variance; and Simulated Block Coefficient of Variation and Kriged Variance.

5.2.1.1 Simulated Block Variance (SBV)

The Simulated Block Variance (SBV), as defined by Ramos (2016) represents the uncertainty related to the original sampling, the function value is attained after simulating the domain by SGS and attaching all the simulated blocks variance. To be used as an objective function the SBV necessarily needs that a new simulation is made at each iteration, considering the new probable collar infill configuration, while the search for an optimum continues. The algorithm should then minimize the SBV value at the optimization. The SBV function is presented in equation (XIII):

$$Minimize\ SBV = \sum_{i=1}^{N} \sigma_{SGS}^{2}(X_i)$$
$$Subject\ to\ n * Z(X_M): X_M = [X_1, X_2, \dots, X_M];\ (X_M) = (X_M, Y_M, Z_M);$$
$$X_{min} \leq X_M \leq X_{max}; Y_{min} \leq Y_M \leq Y_{max};\ Z_M = Z_{Bmax} \qquad \text{(XIII)}$$

where $\sigma_{SGS}^{2}(X_i)$ is the simulated block variance at a node, N is the number of simulated nodes, $n * Z(X_M)$ are the n collar infill locations that should be optimized, X_M, Y_M, Z_M are the coordinates of the collar location in the axis East, North, and elevation, X_{min} and X_{max} are the limits imposed by the user at the coordinates in the East axis, while Y_{min} and Y_{max} are limits imposed in the North axis and Z_{Bmax} is the defined as the highest elevation value of the block model used, therefore the infill collars are considered with a fixed elevation value.

5.2.1.2 Simulated Block Coefficient of Variation (SBCV)

The Simulated Block Coefficient of Variation (SBCV), as defined by Ramos (2016) uses the same procedure as SBV, simulating each iteration by SGS, but instead of using simulated block variance the value considered is the coefficient of variation. This function is interesting because while optimizing infill collars minimizing SBCV values would not only minimize the uncertainty of the domain but would also raise the mean simulated values, as the coefficient of variation value is inversely proportional to the mean value of the distribution considered. Therefore, the minimization of SBCV should return not only the collars that reduce uncertainty but also rises the grades. The SBCV function is presented in equation (XIV):

$$Minimize\ SBCV = \sum_{i=1}^{N}\left(\frac{\{\sigma_{SGS}^{2}(X_i)\}^{1/2}}{\overline{Z^S}(X_i)}\right)$$
$$Subject\ to\ n * Z(X_M): X_M = [X_1, X_2, \dots, X_M];\ (X_M) = (X_M, Y_M, Z_M);$$
$$X_{min} \le X_M \le X_{max}; Y_{min} \le Y_M \le Y_{max};\ Z_M = Z_{Bmax} \qquad \text{(XIV)}$$

where $\overline{Z^S}(X_i)$ is the mean of the simulated values on the node.

5.2.1.3 Kriging Variance Sum (KVS)

The Kriging Variance Sum (KVS) considers kriging variance as an objective function to be compared to simulation-based objective functions and the compost function, which considers both simulation and kriging in the same function. Minimizing KVS should locate the infill collars in the regions of the domain with low sample density. Nowadays it is relatively easy and fast to estimate a domain with kriging, so in this work, at each iteration, a new estimative will be done to assess how the global kriging variance value, computed as the sum of kriging variance of all estimated nodes, changes from an iteration to another. The KVS function is presented as follows, at equation (XV):

$$Minimize\ KVS = \sum_{i=1}^{N} \sigma_{KO}^2(X_i)$$
$$Subject\ to\ \ n * Z(X_M): X_M = [X_1, X_2, \ldots, X_M];\ (X_M) = (X_M, Y_M, Z_M);$$
$$X_{min} \leq X_M \leq X_{max}; Y_{min} \leq Y_M \leq Y_{max};\ Z_M = Z_{Bmax} \quad \text{(XV)}$$

with $\sigma_{KO}^2(X_i)$ being the ordinary kriging variance of the block.

5.2.1.4 Simulation and Kriging Block Variance (SKBV)

The Simulated and Kriged Block Variance (SKBV) was developed to consider both the uncertainty derived from the simulation and the estimation of the domain, therefore considering the sampling uncertainty and spatial sample distribution. Minimize the SKBV as an objective assessment of the influence of the regions with lower sample density in the locations of infill while still considering the factor of grade uncertainty provided by simulation. When the SKBV is used both simulation and kriging should be made at each iteration and then compute the objective function value, this demand more computational time. Another change from the previous objective functions regards the fact that weighs (a and b) are associated with each term of the function, the simulated and the kriged related sides of the function, that summed should be constrained to 1. The SKBV function is presented in equation (XVI):

$$Minimize\ SKBV = a * \sum_{i=1}^{L} \sigma_{SGS}^2(X_i) + b * \sum_{j=1}^{N} \sigma_{KO}^2\left(X_j\right)$$
$$Subjected\ to\ n * Z(X_M): X_M = [X_1, X_2, \ldots, X_M];\ (X_M) = (X_M, Y_M, Z_M); \quad \text{(XVI)}$$
$$X_{min} \leq X_M \leq X_{max}; Y_{min} \leq Y_M \leq Y_{max};\ Z_M = Z_{Bmax};$$
$$and\ a + b = 1;$$

where a is the weight associated with the simulation side of the function and b is the weight associated with the kriging side of the function; L is the number of simulated nodes on the domain, note that L can be different from the number of kriged blocks N.

5.2.1.5 Simulated Block Coefficient of Variation and Kriging Variance (SBCVKV)

The Simulated Block Coefficient of Variation and Kriged Variance (SBCVKV) is the last function considered and follows the same procedure as the SKBV, with the difference being the use of the coefficient of variation in the place of the variance. The same strong point of the SBCV is considered in this function, when SBCVKV is minimized not only the uncertainty would be reduced but, at the same time, the mean ore percentage should rise, both added to the consideration of regions of the domain with low sample density. The SBCVKV function is presented in equation (XVII):

$$Minimize\ SBCVKV = a * \sum_{i=1}^{L} \left(\frac{\{\sigma_{SGS}^2(X_i)\}^{1/2}}{\overline{Z^S}(X_i)} \right) + b * \sum_{j=1}^{N} \sigma_{KO}^2\big(X_j\big)$$
$$Subject\ to\ n * Z(X_M): X_M = \ [X_1, X_2, \dots, X_M];\ (X_M) = \ (X_M, Y_M, Z_M); \quad \text{(XVII)}$$
$$X_{min} \leq X_M \leq X_{max}; Y_{min} \leq Y_M \leq Y_{max};\ Z_M = Z_{Bmax};$$
$$and\ a + b = 1$$

5.2.2 Optimization algorithm

The algorithm applied to optimize the location of the infill drillhole is based on the simulated annealing (SA) procedure. Optimizing by simulated annealing method in the present thesis was adapted to the interests of the search, considering the different possible objective functions applied, the method of location of the infill collars, and the parametrization while searching for the optimum. The direction of the drillholes located can be optimized in the algorithm, which can modify the azimuth and dip of each collar at each iteration.

In the SA algorithm collar parameters, at each iteration, are changed through a unidimensional approach, meaning that: first, the only collar will be modified during the iterations, and the choice of which will be modified is made at random from the *n* collars defined to be considered as infill; second, the modification of the collar is unidimensional, meaning that only one of the parameters that define the collar (East, North, Azimuth, Dip) shall be modified at the iteration, the choice of which parameter will be changed is made at random too. The elevation parameter is not considered in the modification as the elevation of each collar is considered fixed or is dependent on the topography of the domain, in that way this information must be provided.

Another difference is the value associated with the samples while optimizing the infill. As some of the objective functions consider the simulation model, the value of the samples considered in the neighborhood will change the simulated value, so the value that samples receive when being considered as probable infill data can influence the result of the search. Therefore, the definition of which value shall be associated with the possible infill samples can be defined as one of 4 values: the mean, median, P10, or P90 of the nearest simulated node to the sample considered. This procedure was adopted to assess the influence of the values associated with the samples while optimizing the infill drillholes.

The difference related to the objective function is in the sense of considering the models while optimizing, being kriging, and/or simulation. Originally the algorithm used only simulated models, but with the new objective function proposed the kriging value can be considered by itself or with the simulated value. So, in the case of the direct objective functions, only one of the models is made in each iteration and is considered to calculate the objective function. When the compost objective functions are considered both, kriging and simulation should be made at each iteration to then compute the objective function. Each of those approaches is automatically made when the objective function to locate the infill is defined.

For this thesis, only a fast cooling schedule was considered, which minimizes the system "temperature" by the following equation (XVIII).

$$\frac{\left\lfloor \frac{\mathrm{niter}+1}{\mathrm{i}+1} \right\rfloor}{\mathrm{niter}} \tag{XVIII}$$

where the temperature of the present iteration is dependent on the present iteration number, i, and the total number of iterations, niter.

The temperature value is compared with the control value, a random number in the interval [0,1], only when a worse objective function is obtained. The worse value will only be accepted as a new optimum when the temperature is higher than the control value.

6 RESULTS

To assess the competence of the infill sampling optimization proposed methods a series of tests were performed. The workflow is: first comparing the SBV and KVS objective function in optimizing the infill is done; secondly, a comparison considering the five objective function competence in optimizing the infill position was performed; third the influence of the value associated with the infill sampling was tested while optimizing; lastly the influence of optimizing the drillhole direction while searching the best infill configuration.

The first and second tests were completed, with the results of minimizing the objective function and the optimization being presented and compared. The third comparison was only made in the synthetic data set. The last results are presented only to show the effect of the method in minimizing the objective function.

Each of those tests was made to assess which optimization parametrizations could perform better in terms of infill configuration, in the sense of furthering the original sampling populational representativity. A series of comparisons considering the population statistical parameters were made to that end. By the end of this section, a discussion of the best parametrization will be presented to demonstrate which reached the best results, i.e., infilled samples with higher representativeness of the population.

6.1 Paper 1: comparisons of infill sampling optimization using simulated, and kriging based objective functions

6.1.1 Introduction

The process of sampling is basic to further the knowledge of a mineral deposit. Using samples as a guide one can interpret and predict (by estimation or simulation) values to assess the population, i. e. the exhaustive data information of the phenomena. During a mine, life is common to produce new sampling campaigns to further, even more, the data representability. Problems related to where locate new samples rises, mostly, considering cost and accessibility to acquire more information. One common reason behind infill samples is the uncertainty of the original data. Model uncertainty must be considered, but more so one must infer the data uncertainty to better represent the reality of the phenomena studied. Using models one can assert areas or portions of the domain with greater uncertainty to consider increasing sampling at those locations. The problem arises from the doubt if those new samples will add the information gathering positively, i. e. the new data have a higher representativity of the population facing the economic cost of obtaining said data. Therefore, it is necessary to test methods to locate the new samples and, also, the representativeness of this new information regarding the population being sampled.

Kriging variance is an uncertainty measurement derived from the kriging procedure, which was tested as one of these methods by several authors. Gershon (1987) minimizes the kriging variance to locate new samples. Lloyd & Atkinson (1999) compared the location of new drillholes using the mean of ordinary kriging variance and the conditional variance of different thresholds, based on the indicator kriging. Wilde (2009) uses six different algorithms to optimize the objective function that minimizes the simple kriging variance sum of all estimated blocks. Soltani *et al.* (2011) minimize the average kriging variance to guide the genetic algorithm in locating new drillholes in 3 dimensions. Mohammadi *et al.* (2012) optimized sampling locations through a simulated annealing algorithm guided by the average of kriging variance weighted by block estimate. The average kriging variance considering drillhole dip in locating the infill was used by Soltani & Hezarkhani (2013) to maximize the information amount inside the orebody. Silva & Boivest (2013) minimize the kriging variance average to locate the new drillholes using 4 optimization algorithms, in addition for search the objective function considers the block classification and maximization of the tonnage of each block. Safa & Mohammadi (2017) applied the minimization of the kriging variance added with a local variance, consisting of the weighted average of the data samples used in the estimative and the estimated value difference, to guide the search for new drillholes location. The kriging variance is minimized in search of the infill configuration by the particle swarm optimization method in Fatehi *et al.* (2017). Dutaut & Marcotte (2020) applies the kriging estimative and variance to determine the new drillholes position, considering the distance of the new data from the block center as a selection of possible candidates for the infill position.

Some of the cited papers try to develop the representation of the kriging variance as a decision maker to locate new drillholes. This procedure is important due to kriging variance being limited in the sense of uncertainty representation, as its value is not dependent on the sample data used in the estimation, rather is dependent on the variogram model and the data configuration, i. e. distance on the data to the point estimated (Armstrong, 1994; Goovaerts, 1997). For the reasons previously exposed the use of kriging variance could point out sub-sampled portions in the domain, rather than provide an assessment of the uncertainty of the data. Being the focus of the infill further the representativity the use of kriging variance is recommended in association with another tool in other to decide the best locations for infilled drillholes.

Another approach to infill location optimization is to apply simulated models to guide the search for the best sampling configuration. The simulation can either be used in the objective function or in the search for the best infill configuration. Usually, the values of the possible new drillholes are drawn from the simulation and then the objective function is calculated. Some papers that use simulation to locate infill are: Goria *et al.* (2001) that simulated golden ore and uses 3 different values derived from the simulation to attribute the ore data to new drillholes, with the new samples a new simulation is calculated for each scenario to analyze the dispersion attained; Pilger *et al.* (2001) simulate the original data and uses the interquartile range to assess the local and global uncertainties and locate the new samples in the blocks with higher range value, after the location a new simulation is computed with the new data set to observe if the local and global uncertainty value was reduced. Martinéz-Vargas (2017) uses simulation to define values to different drillhole configurations and calculate the cost associated with the information, regarding the errors associated with block misclassification and the drillholes sampling cost; Pinheiro *et al.* (2017) use as an objective function the sum of average block variance and the width of the 95% probability interval of the simulation to guide the infill location. Zagré *et al.* (2018) use the variance of the Bernoulli variable based on the probability of a given block being ore or waste to assess the uncertainty, using simulation to calculate the distribution of possible value and from that derive the probabilities needed, locating the new drillholes on portions with high uncertainty; In Dirkx & Dimitrakopoulos (2018) new drillholes receive values drawn from a simulation which is then re-simulated, classified and compared with the previous classification; Nowak & Leuangthong (2019) calculate the confidence limit of panels simulated with new drillholes to assess the uncertainty, used to classify the information, with less uncertainty the panel class is higher.

The present paper seeks to compare the use of kriging and simulation approaches in locating new drillholes. As such it is proposed the application of two objective functions that have the shared intent: locate, under the same constraints, the best infill position drawn from the same original dataset. The objective function based on simulation results is the sum of the simulated block variance and the one based on kriging is the sum of the kriged variance. The first objective function is given that for each simulation block there are L equiprobable values, considering that it is possible to calculate the variance of the simulated block, i. e. the local uncertainty of the simulated value. Applying the sum of the simulated block variance one can represent the global uncertainty associated with the original sampling information. The second objective function uses kriging variance as the guide to search for the best new drillholes locations through a global approach using the sum of kriging variance calculated for each block in the domain. The simple idea is that the second function shall guide the search towards portions with less information on the domain, while the first function guides the search towards areas with higher uncertainty based on the original samples. The search was developed using an algorithm based on the simulated annealing (SA) optimization, and the parametrization was made considering the number of new drillholes sampled and the number of iterations used to complete the search.

The objective of these comparisons is to point out how capable each approach is to further the representability of the original sampling regarding the population. Tests were made in a synthetic 3D body representing a geological occurrence of copper. Therefore, for each optimized configuration, the new drillholes can be made easily and the results could be compared to attest efficacy.

6.1.2 Methods and Materials

6.1.2.1 kriging variance sum (KVS)

The object function based on the estimative is the block kriging variance sum (KVS) of the domain. The ordinary kriging variance is presented in equation (I) (Armstrong 1998, Olea 1999):

$$\sigma_{KO}^2(X_0) = \sum_{i=1}^{n} \lambda_i * \gamma(X_i, X_0) + \mu \qquad \text{I)}$$

where $\sigma_{KO}^2(X_0)$ is the kriging variance calculated regarding the estimated location X_0. λ_i is the weight associated with the i-est sample. $\gamma(X_i, X_0)$ is the spacial variance between the sample X_i and X_0. μ is the Lagrange term attained from the ordinary kriging equations. For more details about ordinary kriging and its variance proceedings, equations, and parametrizations the authors refer to Isaaks & Srivastava (1989); Deutsch & Journel (1998); Chilès & Delfiner (1999); Deutsch (2002).

From equation (I) the kriging variance sum is presented in equation (II):

$$KVS = \sum_{i=1}^{n} \sigma_{KO}^2(X_i) \qquad \text{II)}$$

The application of the KVS as an objective function is presented in equation (III):

$$\begin{aligned} &Minimize\ KVS(x) \\ &Subject\ to\ nZ(x), x = X, Y;\ X \subset [Xmax, Xmin];\ Y \subset [Ymax, Ymin] \end{aligned} \qquad \text{III)}$$

where n is the number of new drillholes Z(x) that should be located in the positions x; x is decomposed in two coordinates X for east and Y for north, both limited by the domain where the infill will be located, the superior and inferior limits of the east (Xmax, Xmin) and north (Ymax, Ymin). The lack of a third coordinate is related to the fact that the altitude of the new holes is constant and defined by the user as the highest value used in the block model.

6.1.2.2 simulated block variance (SBV)

The second objective function applied is dependent on simulating the data to guide the optimization. To each simulated block, there are (*l*) equiprobable values. The simulation method applied is the sequential Gaussian simulation (SGS). Each simulated point in the domain is given by equation (IV) (Olea 1999):

$$Z_{SGS}^{(l)}(x_0) = Z_{KS}^{*}(x_0) + \sigma_{ks}(x_0) * e \qquad \text{(IV)}$$

where: $Z_{SGS}^{(l)}(x_0)$ is the simulated value, of l-est realization, given by the simple kriging estimative $Z_{KS}^{*}(x_0)$ plus the estimation standard deviation $\sigma_{ks}(x_0)$ times the value e which is randomly obtained in the interval [-1,1] by Monte-Carlo simulation. For more details on SGS, the authors recommend: Deutsch & Journel (1998); Olea (1999); Deutsch (2002).

The simulated block variance is given by equation (V):

$$\sigma_{SGS}^{2}(x_0) = \frac{1}{n}\sum_{l=1}^{n}(Z_{SGS}^{(l)}(x_0) - \mu_{sgs}(x_0))^2 \quad \text{(V)}$$

where $\mu_{sgs}(x_0)$ is the mean of the simulated values at x_0.

The sum of simulated block variance (SBV) is computed as (equation VI):

$$SBV = \sum_{i=1}^{n}\sigma_{SGS}^{2}(x_i) \quad \text{(VI)}$$

The SBV as an objective function is presented in equation (VII)

Minimize SBV(x)
Subject to nZ(x), x = X, Y; X ⊂ [Xmax, Xmin]; Y ⊂ [Ymax, Ymin] VII)

Both objective functions use the same constraints to enable comparisons between functions. The only difference is in how the functions are calculated.

6.1.2.3 Optimization algorithm

The optimization algorithm implemented is based on simulated annealing (SA), a heuristic stochastic optimization method. SA is based on Metropolis *et al.* (1953) who presented the calculation of equations of state using a computer, and Kirkpatrick *et al.* (1983) who developed the SA optimization procedure. What differentiates the SA from other optimizations is the probability that allows the acceptance of worst results of objective function value as new optima. This SA characteristic is interesting for evading the occurrences of local optimal during the search. The acceptance of worst values is not completely random but is stochastic in nature, the algorithm decides for acceptance with a higher chance when the search is in the initial stage. At the beginning of the annealing procedure, the entropy of the system shall be the maximum and gradually decrease, emulated by what is named as cooling procedure in the algorithm. During the high entropy state, some particles may tend to move to the worst location overall, which is reproduced by the algorithm when worst configurations are taken as new optima. The acceptance is given by the temperature at the iteration and a control value that is obtained randomly from an interval predefined, therefore given the stochastic portion of the optimization.

Two stochastic components were implemented in the algorithm, the acceptance previously appointed, and the movement given between each iteration. At each iteration, the movement is dependent on a succession of 3 choices randomly made by the algorithm. The first choice is which drillhole shall be moved between the *n* new drillholes that will be located by the search. The second choice is which coordinate shall be moved, in this paper the choice is given between East and North coordinates, X and Y respectively. The final choice is the new position of the coordinate, given randomly in the complete interval given by the range of possible and acceptable coordinates of the domain. Those choices were implemented to maintain some sequential aspect in the drillhole configurations tested, evading the complete random approach of the research that can be a hindrance.

The cooling schedule utilized considers the total number of iterations and the present iteration to calculate its value. In equation (VIII) is presented the cooling schedule is applied.

$$[(n\text{iter}+1)/(\text{i}+1)]/n\text{iter} \qquad \text{(VIII)}$$

Where *n*iter is the number of iterations applied in the research and *i* is the present iteration. The limits imposed by this function tends to 1 and 0, i. e. the temperature value does not represent the control interval that is given between [1,0]. Regarding this limitation the temperature tends to the control limits when the number of iterations is higher, therefore it is recommended that the user gives a high enough iterations range to best develop the search, which results in a higher elapsed time of processing and probability of obtaining best results of the optimization.

The algorithm result is the configuration where the lowest objective function during the optimization and the function value was obtained. The only difference applied to each objective function is the theoretical variogram model adjusted to each, simulation, and kriging, which are different. To allow the full comparisons between methods the parametrization of the search was the same, and the number of new drillholes and interval of possible collar coordinates was the same, with constant altitude, azimuth, dip, and length of the drillholes.

6.1.2.4 Programming languages and software

The optimization algorithm applied was developed in Python™. No SA optimization library was used, the algorithm was developed by the authors. The libraries used were NumPy, random, math, DateTime, and subprocess. Subprocess was used to obtain the simulation and kriging results necessary for each iteration, used to spawn the Geostatistical Software Library (GSLIB) developed by Deutsch & Journel (1992).

The sgsim program of GSLIB performs SGS on the spatial data analyzed, its results are the realization in Gaussian or original distributions. There were applied 2 major differences to the original sgsim algorithm: maintaining the output in binary to attain a higher velocity of the processing; and the association of convex hull information to the simulation, limiting the blocks to be simulated inside a hull.

6.1.2.5 Synthetic data

The synthetic dataset developed by Takafuji (2015) and Takafuji *et al.* (2017) and mimics a cooper occurrence genetically related to a fault system. For each block that composes the mineralized body, there is information regarding the lithology and the copper percentage. The initial exploratory sampling was composed of 32 drillholes, all of them intercepting mineralized bodies. Those samples were preferentially taken perpendicular to the ore body, with an azimuth of 270 degrees and a dip of 45 degrees for each drillhole. Sampling location was taken randomly at the domain and shows some areas without information, and some portions with clustered drillholes, this proceeding was made to represent the need for infill to better represent the population data. The base map of the drillhole samples is presented in Figure P01-05.

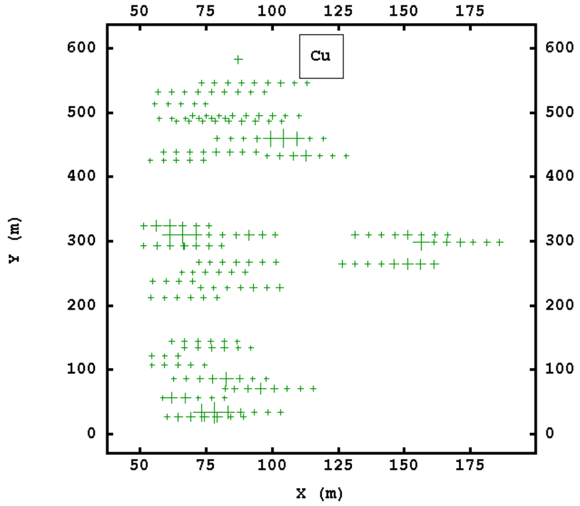


Figure P01-05 – Base map representing the samples' location. The size of the cross indicates the ore percentage.

6.1.3 Results and discussion

The first comparison between the objective functions applied regards the minimization attained by optimizing each function. To compare results, 1 to 15 new drillholes were considered, each one of them located after 1000 iterations, in Figure P01-06 the proportional objective function values are presented. The proportional objective function value is a proportion taken from the optimum result to the original sampling objective function. The KVS objective function achieves a smaller value by optimizing a single drillhole, with almost half of the original sampling KVS value. But the degree of the minimization is lower than the SBV objective function, for those reasons from 7 new drillholes onward SBV attained lower optimal values.

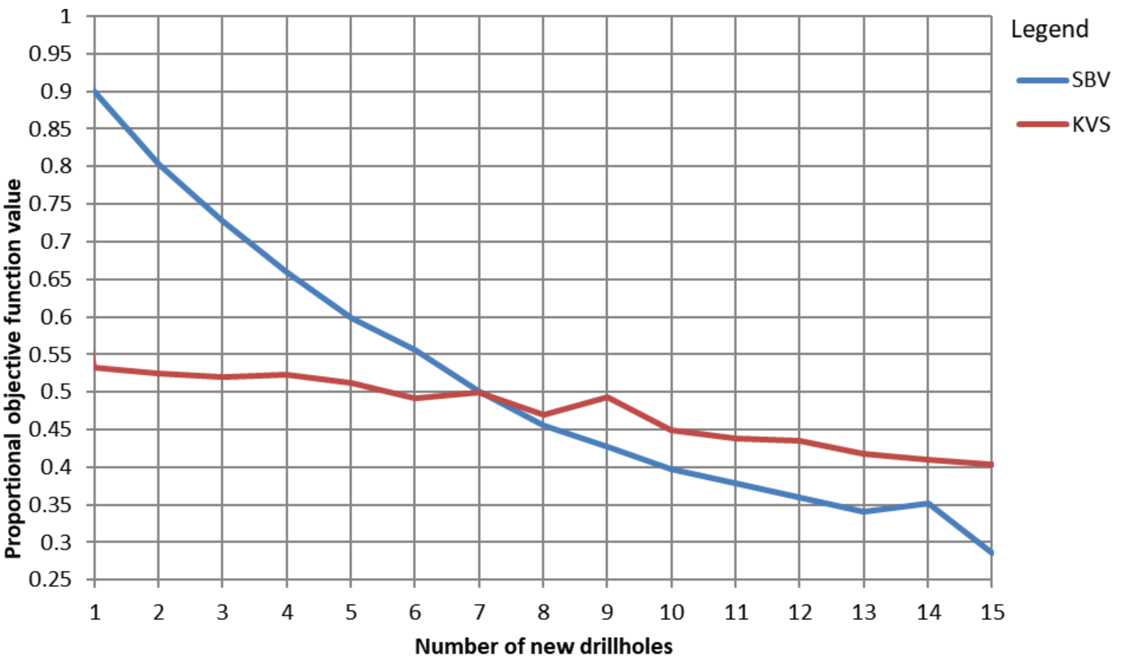


Figure P01 -06 – Minimization comparison of SBV and KVS under the same circumstances. An interval of 1 to 15 new drillholes was located, each new configuration attained after 1000 iterations.

The second comparison made regards the infill results. The comparisons were made with 2 infill configurations and 3 different iteration numbers, with a total of 6 scenarios applied to each objective function. After the drillholes were made, the results were compared to observe the representativity attained by each objective function and scenario. Each result is compared with the population distribution – from the synthetic deposit – statistically, by using the mean and standard deviation, and the Kolmogorov-Smirnov distance – the maximum distance between 2 distributions. The population mean and standard deviation are the base values that all optimized infill should approximate. The Kolmogorov-Smirnov distance is given zero to the population, as the distribution should have no difference in itself. The second base value is the exploratory sampling data, used to observe if representativity was enhanced in comparison to the initial data. Each scenario is repeated 10 times due to the stochastic approach given by the optimization method applied. To fully show the competence of the stochastical search of optima there is a need to repeat and assess if the range of results approximates the infilled sampling to the population.

In Figure P01-07 the mean of each distribution is compared. Figure P01- 07 A compares infill with 10 new drillholes located where each optimum attained was after 1000, 10000, or 50000 iterations. Figure P01-07 B presents the same iteration number but locates 15 new drillholes. The infill defined by SBV minimization did values closest to the population mean with both 10 and 15 infill drillholes. It is noticeable that with 15 new drillholes both objective functions presented a closer mean value regarding the population mean than with 10 new drillholes. The KVS minimization for 15 new drillholes has results proximal to each other, with a small dispersion in 3 of the 10 tests, a behavior that does not occur with 10 new drillholes nor when minimizing SBV. Regarding the number of iterations, it is not clear if the higher number of iterations attained the best results overall. The minimization of SBV to locate 10 new drillholes obtained the best results with both 1000 and 10000 iterations, with the best overall result with 10000 iterations. Locating 15 new drillholes by minimizing SBV the best number of iterations, regarding proximity to the population mean, is given by 10000 iterations, a value that surpasses the population mean, probably due to outlier occurrences. Minimizing KVS attained the best results with 50000 iterations for both infill configurations.

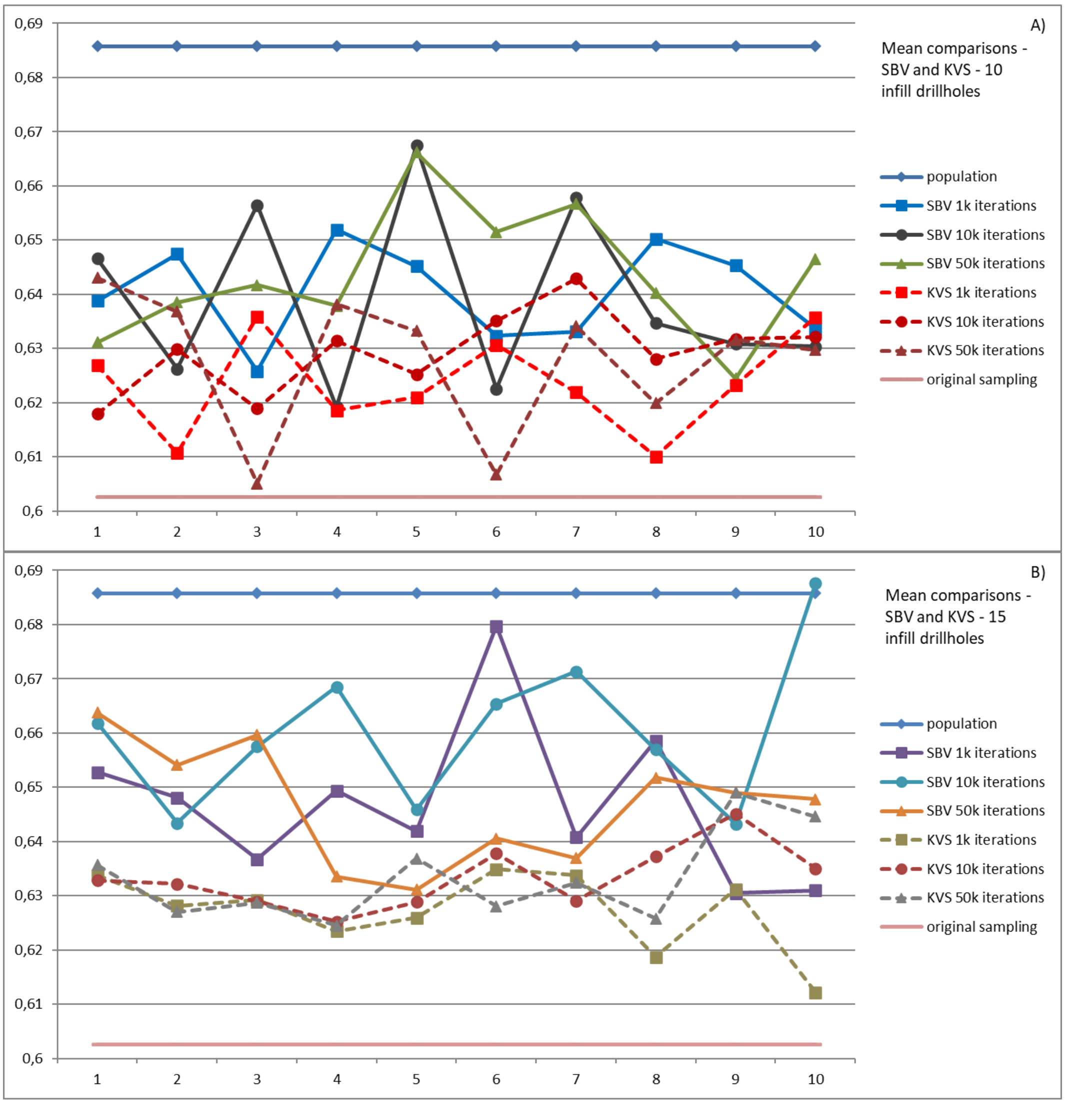


Figure P01 -07 – Infill comparisons to population and original sampling considering the mean. Comparisons with 10 (A) and 15 (B) new drillholes. The infill location was attained by either SBV or KVS minimization. The number of iterations necessary for each optimum is 1000, 10000, or 50000.

The second comparison was made between the standard deviation of the infill located in each scenario. Figure P01-08 presents the results of locating 10 (A) and 15 (B) new drillholes considering the same scenarios. The first remarkable difference to the mean is the more behavior that the standard deviation has between the different optimums obtained. Overall, the SBV returned the most approximated values to the population when locating 10 new drillholes, but in some cases, the KVS minimizations presented better or equal values to those obtained by the SBV function. The results considering only one scenario proves that there is not a trend in the results, for example minimizing SBV with 50000 iterations there are results closer to the population and the original data, the same can be said of 10000 iterations with the same function, or even minimizing KVS with 50000 iterations. It is important to notice that the infilled data tends to approximate more to the population standard deviation than its mean, this proves that the original standard deviation is greatly underrepresented, pointing out the occurrence of outliers in the population. Some of the optimums were able to attain outliers, which sometimes affected the representability negatively, as in the cases where the infilled data standard deviation surpasses the population standard deviation. When considering the location of 15 new drillholes there is again a remarkable difference between both objective functions, the KVS resulted in a lower dispersion of results that are overall worse than those attained by the minimization of SBV. It is interesting to notice that the same scenario – index-wise – that obtained the best result regarding the mean, repeats the feat when the standard deviation is considered. This does not disprove that the better option to optimize the location of 15 new drillholes is given by minimizing SBV with 1000 iterations, as the other results obtained by the same scenarios disperse closer to the population and the original sampling. If the range is overall considered the SBV function with 10000 iterations, when locating 15 new drillholes is the parametrization that comes closest to the population, overall.

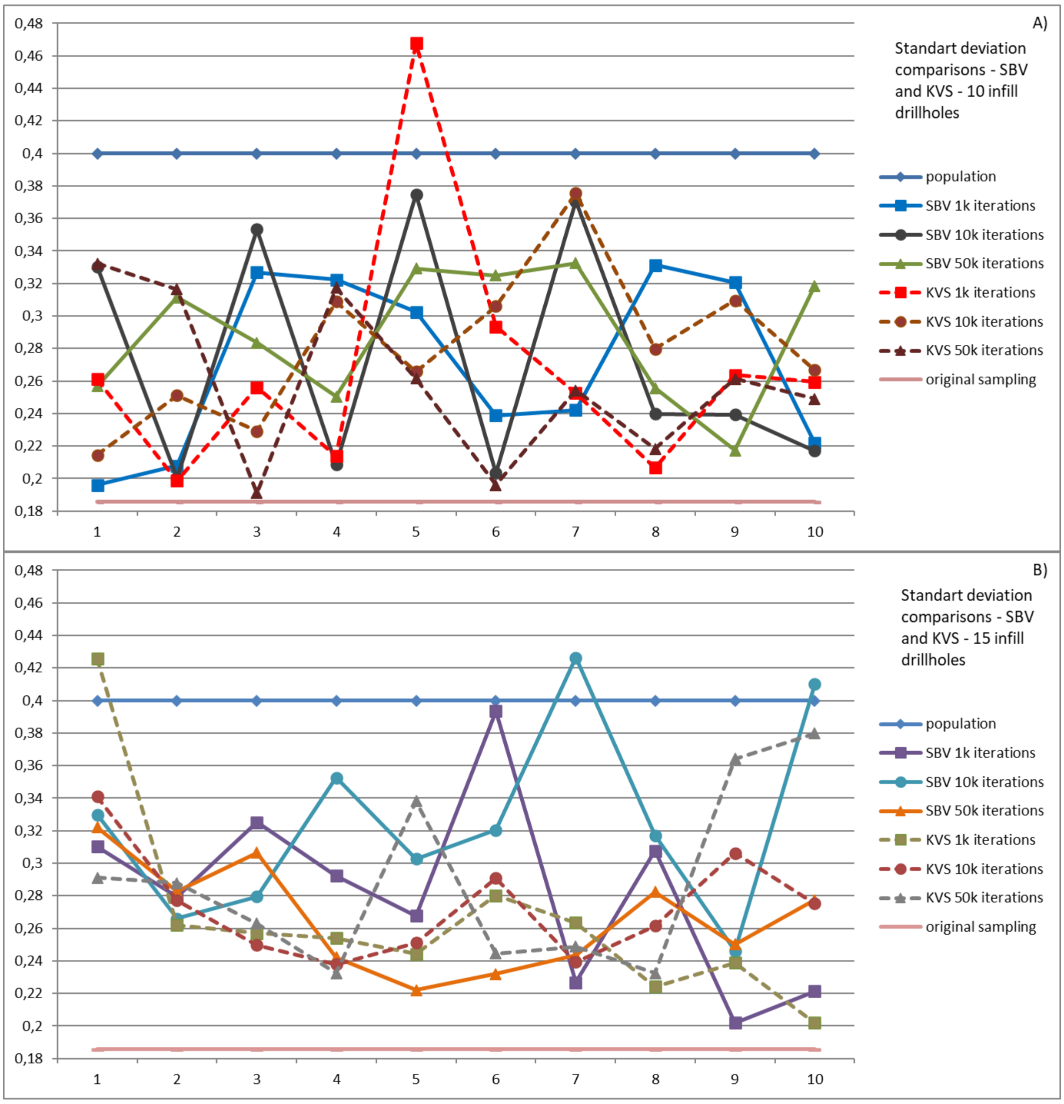


Figure P01-08 – Infill comparisons to population and original sampling considering the standard deviation. Comparisons with 10 (A) and 15 (B) new drillholes. The infill location was attained by either SBV or KVS minimization. The number of iterations necessary for each optimum is 1000, 10000, or 50000.

The last comparison was made considering the Kolmogorov-Smirnov distance to the populational distribution. In Figure P01-09 the comparisons are presented considering an infill of 10 (A) and 15 (B) new drillholes. The first difference between the comparisons using statistics and the Kolmogorov-Smirnov distance is the fact that both objective functions have a distinctive range of occurrence, and SBV distances are lower than those of KVS. This means that the minimization of SBV tends to provide a more proximate distribution regarding the population. Another great difference is the overlap between objective functions, that in the Kolmogorov-Smirnov distance tends to be minimal, each function separates from the other results. It is also noticeable that even infilling more drillholes, most of the results minimizing KVS did not reach the 10 new drillhole responses attained by the minimization of SBV. At last, it is interesting to notice that the ranges obtained by the Kolmogorov-Smirnov distance are smaller when considering the mean and especially the standard deviation. Considering minimization of SBV there is not a trend when the number of iterations is considered, with good results coming from both the lowest, 1000, and highest, 50000, iterations number.

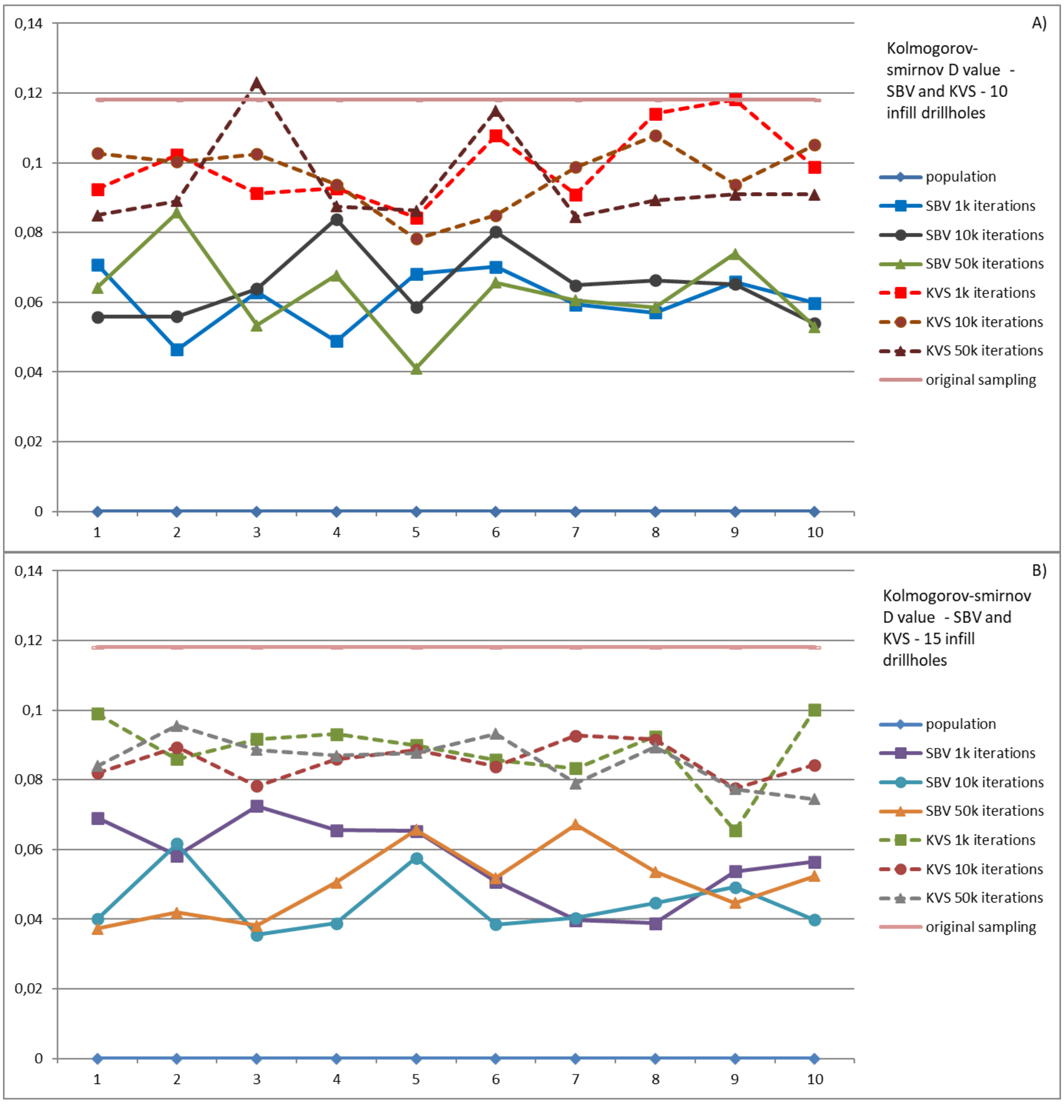


Figure P01-09 – Infill comparisons to population and original sampling considering the Komolrov-Smirnov distance. Comparisons with 10 (A) and 15 (B) new drillholes. The infill location was attained by either SBV or KVS minimization. The number of iterations necessary for each optimum is 1000, 10000, or 50000.

6.1.4 Final considerations and conclusions

Considering both the SBV and the KVS minimization as objective functions to locate infill drillholes proves to be functional in further the representability of the original data. The infill located by minimizing SBV tends to approximate the sampling distribution to the population, given the different proposed optimization scenarios applied. The competence of the SBV is evident considering the Kolmogorov-Smirnov distance, where the results of KVS presented a subpar, i. e. more proximal to the exploratory data than the population. Even so, the use of KVS is possible to guide the search for infill, as it can further the representability of the original sampling.

Utilizing both methods should not mean that different approaches could not be taken or explored. Using the SBV or KVS to help guide the infill location, or even present possible areas of the domain where a higher uncertainty occurs, especially when the SBV is used, is commendable. Therefore, it is interesting to explore different approaches or try to integrate the results of both objective functions.

6.2 Paper 2: objective function utilizing estimative and simulation data to optimize infill 3d dillhole samples

6.2.1 Introduction

The process of sampling is inherent to exploring and developing geoscientific knowledge. A good sampling campaign must be able to fully represent the variables under study. All sampling shall be made in a way to better represents the population, which raises the question of how to infer the uncertainty of the data regarding the population and what to do if that value is higher than a theoretical optimal threshold. One way to assess the uncertainty is by modeling the population through samples. In earth sciences a usual modeling tool is geostatistics. Geostatistical methods enable not only estimation or simulation of unsampled locations as it is possible to quantify and assess uncertainties associated with predicted values. Those uncertainties are kriging variance and simulation variance.

Several papers can be cited as examples of the use of kriging variance to locate infill sampling: Gershon (1987) presented four methods of optimizing drillhole location while minimizing the kriging variance. Wilde (2009) minimized the kriging variance sum for all estimated data. Soltani *et al.* (2011) minimized the average kriging variance to locate 3D infill drillholes utilizing a genetic algorithm. Fatehi *et al.* (2017) minimized the average kriging variance to guide the optimal infill configuration searched by the particle swarm optimization algorithm. Some researchers have applied different parameters with the kriging variance to further the search for infill sampling configurations as examples of such methods: Mohammadi *et al.* (2012) have applied a weighted average kriging variance, where the weight is the estimated grade. Soltani & Hezarkhani (2013) minimized the mean kriging variance while searching for drillholes that have the maximum length inside the orebody while having a minimum length outside. Silva & Boisvert (2013) minimized the kriging variance while maximizing ore tonnage. Safa & Soltani-Mohammadi (2018) utilized the minimization of the combined variance sum, composed of kriging variance times local variance, obtained as the weighted variance between the estimated value and the samples. Dutaut & Marcotte (2020) applies kriging estimate and variance with indicator kriging estimate to assess the probability of a drillhole being considered as an infill sample.

Geostatistical simulation methods results are always a set of equiprobable scenarios that reproduces sample statistics and spatial variance. In that way, areas with higher uncertainty can be assessed and used to guide infill optimization, or simulation scenarios can be utilized to compare possible sampling configurations in the domain.

Examples of geostatistical simulation applied to guide infill location are many and diverse. As examples can be cited: Goria *et al.* (2001) used simulation to represent the possible values the new information could have and assess the impact of the new data compared to the initial sampling values. Pilger *et al.* (2001) simulated the data and utilized the fact that at each simulated node there are *l* possible values to assess an uncertainty index through the interquartile value of the node's distribution. Pinheiro *et al.* (2017) used the average variance and the 95% interval width of the simulated nodes distribution to assess the uncertainty and guide the location of new drillholes using a simulated annealing algorithm. Dirkx & Dimitrakopoulos (2018) applied simulation to guide the search for drillhole location using the mineral classification of the blocks as a guide for the best sampling configuration. Zagré *et al.* (2018) utilized simulation to assess the probability of each node being of a given geological facies and from that value derive the Bernoulli random variable variance, used as the uncertainty index that guides the location of new samples.

The present work uses both estimate and simulation to define the location of infill drillholes. This approach is useful to guide the search for new samples considering the data value uncertainty and sampling density. The compost objective functions are compared to functions composed of just estimated values or simulated ones (single approach), therefore appointing the competence of the new proposed approach regarding the specific methods.

6.2.2 Objective

This paper aims to assess the competence of the proposed objective function and check if its use improves sample representativity when compared to the population. Tests and comparisons are held considering a synthetic orebody, which will be detailed in the next section and using the dataset of an actual iron ore mine as a study case. Comparisons are made considering the goodness of the objective function minimization, the elapsed time of each method (compost and single approach), and statistical comparisons between infilled sampling, original sampling, and population.

6.2.3 Materials and Methods

In this section used software, programming languages, and data sets will be presented. The objective functions, the optimization algorithm, the search methodology applied, and the constraints imposed on the search for the new collars will be shown and explained.

6.2.3.1 Programming languages and software

The optimization algorithm was developed in Python™, with no simulated annealing library being used. The libraries applied are limited to:

- NumPy is an algebra and array program used to read the original sampling and block data, generate the arrays of new collars and drillholes, and even calculate the objective functions;
- random to obtain the random values, from the choice between which collar to move and which coordinate of said collar, to calculate new coordinates for each collar;
- subprocess to use the geostatistics that was not from the Python™ library;
- DateTimeme to calculate the processing time of each iteration and the complete optimization;
- math to use the trigonometric function used to project each collar drillhole and obtain the barrel length and position of said drillhole;
- tempfile used to generate temporary files to proceed with the kriging/simulation at each iteration;
- os that uses the operating system interface and makes actions;
- scipy.spatial library was utilized to compute the convex hull of the samples over the block data.

The last Python™ library is relative to the geostatistical library applied, in this case being the GSLIB. GSLIB is a geostatistical library written by Deutsch & Journel (1992) in Fortran programming language. The GSLIB is robust, even with its age, and functionality, and as an open-source one can read and change the algorithms in the program, what was done in this work. Of all the programs in GSLIB only two, the kt3d and sgsim, were used. The kt3d is a tridimensional kriging program used to calculate kriging and kriging variance. The sgsim is used to calculate the sequential Gaussian simulation over the domain. Defaulted information on how to use GSLIB routines can be seen in Deutsch & Journel (1992 and 1998).

Two changes were made to the sgsim program, the first one is to the program to use a convex hull limiting the blocks that shall be simulated. The second modification was applied to the results, which were maintained in the binary form to decrease the computational time needed to simulate the block data. This gain of computer processing time is interesting since it can allow the optimization to have less elapsed time in the research, allowing its application as an analysis of infill data location.

The last program used was Studio RM® of Datamine Software. The program is used to assign the drillhole values of the synthetic deposit. The synthetic data is imported to Studio RM, and with the infilled collar configuration, it was possible to project the drillholes and associate the "real sample value" to each sample that composes the drillholes. As at all positions inside the domain the variable is known, it was possible to associate values to each sample of the infilled data.

6.2.3.2 Synthetic data and its initial sampling

The synthetic data was developed by Takafuji (2015) and Takafuji *et al.* (2017) to emulate a copper deposit with a complex disposition of lithologies where occurs faults and folds. The copper is genetically associated with the fault, with mineralized rocks occurring in quartzite, and some of the copper is disseminated at the embedding rocks.

The first sampling is composed of 32 drillholes most of them perpendicular to the geological fault, so the drillholes have a direction of 270° azimuth and 45° dip. The sampling schema was designed to present some clusters and gaps over the domain. Locate the sampling in that configuration was made to justify the importance of infill. Figure P02 - 01 presents the spatial map of the samples made in the synthetic data.

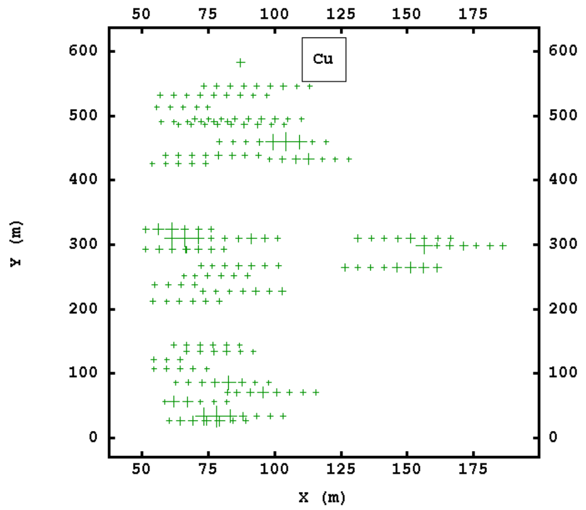


Figure P02-10 - Base map of the sampling made in the synthetic data. Lined-up crosses indicate the horizontal projection of samples along drillholes.

Each of the proposed objective functions shall produce different results of infilled collar configurations that shall be compared to define which result is the best in terms of representativity of the population searched. Knowing the population, it is possible to compare the infill schema reached by each optimization and then classify which one fared in further representativity of the population in the new samples. The objective is to rate all different results and point out which one performed best in population reproduction and what objective function was better to achieve that. This approach is practically impossible using real data, as the population is never attainable, but one can use the synthetic data to infer the competence of the method.

6.2.3.3 Real data set

To test the methodology developed is important to see how well it fares in a real case. The problem previously exposed is how to infer the infilled data quality if the population is unknown. One way to overcome this is by using different drilling campaigns. In that way, the Capanema mine drillhole data were used as initial sampling information possible of being infilled. And as a representation of the population, the blast drilling grade information was used. Figure P02-11 presents the base map of the Capanema mine drillholes.

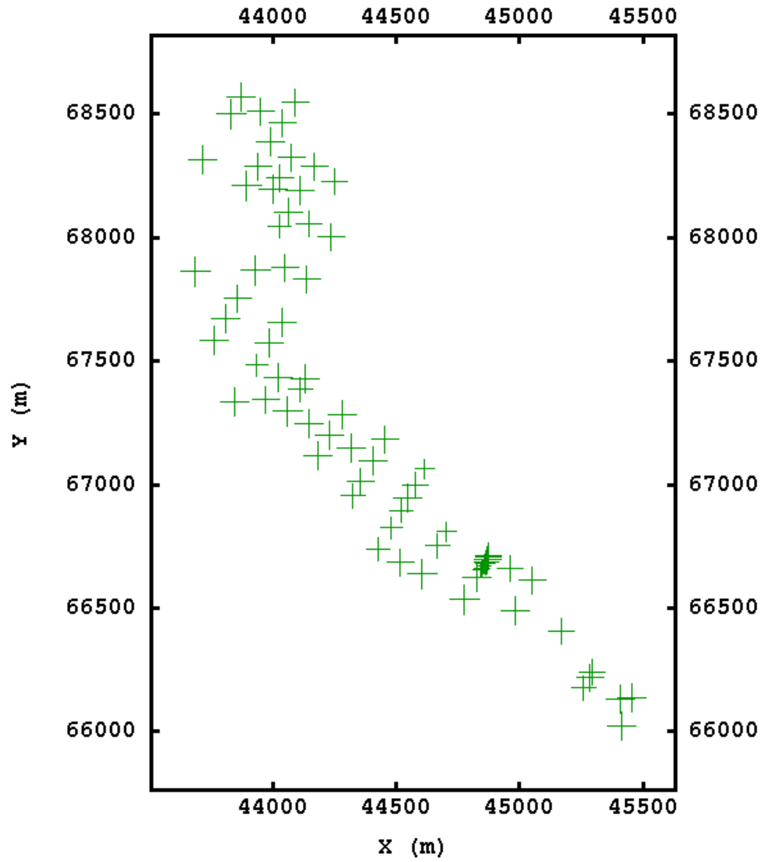


Figure P02-11 - Base map of the Capanema mine samples. The size of the cross indicates the iron grade.

Capanema is an iron ore mine in the so-called Quadrilatero Ferrífero of Brasil, in Minas Gerais state. The deposit is a folded BIF that was sampled by 69 drillholes (for more information about the geology of Capanema Mining the authors recommend Massahud & Viveiros 1983; Fonseca 1990; Rocha 1999). Those drillholes are mostly vertical, with one exception. Sampling design follows the ore body direction with the main axis with an azimuth of 135°.

6.2.3.4 Kriging variance sum (KVS)

The application of kriging is extensive in mining, details can be seen in several papers and books, such as: Matheron (1963), Matheron (1971), Isaaks & Srivastava (1989), Goovaerts (1997), Armstrong (1998), Deutsch & Journel (1998), and Olea (1999).

Unsampled points of the domain are estimated by kriging, a weighted mean, where the weights are calculated through the spatial variance between the samples and the estimated point. From the weights and spatial variance between samples and estimated points, it is possible to compute the kriging variance associated with the kriged value. The kriging variance does not consider the sampled grades, considering only the samples spatial distribution.

The authors have chosen to use ordinary kriging, and the ordinary kriging variance, as they are the main technics applied in mining. Ordinary kriging variance can be seen in equation (I):

$$\sigma_{OK}^2(X_0) = \sum_{i=1}^{n} \lambda_i * \gamma(X_i, X_0) + \mu \qquad \text{(I)}$$

where λ_i is the weights obtained solving the kriging system; $\gamma(X_i, X_0)$ is the spatial variance between the sample data (X_i) and estimated point (X_0); μ is the Lagrange multiplier added to guarantee the constrain required by ordinary kriging, where the sum of weights shall be 1 ($\sum_{i=1}^{n} \lambda_i = 1$).

The kriging variance is used as an objective function in the present paper as the sum of all kriged variance obtained in the domain to appoint a global indicator of areas with low sampling density. Kriging variance sum is presented in equation (II):

$$KVS = \sum_{i=1}^{n} \sigma_{OK}^2(X_i) \qquad \text{(II)}$$

Each new infill sampling configuration provides actualized values of kriging variance to different estimated points; in that way, it is necessary to re-estimate the deposit to obtain the new value of kriging variance that then shall be summed. The objective is to minimize the kriging variance sum computed at each iteration to attain the optimal infill drillhole configurations. To formalize the kriging variance sum as an objective function the formula applied in the research is presented in equation (III):

Minimize KVS(x)
Subject to nZ(x), x = X, Y, Z; X ⊂ [Xmax, Xmin]; Y ⊂ [Ymax, Ymin]; Z = Zmax (III)

where n is the number of new drillholes located by each optimization, it shall be the same at the origin and end of the search for the best infill configuration; Z(x) is the sample located, that shall be limited by the drillhole collar it is represented by; the collar value is given by the X and Y coordinates, limited by any possible value between the maximum and minimum coordinate values defined by the user; the Z collar coordinate is fixed as the maximum Z value of the block model were estimative where made.

After the possible infill location is considered a new estimation and its variance is computed, which is summed and compared to the objective function of the previous iteration value. If the new value is lower than the previously attained objective function value, the new infill configuration is considered the best, and the optimal value is actualized.

6.2.3.5 Simulated block variance (SBV)

Geostatistical simulation differs from kriging in different factors, some of those are the number of models computed, the statistics and geostatistics of those models, and others. The simulation shall provide *l* equiprobable realizations of the domain. As simulation results must reproduce original sampling data, not resulting in a smoothed distribution if compared with the original sampling distribution. For more details on geostatistical simulation please refer to: Deutsch & Journel (1998), Chilès & Delfiner (1999), Olea (1999), and Deutsch (2002).

In the present paper sequential Gaussian simulation is applied in the objective function. To simulate each node of the block data shall be visited at random in the realization, the block is simulated by the function shown in equation (IV) (modified and adapted from Olea 1999 and Deutsch 2002) until all nodes are visited. This procedure is repeated *l* times utilizing a new random path for each realization (Deutsch & Journel 1998; Deutsch 2002).

$$Z^{l}_{SGS}(x_0) = Z^{*}_{SK}(x_0) + R(x_0) \qquad \text{(IV)}$$

$Z^{*}_{KS}(x_0)$ is the simple kriging estimator; the random residual ($R(x_0)$) results from a Monte Carlo simulation of a distribution given by mean 0 and variance $\sigma^2_{SK}(x_0)$ (Deutsch 2002); as the distribution is Gaussian, the simulated value is randomly drawn from the interval given by the know mean, $Z^{*}_{KS}(x_0)$, and variance $\sigma^2_{SK}(x_0)$ (Deutsch 2002).

As each node has *l* simulated values it is possible to calculate the mean and variance for them. The authors opted to use the node variance as local uncertainty of the simulation: Equation (V) is the node variance function:

$$\sigma^2_{SGS}(x_0) = \frac{1}{n}\sum_{l=1}^{n}(Z^l_{SGS}(x_0) - \bar{Z}^l_{SGS}(x_0))^2 \quad \text{(V)}$$

$\bar{Z}^l_{SGS}(x_0)$is the node distribution mean.

To assess the global indication of uncertainty derived from the simulation model it is proposed the sum of all nodes variance, which is presented in equation (VI), the simulated block variance (SBV):

$$SBV = \sum_{i=1}^{n}\sigma^2_{SGS}(x_i) \quad \text{(VI)}$$

The value of SBV is the proposed objective function that the search algorithm shall minimize, as formalized in equation (VII):

Minimize SBV(x)
Subject to nZ(x), x = X, Y, Z; X ⊂ [Xmax, Xmin]; Y ⊂ [Ymax, Ymin]; Z = Zmax (VII)

6.2.3.6 Simulated block coefficient of variation (SBCV)

Another proposed objective function based on the simulated model is the minimization of the simulated block coefficient of variation (SBCV). This objective function follows the same procedure as the SBV differing from that by utilizing the coefficient of variation. The coefficient of variation function applied to each simulated block is presented in equation (VIII).

$$CV_{SGS}(x_0) = \frac{\sigma_{SGS}(x_0)}{\bar{Z}^l_{SGS}(x_0)} \quad \text{(VIII)}$$

$\sigma_{SGS}(x_0)$ is the simulated node standard deviation value.

The same proposed procedure is followed to apply the coefficient of variation as a global indicator of uncertainty, the sum of all simulated nodes' coefficient of variation. At equation (IX) is presented the proposed objective function, simulated block coefficient of variation (SBVC), and at equation (X) the formal objective function:

$$SBCV = \sum_{i=1}^{n} CV_{SGS}(x_0) \tag{IX}$$

Minimize SBCV(x)
Subject to nZ(x), x = X, Y, Z; X ⊂ [Xmax, Xmin]; Y ⊂ [Ymax, Ymin]; Z = Zmax (X)

The objective function SBCV is formalized equally to the previous objective functions presented, but what sets it apart from both is the result of minimizing the SBCV. A remarkable aspect of utilizing the coefficient of variation and not the variance as an objective function is the consideration of the mean value of realizations at each simulated block.

6.2.3.7 Simulated block variance and kriging variance (SBVKV)

A proposed compost objective function minimizes the sum of the simulated block variance and kriging variance or SBVKV. To calculate this objective function value, the domain shall be kriged and simulated at each iteration, and each simulated block variance and kriging variance are calculated and summed to give the uncertainty index that guides the search for infill collars. The procedure follows what was previously presented, and the difference is how the objective function is computed. To better represent the problem, each portion of the function shall receive a weight. The weights represent the importance given by each factor over the final objective function value. The SBVKV is presented in equation (XI), while in equation (XII), the formal objective function is presented.

$$SBVKV = \sum_{i=1}^{n} w_S * \sigma^2_{SGS}(x_i) + w_K * \sigma^2_{OK}(X_i) \tag{XI}$$

Minimize SBVKV(x)
Subject to nZ(x), x = X, Y, Z; X ⊂ [Xmax, Xmin]; Y ⊂ [Ymax, Ymin]; Z = Zmax; (XII)

$$w_S + w_K = 1$$

w_S is the weight associated with simulation, and w_K is the weight associated with kriging. The sum of weights shall be 1, as each weight is a representation of the proportion given to both simulation and kriging during the search for the optimal infill collar configuration. The user can assign which parameters are more important, or if both shall be considered with the same importance during the search.

6.2.3.8 Simulated block coefficient of variation and kriging variance (SBCVKV)

The last objective function presented is another compost function, the sum of simulated block coefficient of variation and kriging variance (SBCVKV). This objective function follows the same procedure as SBVKV, but instead of simulation variance, it uses the simulation coefficient of variation. Equations (XIII) and (XIV) present the SBCVKV function and the formal objective function proposed:

$$SBCVKV = \sum_{i=1}^{n} w_S * CV_{SGS}(x_0) + w_K * \sigma^2_{OK}(X_i) \qquad \text{(XIII)}$$

Minimize SBCVKV(x)
Subject to nZ(x), x = X, Y, Z; X ⊂ [Xmax, Xmin]; Y ⊂ [Ymax, Ymin]; Z = Zmax; (XIV)

$$w_S + w_K = 1$$

The same approach is given to the SBCVKV and SBVKV, as they use the same idea, considering both simulation and estimate when quantifying uncertainty. The difference is the fact that, once again, the SBCVKV can consider not only uncertainty when minimized, but the mean value of the realizations, which can be maximized during the search, the same point made when the SBCV is considered.

6.2.3.9 Optimization algorithm

The search for the optimal infill collar is realized by an algorithm developed by the authors based on the simulated annealing optimization method. The simulated annealing was developed by Kirkpatrick *et al.* (1983) based on the work of Metropolis *et al.* in 1953. Metropolis *et al.* (1953) presented a methodology to compute equations of the state where the system molecules interact between themselves. Kirkpatrick *et al.* (1983) borrowed this concept to apply to optimization problems, using the concept of annealing to illustrate the methodology proposed. At the initial state of the annealing procedure, the temperature is the highest and is minimized during the procedure. The temperature controls the acceptance of the new configuration obtained by the search, in the sense of adding a probability of accepting a worse configuration than the previous optima regarded. This mechanism set apart the simulated annealing from other known optimization methodologies, the controlled acceptance of worse scenarios that can be used to scape local optima that would hinder the search algorithm. At the start of the annealing, the probability of acceptance of worse optima is higher than at the end of the optimization, considering that system entropy is highest at the start and decreases along the search. Facing a configuration that resulted in a worse value, the control value is randomly drawn from an interval between [0,1] and compared with the temperature at the current iteration. If the temperature is higher than the control obtained, the worse result and its configuration are accepted as the new optima, and the search continues.

The definition of the temperature is given by a cooling schedule that shall be considered by the algorithm. There are different applications and forms to define the cooling schedule to choose from, each one is dependent on the problem at hand. Implanted in the cooling schedule is the stop criteria, given by where the differences of each configuration tend to be constant and so to attain system stability. In the present work, this procedure was not considered in the cooling schedule or the optimization, with the minimum temperature, temperature equal to 0, obtained at the last iteration. For that purpose, the schedule is given by the equation (XV):

$$[(n\text{iter}+1)/(\text{i}+1)]/n\text{iter} \qquad \text{(XV)}$$

*n*iter is the number of iterations to be made by the algorithm, and i is the iterations of the search. The definition of the schedule proposed is to minimize the temperature following an exponential approach, with a higher probability at the start, that rapidly decreases at the initial iterations. The temperature interval is not [0,1], equal to the control, but tends to those values the highest the number of iterations used. The implemented approach has the function of forcing the user to set a high number of iterations, which is interesting as a random approach to guarantee a better result. As simulated annealing is a heuristic optimization method, there are no guarantees that the best result overall is obtained, but a good enough result is attained in a relatively smaller time and computational demands.

The algorithm utilized in this work has other random processes, besides the acceptance of worse configurations. The algorithm initiates by obtaining *n* collars with random X and Y coordinates, limited to the user-defined domain. The number of collars is maintained until the end of the optimization. The algorithm projects the drillholes based on the collar location with the total length given by the user. Each drillhole is then partitioned in the barrel length desired, and each one of those is considered a sample that shall receive a value drawn from the closest block. If the block value was simulated, the sample value is the mean of the nearest block realizations. If kriging was made, the value is the kriging block value. With the infilled data defined, a new kriging/simulation (or both) is calculated to obtain the proposed configuration objective function value. After the sequence ends, the optimization initiates. After the temperature is calculated, the first random decision is made, which collar shall be moved in the present iteration. This guarantees the sequential procedure as the movement is not attaining completely random collar values at each iteration, with only one of those to be moved in the iteration. Selecting which collar, then the coordinate to be changed is randomly defined too, between X or Y value of the selected collar. The variation of each configuration is given in a one-dimension approach. The new coordinate value is drawn randomly inside the limits of the domain. Having the new collar position, the drillhole is produced in the way previously described, with a small change regarding the values assigned with each sample. If the new collar does not have at least one new sample with a value associated, a new collar is drawn randomly in the domain until a drillhole with at least one value is obtained. Once the new drillhole is obtained, the objective function is calculated in the same way previously explained. The objective function value is then compared to the optimum, being the new value smaller than the optimum, the new value and its configuration are defined as the new optima. If the new value is higher than the optima, the control is calculated and compared with the temperature. If the temperature is smaller than the control, the iteration ends. Being the control smaller than the temperature, the worse objective function value and its collar configuration are considered the new optima and the iteration ends. After there are no more iterations to be done, the optimization finalizes with the algorithm returning the optimal objective function value, the smallest overall value attained during the search, and the collar configuration that obtained that value. A flowchart of how the algorithm proceeds with the optimization is presented in Figure P02-12.

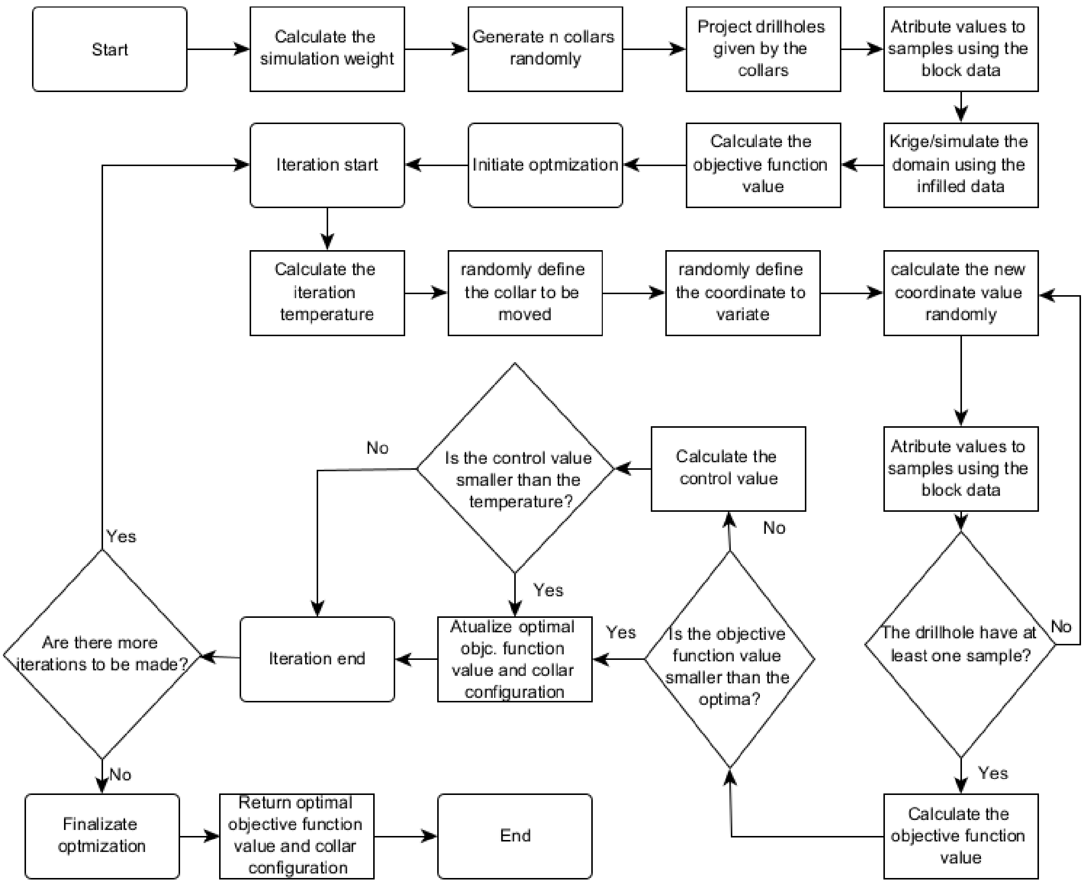


Figure P02-12 – Flowchart of the algorithm implemented to optimize the infill collar configuration.

6.2.3.10 Imposed optimization constraints

To guarantee that each objective function optimization is treated equally some constraints were imposed in the search for infilled collar configurations. The first one regards the direction of the drillholes fixed perpendicular to the synthetic data fault and vertical in the Capanema drillholes infill. The second one is to fix the z coordinate of the collar as the highest z value from the data set. All drillholes have the same total length, given as the domain direction total length, taken at each case, synthetic and Capanema, drillhole direction. Each new probable collar must have at least one non-empty sample composing it. The number of new collars located is maintained from the start till the end of the optimization. The variogram model is the same over the optimization iterations, and the domain data is considered stationary. All simulations and kriging were restricted to the boundaries of the convex hull computed on data sets. No changes are made in the block data parametrization, which shall be maintained during the search and be informed by the user. The random seed of the simulation is maintained during the optimization and must be informed by the user. Data value limits must be respected, with the maximum and minimum values informed by the user.

6.2.4 Results and discussion

6.2.4.1 Synthetic data

The first analysis considered objective function minimization assessed through the competence of the algorithm to optimize collar infill. To compare each objective function an interval of 1 to 15 new collars were located after 1000 iterations. Figure P02-13 is a graph presenting the minimal value obtained by each objective function at the end of the optimization procedure, the objective function value is a ratio given to the initial function value, taken directly from the original sampling with no infill done. The first remarkable difference is how a single new drillhole changes the objective function value, with the reduction of KVS to almost half the initial value. The objective functions that only depend on the simulation do not have a sudden drop in the objective function, with the SBV reducing 10% of the initial value. The difference between the simulation and kriged-based functions affects the compost objective functions too, while in the case of SBVKV the original value was reduced to over 25% of the original sample value, and the SBCVKV function was the closest to the KVS over the objective function value, to the point the results are superposed to the KVS optimization for most of the analyzed interval. That can be explained by the fact that the SBCV produced the worst, i.e., highest minimization optimal, systematically. The superposition of KVS over the SBCV side of the SBCVKV objective function can be prejudicious to the results obtained by the optimization, but as the only comparisons made are between the objective function minimizations, it is not possible to assess the competence of either objective function over another. The only noticeable fact is that the KVS regardless of having a higher minimization its competence is not kept when more drillholes are located, decreasing less when compared with the SBCV. From the beginning to the end of the interval, the decrease of KVS was almost 15%, while in the case of the SBV objective function it was greater than 60%. Therefore, the degree of minimization of the SBV objective function is the highest, obtaining the smaller values systematically after locating 8 infill collars at the original sample data. The SBVKV presents an interesting result when compared to the SBV and KVS objective functions, which appear to approximate to the point of even superimpose over the SBV, the addition of KVS decreases the degree of minimization of the compost function. The last comparison considering the minimization is over the optimization total processing elapsed time for each objective function, which is presented in Table P02-02. The KVS has a smaller processing time by far, however, the compost objective functions do not demand much more time to optimize.

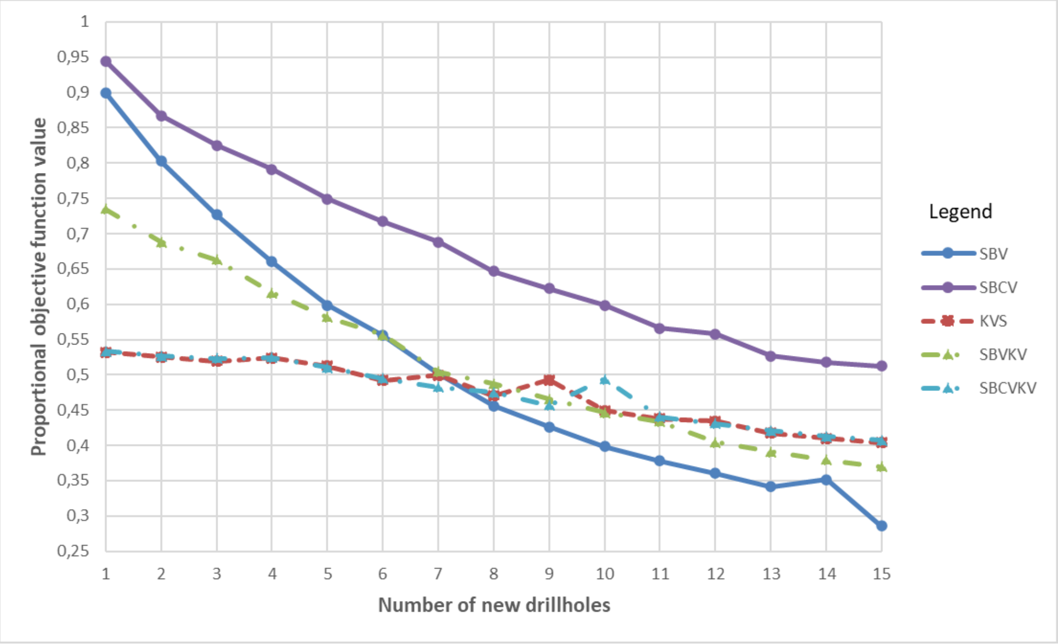


Figure P02-13 – Comparisons between each objective function optimization, obtained after 1000 iterations, under the same parametrizations. The objective function value is a ratio between the initial value, taken directly from the original sampling model (kriged and/or simulated), and the optimal value obtained with the number of collars.

Table P02-03 – Mean elapsed processing time to optimize, after 1000 iterations, each objective function proposed.

	SBV	SBCV	KVS	SBVKV	SBCVKV
elapsed time	00:26:05	00:25:46	00:08:25	00:32:25	00:32:34

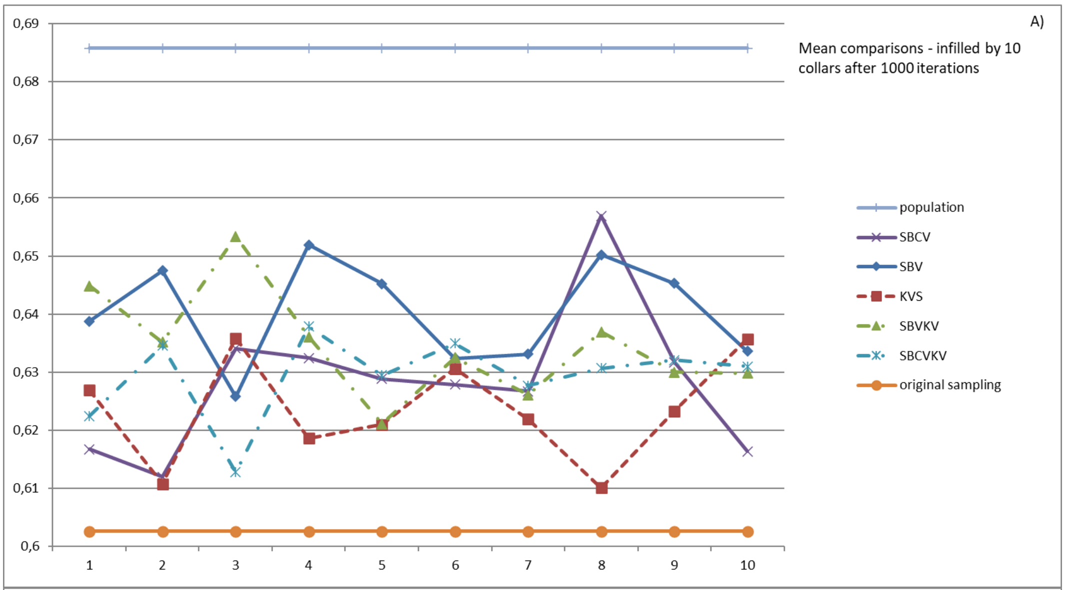
The second comparison considered the results of evaluating the optimized infilled samples. To better compare the results two scenarios were proposed: first 2 different numbers of infilled drillholes, 10 or 15 new drillholes; second, 3 iteration numbers to optimize the collar locations, respectively 1000, 10000, and 50000 iterations. Each objective function was optimized by both scenarios proposed 10 times. The use of more than one optimization for scenarios is important for the same reason as using a high iteration value, as the method is stochastic in nature, only one result per scenario could either provide a high array of possible results, therefore the higher the number of procedures more reliable are the comparisons of the results.

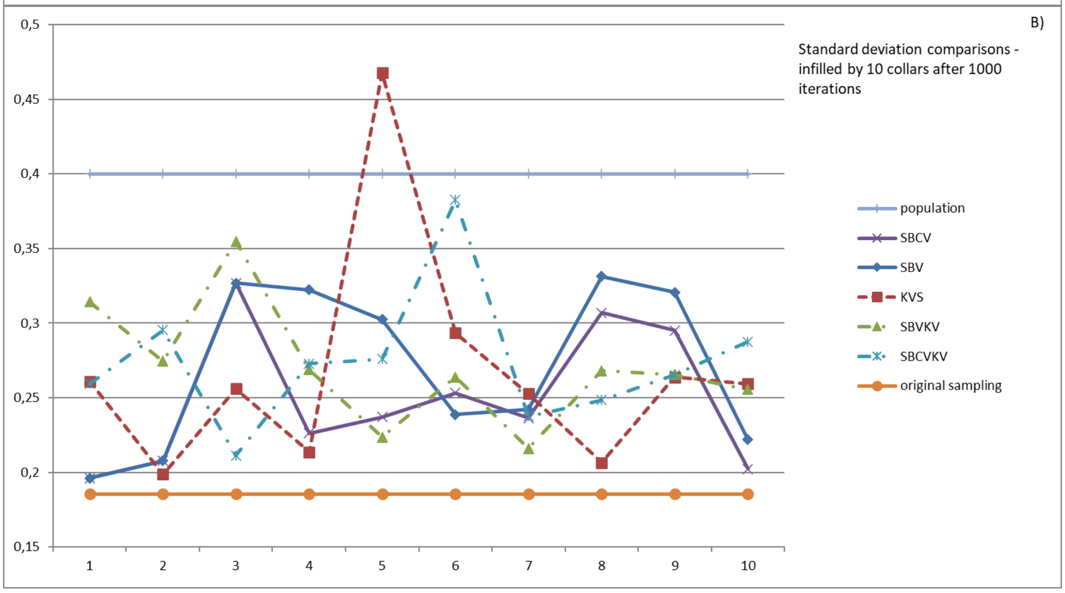
The first scenario applied to compare the results considered 10 infill collars optimized by 1000 iterations. In Figure P02-14 the comparisons considering the mean (A), standard deviation (B), and Kolmogorov-Smirnov distance (C) are presented. The comparisons are made between each objective function's optimal infill collar configuration, the population, and the original sample values. In the Kolmogorov-Smirnov distance, the values were taken considering the population distribution. The population value on the graph is equal to zero because it is measuring the population to itself. The comparisons are presented in Table P02-03 focusing on the whole interval of the 10 results obtained and the best overall value for each objective function.

The objective function that obtained the closest value to the populational mean was the SBCV minimization. However, this objective function has a higher value of interval, as seen in Table P02-03. Therefore, the best function would be the SBV which has a smaller interval with a maximum value close to the best. The worse results regarding the mean were attained by minimizing KVS, which returned the smaller best mean value. Even if the KVS and SBV minimization have the same interval, which is small, the fact that SBV has a higher maximum proves that this objective function fared better in furthering the mean representability.

The SBCVKV minimization obtained the closest standard deviation to the populational one. However, its interval is one of the highest, which shows that this function is not as accurate as the others in furthering the standard deviation representativity. This is the same occurrence that happens when minimizing the KVS function, which has a standard deviation higher than the populational but with the highest interval. The best function in furthering representing the populational standard deviation was the SBVKV, considering its best value and interval.

Regarding the Kolmogorov-Smirnov distance, the SBCV minimization has the smallest value overall. However, it has the highest interval too. The best values were obtained by minimizing the SBV, with a small interval and minimum. The worse results were those of KVS minimization.


A)
Mean comparisons - infilled by 10 collars after 1000 iterations
0,69
0,68
0,67
0,66
0,65
0,64
0,63
0,62
0,61
0,6
1 2 3 4 5 6 7 8 9 10
population
SBCV
SBV
KVS
SBVKV
SBCVKV
original sampling


B)
Standard deviation comparisons - infilled by 10 collars after 1000 iterations
0,5
0,45
0,4
0,35
0,3
0,25
0,2
0,15
1 2 3 4 5 6 7 8 9 10
population
SBCV
SBV
KVS
SBVKV
SBCVKV
original sampling

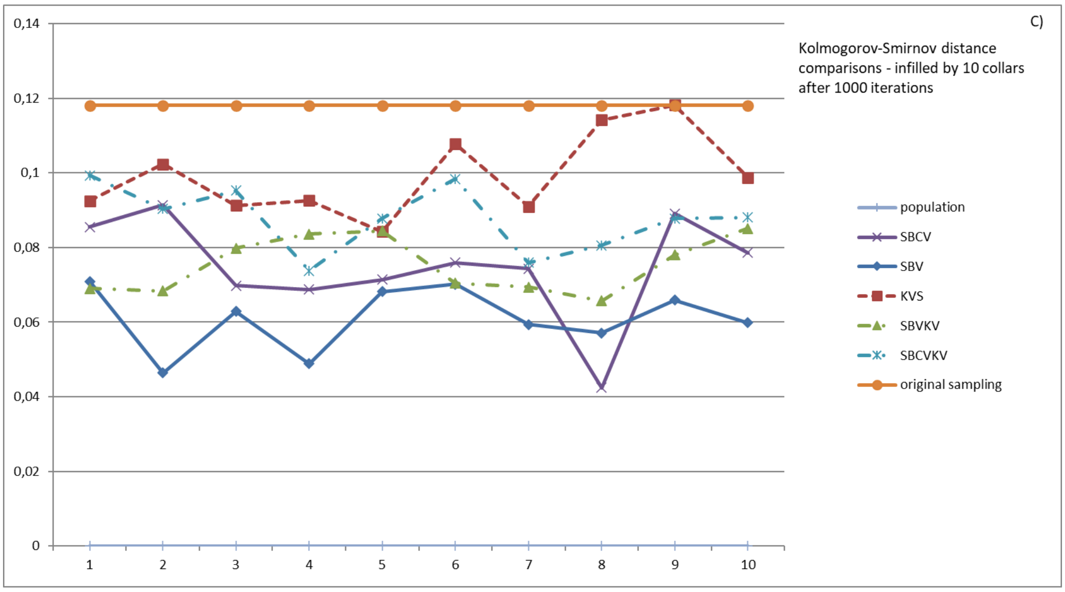
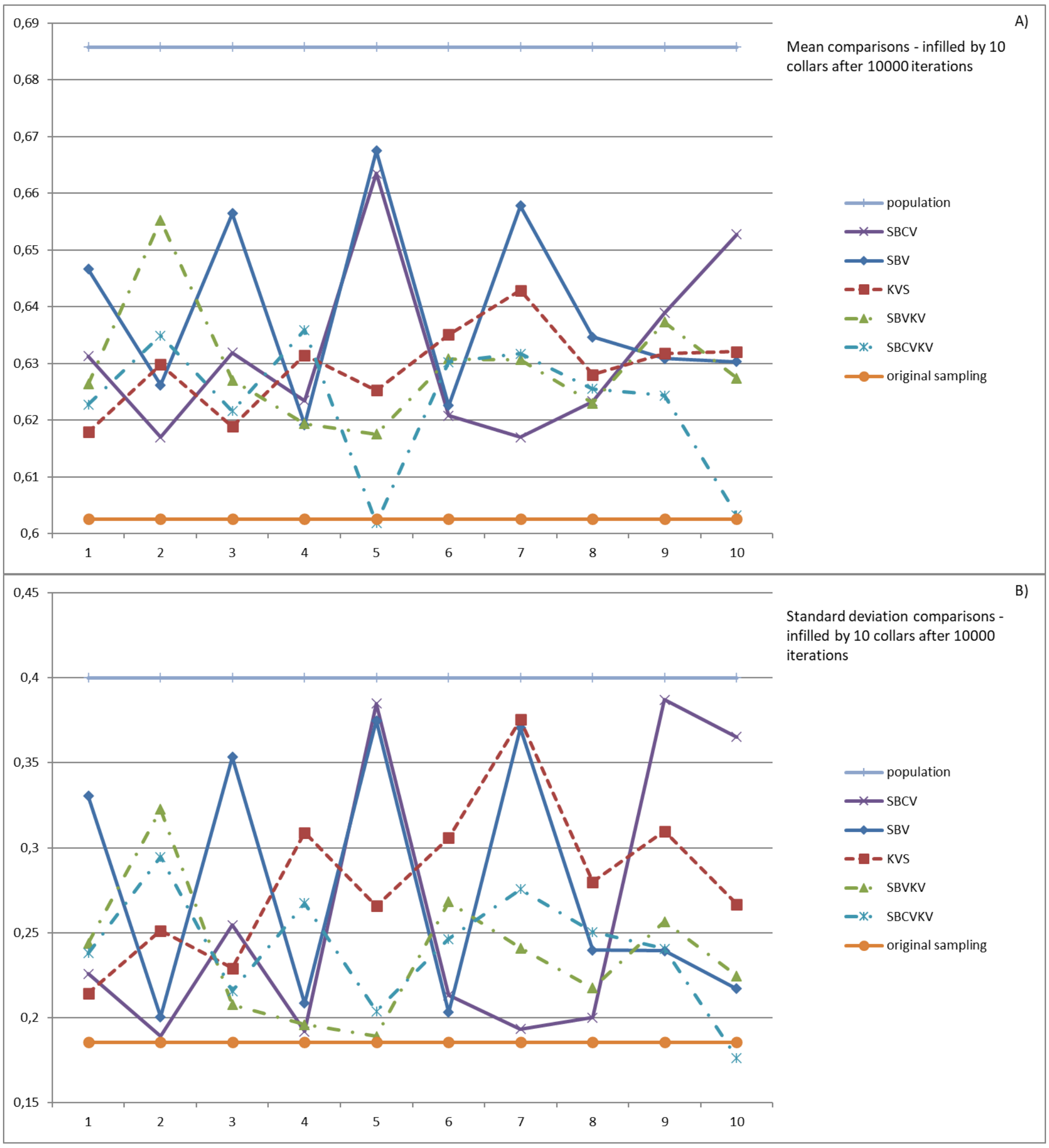


Figure P02-14 – Infill optimization results comparisons. The values compared are mean (A), standard deviation (B), and Kolmogorov-Smirnov distance (C). The scenario considered is to locate 10 infill collars optimized by 1000 iterations. The Kolmogorov-Smirnov distance value is obtained by comparing each distribution to the population, hence the population being zero at (C).

Table P02-04 – Best parameter value, and its distribution spread, for each objective function applied to infill 10 collars, optimized by 1000 iterations. The distribution spread was obtained by subtracting the minimum value from the maximum, considering the 10 optimizations applied with each objective function.

	SBV	SBCV	KVS	SBVKV	SBCVKV
Distribution spread (mean)	0.026	0.045	0.026	0.032	0.025
Distribution spread (std. dev.)	0.135	0.131	0.269	0.139	0.171
Distribution spread (Kolmo.-Smir. dist.)	0.024	0.049	0.034	0.019	0.026
Maximum mean	0.652	0.657	0.636	0.653	0.638
Maximum std. dev.	0.331	0.327	0.468	0.355	0.382
Minimum Kolmogorov-Smirnov dist.	0.046	0.042	0.084	0.066	0.074

The second scenario infilled 10 new drillholes after 10000 iterations. The results of those optimizations are shown in Figure P02-15 and further resumed in Table P02-04. When the mean is considered (Figure P02-15 A), the minimization of SBV obtained the best mean representability regarding the population. However, that function is the one with the highest spread for the mean. Therefore, the SBVKV produced better optima when considering its maximum and spread. The worse function regarding the mean was the SBCVKV, with the smallest maximum. Regarding the standard deviation (Figure P02-15 B) the function with the best representability was the KVS, with a maximum equal to that of the SBV but in a smaller spread. Even if the SBCV has a maximum value closer to the populational standard deviation, it has obtained the highest interval. The function with the worse results in terms of standard deviation representativity is the SBCVKV. The last comparison, the Kolmogorov-Smirnov distance (Figure P02-15 C), shows the SBCV has the closest distribution to the populational one, with the smallest distance and a feasible spread. Again, the SBCVKV has worse results to represent the population than the other functions.



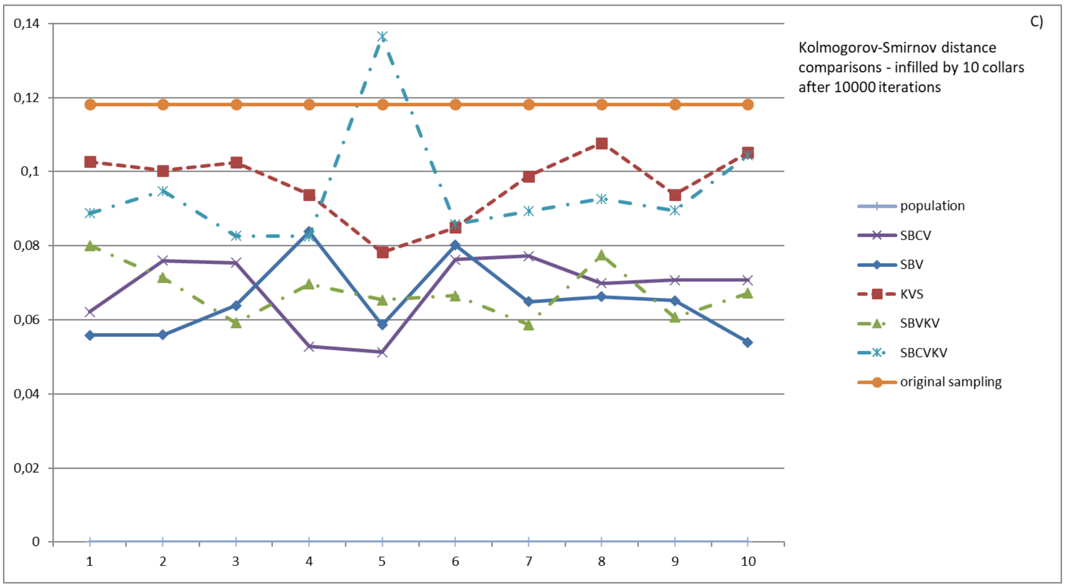
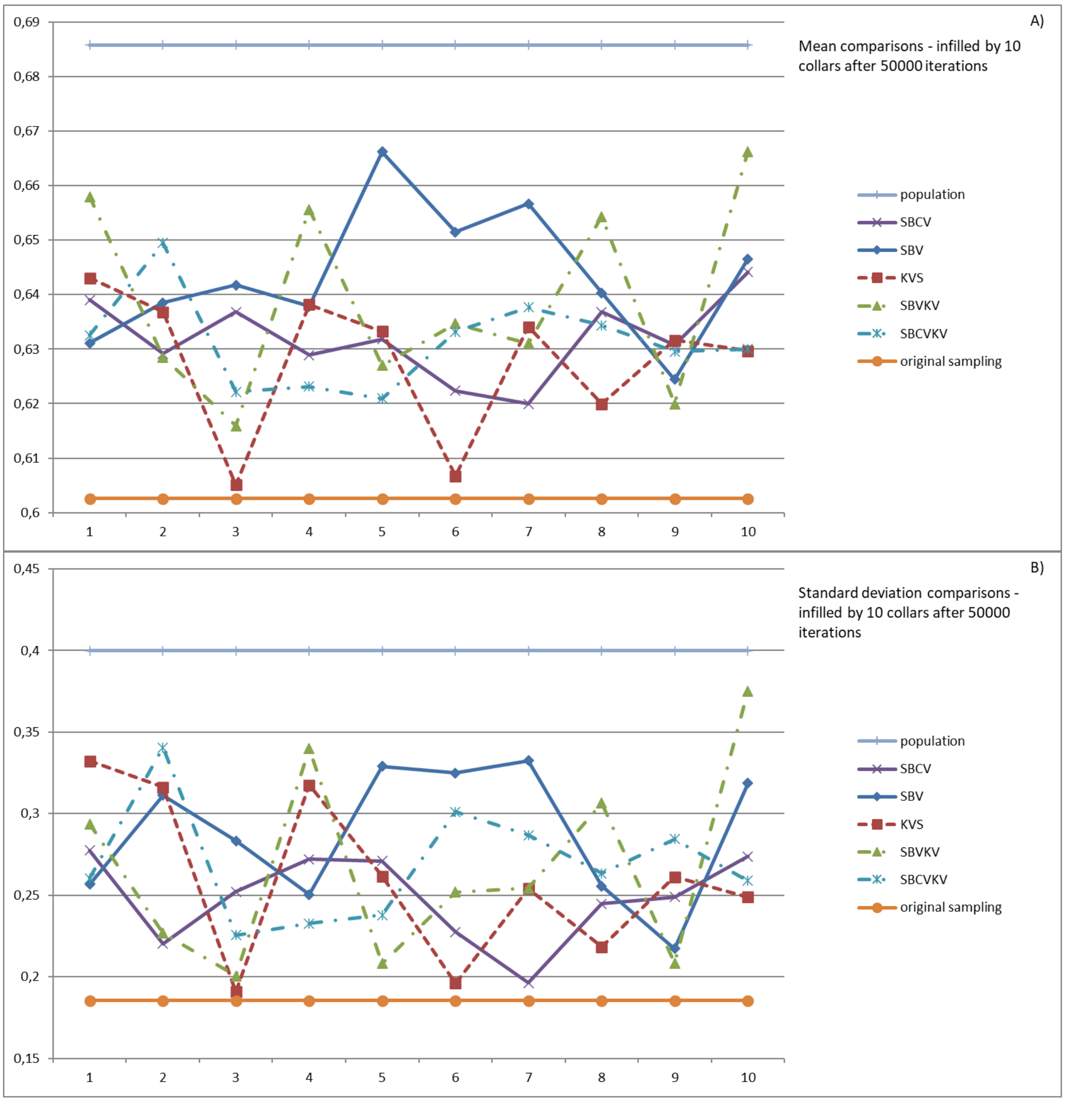


Figure P02-15 – Infill optimization results comparisons. The values compared are mean (A), standard deviation (B), and Kolmogorov-Smirnov distance (C). The scenario considered is to locate 10 infill collars optimized by 10000 iterations. The Kolmogorov-Smirnov distance value is obtained by comparing each distribution to the population, hence the population being zero at (C).

Table P02-05 – Best parameter value, and its distribution spread, for each objective function applied to infill 10 collars, optimized by 10000 iterations. The distribution spread was obtained by subtracting the minimum value from the maximum, considering the 10 optimizations applied with each objective function.

	SBV	SBCV	KVS	SBVKV	SBCVKV
Distribution spread (mean)	0.048	0.047	0.025	0.038	0.034
Distribution spread (std. dev.)	0.174	0.198	0.161	0.134	0.118
Distribution spread (Kolmo.-Smir. dist.)	0.030	0.026	0.030	0.021	0.054
Maximum mean	0.667	0.664	0.643	0.655	0.636
Maximum std. dev.	0.375	0.387	0.375	0.323	0.295
Minimum Kolmogorov-Smirnov dist.	0.054	0.051	0.078	0.059	0.083

The last tested infill scenario considers 10 new drillholes and is optimized after 50000 iterations and is presented in Figure P02-16 and resumed in Table P02-05. The comparison considering the mean (Figure P02-16 A) shows that the SBV was the best function, where the mean is closer to the population mean. The SBV mean value spread is high, compared to the results of the other functions, however, the results are the closest to the population overall. The worse results regarding the mean are obtained by minimizing the KVS. Considering the standard deviation (Figure P02-16 B) the best results are obtained by the SBVKV function, but this is the function where the highest spread is found. The best result is attained by the minimization of the SBCVKV. The worse result regarding similarity with populational standard deviation is obtained by the SBCV function. The Kolmogorov-Smirnov distance (Figure P02-16 C) with the smallest value to the population distribution is obtained by the SBV function. The highest distance and overall worse results are obtained by the KVS minimization.



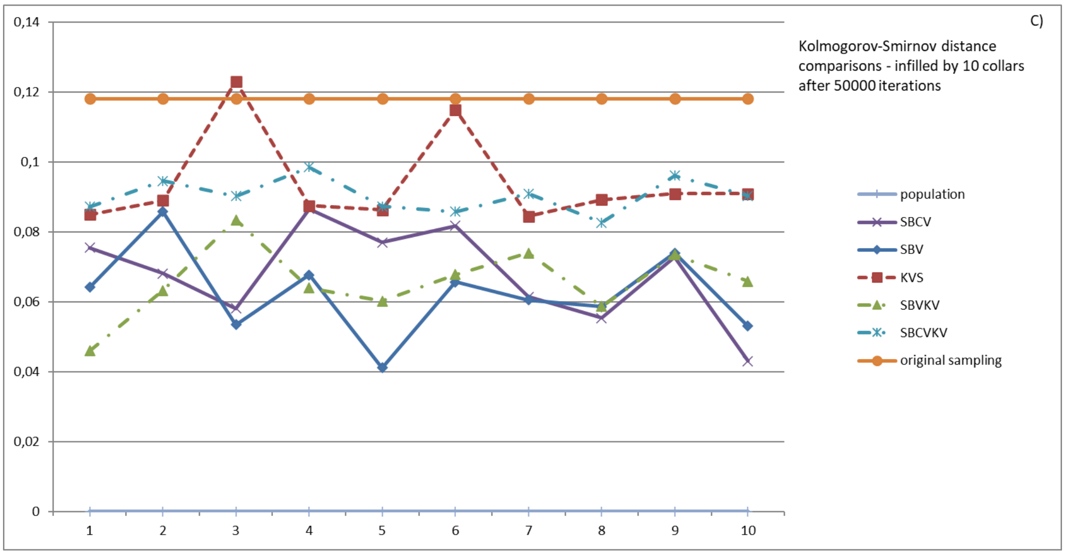
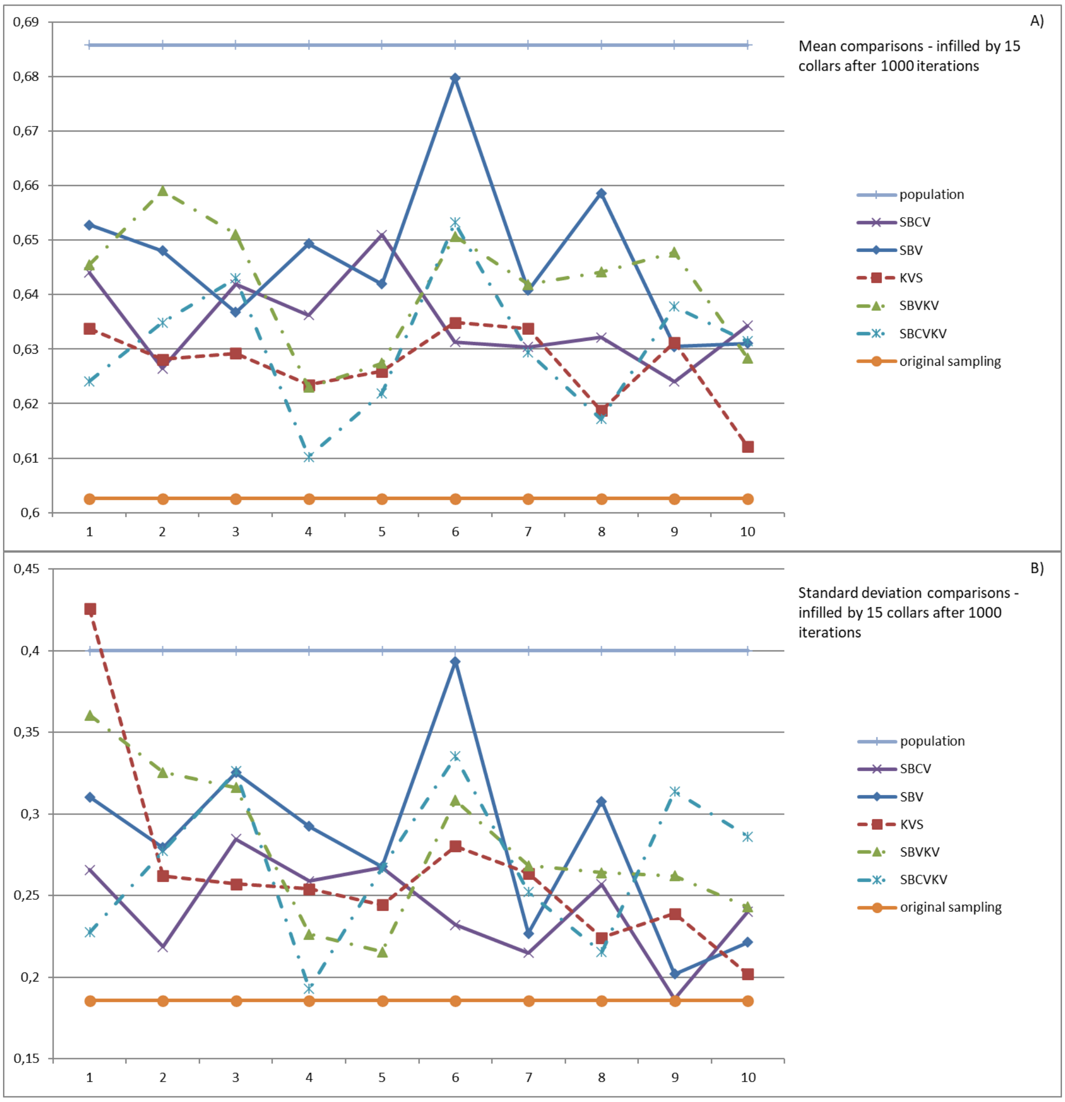


Figure P02-16 – Infill optimization results comparisons. The values compared are mean (A), standard deviation (B), and Kolmogorov-Smirnov distance (C). The scenario considered is to locate 10 infill collars optimized by 50000 iterations. The Kolmogorov-Smirnov distance value is obtained by comparing each distribution to the population, hence the population being zero at (C).

Table P02-06 – Best parameter value, and its distribution spread, for each objective function applied to infill 10 collars, optimized by 50000 iterations. The distribution spread was obtained by subtracting the minimum value from the maximum, considering the 10 optimizations applied with each objective function.

	SBV	SBCV	KVS	SBVKV	SBCVKV
Distribution spread (mean)	0.042	0.024	0.038	0.050	0.029
Distribution spread (std. dev.)	0.115	0.081	0.141	0.174	0.115
Distribution spread (Kolmo.-Smir. dist.)	0.045	0.044	0.039	0.037	0.016
Maximum mean	0.666	0.644	0.643	0.666	0.649
Maximum std. dev.	0.332	0.278	0.332	0.375	0.341
Minimum Kolmogorov-Smirnov dist.	0.041	0.043	0.084	0.046	0.083

The first scenario where the infill is done with 15 drillholes is optimized with 1000 iterations and its results comparisons can be seen in Figure P02-17 and Table P02-06. Comparisons of the mean (Figure P02-17 A) show that the best value is obtained by the SBV minimization, and even the highest interval has the closest values to the population. The worse results are obtained by the KVS minimization. Regarding the standard deviation comparisons (Figure P02-17 B) the SBV has the closest value to the population one, even with the highest spread. The minimization of the SBCV obtained the worse standard deviation values regarding the population. The comparisons with the Kolmogorov-Smirnov distance (Figure P02-17 C) show that the SBV minimization obtained the best results. The worse results, with a higher distance to the population distribution, are obtained by the SBCVKV function.



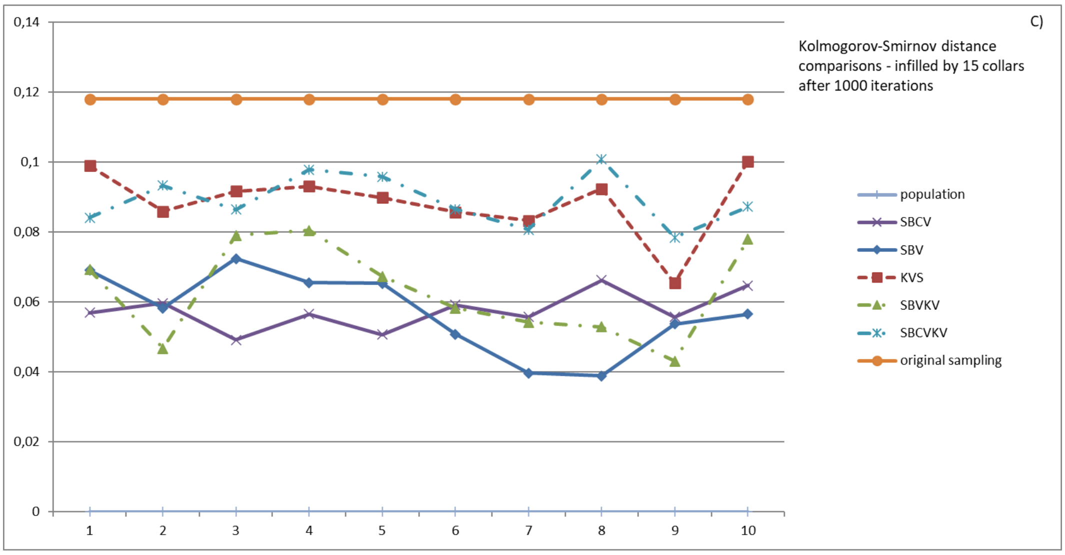


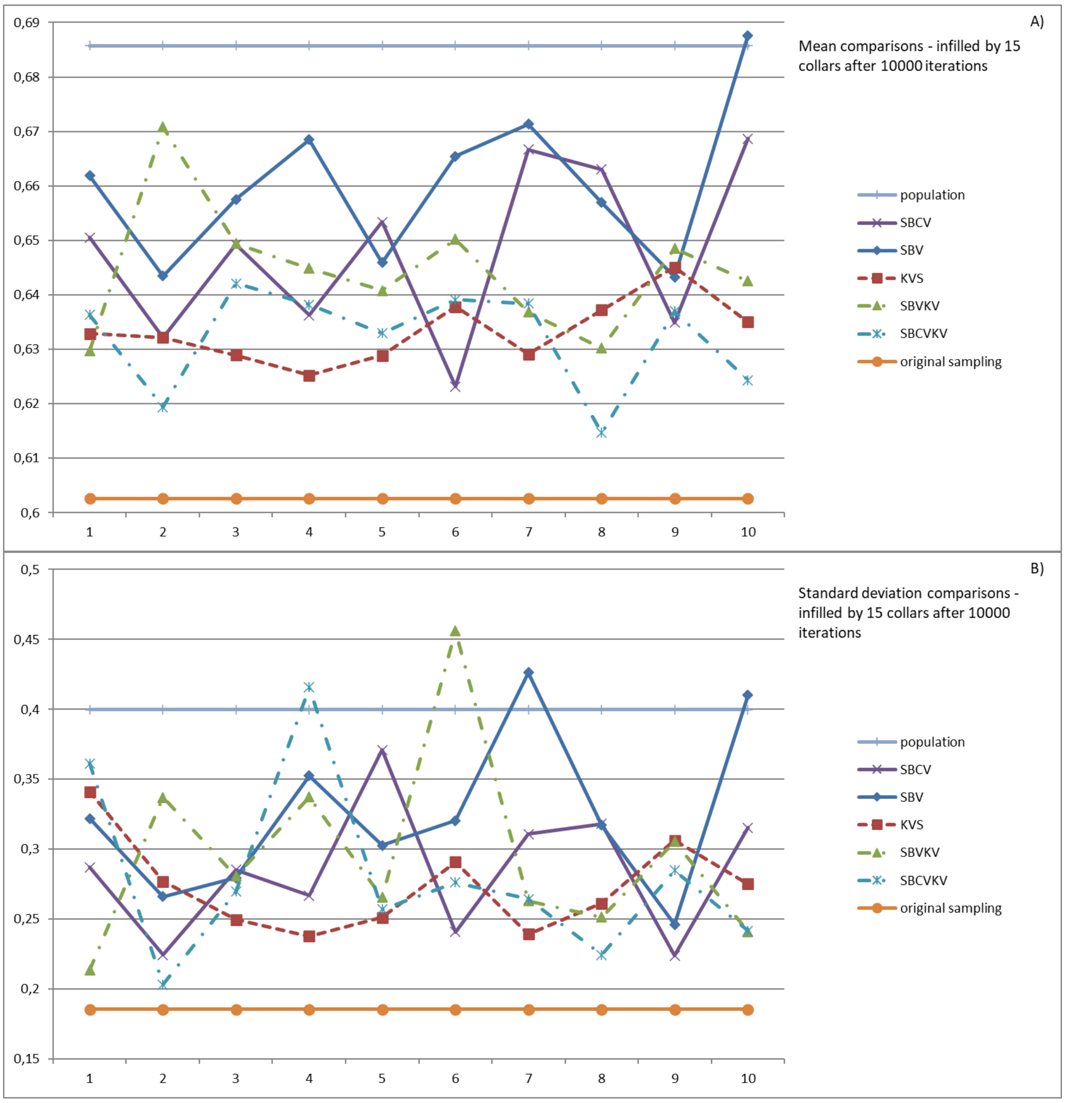
Figure P02-17 – Infill optimization results comparisons. The values compared are mean (A), standard deviation (B), and Kolmogorov-Smirnov distance (C). The scenario considered is to locate 15 infill collars optimized by 1000 iterations. The Kolmogorov-Smirnov distance value is obtained by comparing each distribution to the population, hence the population being zero at (C).

Table P02-07 – Best parameter value, and its distribution spread, for each objective function applied to infill 15 collars, optimized by 1000 iterations. The distribution spread was obtained by subtracting the minimum value from the maximum, considering the 10 optimizations applied with each objective function.

	SBV	SBCV	KVS	SBVKV	SBCVKV
Distribution spread (mean)	0.049	0.027	0.023	0.036	0.043
Distribution spread (std. dev.)	0.192	0.098	0.224	0.145	0.142
Distribution spread (Kolmo.-Smir. dist.)	0.034	0.017	0.035	0.037	0.022
Maximum mean	0.680	0.651	0.635	0.659	0.653
Maximum std. dev.	0.394	0.285	0.426	0.361	0.335
Minimum Kolmogorov-Smirnov dist.	0.039	0.049	0.065	0.043	0.078

A second optimization is made considering 15 new drillholes, located after 10000 iterations. The results of this set of optimizations are presented in Figure P02-18 and resumed in Table P02-07. Considering the mean (Figure P02-18 A) the minimization of SBV obtained the best representativity of the population. The worst results are obtained by the SBCVKV objective function. When the comparison is made regarding the standard deviation (Figure P02-18 B) the SBCVKV obtained the value closest to the populational one. However, this objective function has one of the highest distribution spreads. Therefore, the SBV fared better in furthering the representativity of the standard deviation regarding the population. The worse objective function is the KVS minimization. The last comparison using the

Kolmogorov-Smirnov distance (Figure P02-18 C) shows the SBV as the objective function and the results tend to be closer to the populational distribution. And again, the KVS minimization has the worse overall results.



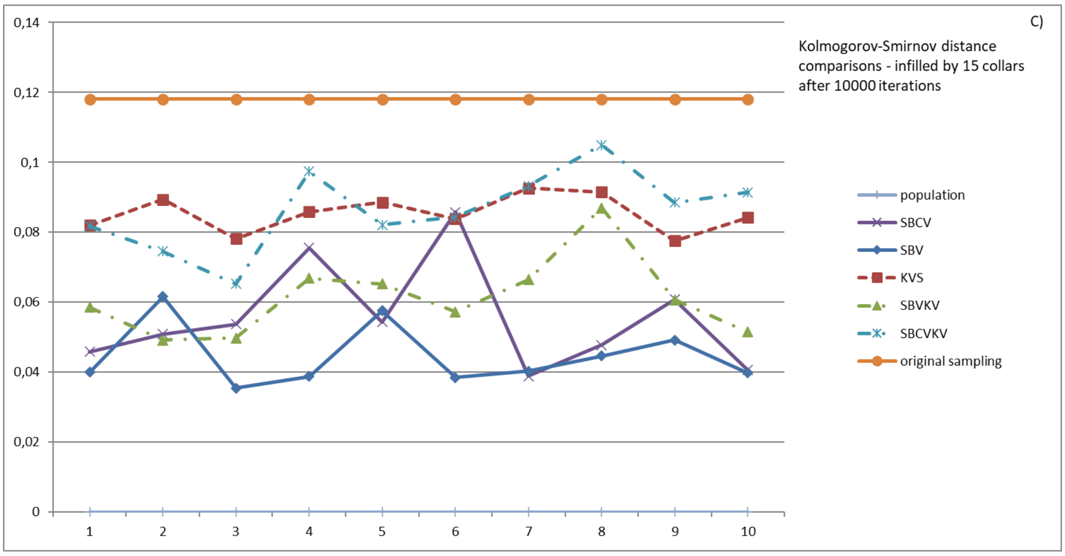


Figure P02-18 – Infill optimization results comparisons. The values compared are mean (A), standard deviation (B), and Kolmogorov-Smirnov distance (C). The scenario considered is to locate 15 infill collars optimized by 10000 iterations. The Kolmogorov-Smirnov distance value is obtained by comparing each distribution to the population, hence the population being zero at (C).

Table P02-08 – Best parameter value, and its distribution spread, for each objective function applied to infill 15 collars, optimized by 10000 iterations. The distribution spread was obtained by subtracting the minimum value from the maximum, considering the 10 optimizations applied with each objective function.

	SBV	SBCV	KVS	SBVKV	SBCVKV
Distribution spread (mean)	0.044	0.046	0.020	0.041	0.027
Distribution spread (std. dev.)	0.180	0.147	0.103	0.242	0.212
Distribution spread (Kolmo.-Smir. dist.)	0.026	0.047	0.015	0.038	0.040
Maximum mean	0.688	0.669	0.645	0.671	0.642
Maximum std. dev.	0.426	0.371	0.341	0.456	0.416
Minimum Kolmogorov-Smirnov dist.	0.035	0.039	0.078	0.049	0.065

At last, 15 new drillhole optimized after 50000 iterations are presented in Figure P02-19 and resumed in Table P02-08. Considering the mean (Figure P02-19A) the SBV fared better, albeit slightly than the SBCV regarding the mean representativity. The worse results considering the mean are obtained by the SBCVKV function. Regarding the standard deviation comparisons (Figure P02-19 B) the KVS has the best overall results. The worst results are related to the SBCV objective function. The last comparison, the Kolmogorov-Smirnov distance (Figure P02-19 C), shows that the best function in furthering the population representativity is the SBV minimization. The worse function, regarding the distance, is the SBCVKV minimization.

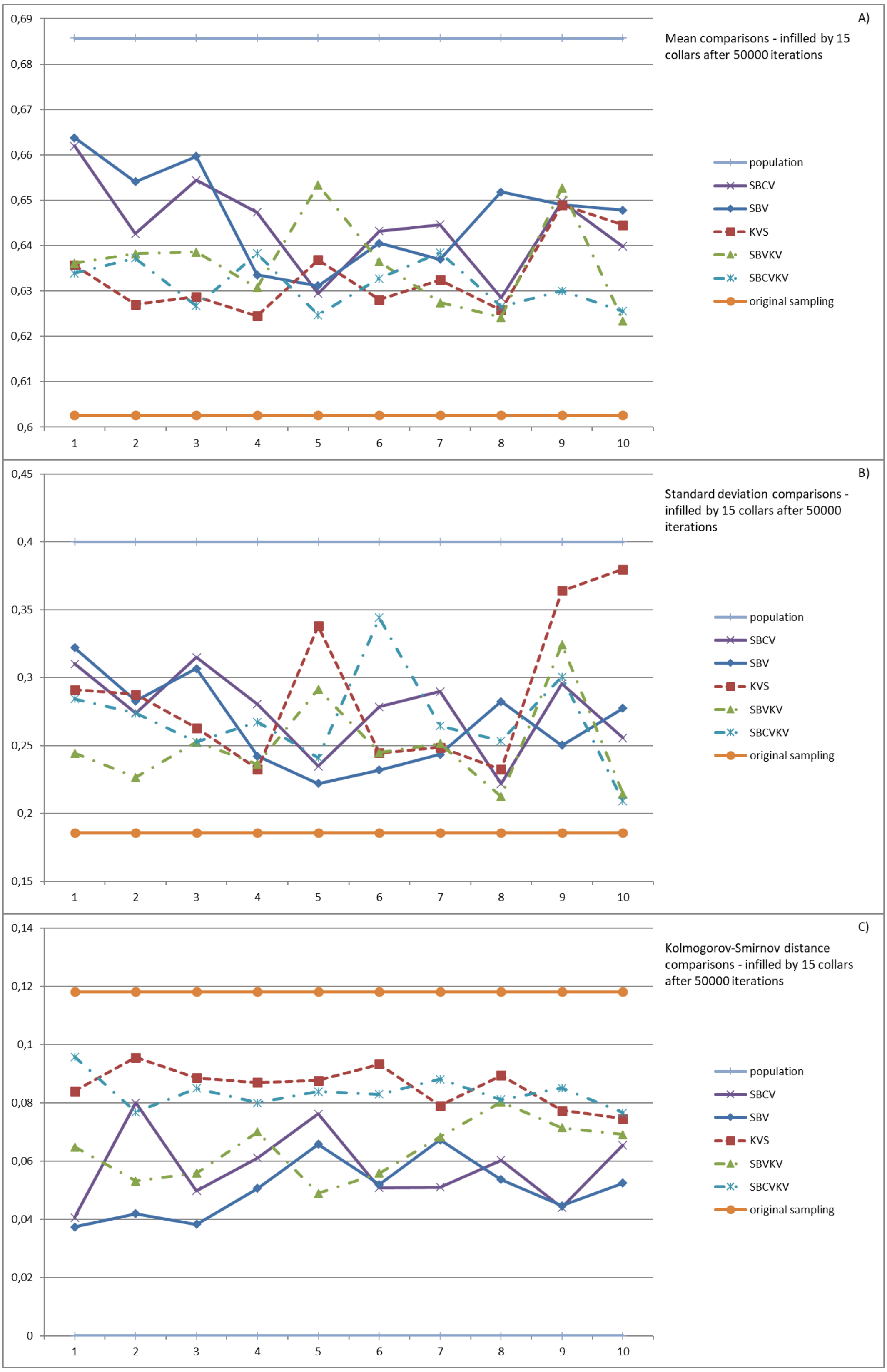


Figure P02-19 – Infill optimization results comparisons. The values compared are mean (A), standard deviation (B), and Kolmogorov-Smirnov distance (C). The scenario considered is to locate 15 infill collars optimized by 50000 iterations. The Kolmogorov-Smirnov distance value is obtained by comparing each distribution to the population, hence the population being zero at (C).

Table P02-09 – Best parameter value, and its distribution spread, for each objective function applied to infill 15 collars, optimized by 50000 iterations. The distribution spread was obtained by subtracting the minimum value from the maximum, considering the 10 optimizations applied with each objective function.

	SBV	SBCV	KVS	SBVKV	SBCVKV
Distribution spread (mean)	0.033	0.033	0.024	0.030	0.014
Distribution spread (std. dev.)	0.100	0.093	0.147	0.112	0.135
Distribution spread (K-S* distance)	0.030	0.039	0.021	0.031	0.019
Maximum mean	0.664	0.662	0.649	0.653	0.638
Maximum std. dev.	0.322	0.315	0.380	0.324	0.344
Minimum K-S distance	0.037	0.041	0.074	0.049	0.077

* K-S is used as an abbreviation of Kolmogorov-Smirnov

The results when dealing with synthetic data show that the objective function SBV fared better than the other functions in furthering the populational representativity. However, this is one of the functions that tend to have higher results distribution spreading. The functions present a clear tendency of separation in the Kolmogorov-Smirnov distance, with a difference in the KVS results, which tend more to the original sampling distance, and the SBV and SBCV functions, which tend to the population more. The SBVKV results tend to approximate more from the simulated functions. Meanwhile, the SBCVKV tends to the KVS results.

The use of a higher number of iterations does not necessarily provide better results, or with smaller results distribution spreading. However, a higher number of drillholes tends to return better results for each parameter, when considering the same number of iterations. However, the same is not perceived when the spread is considered, as there is not a clear trend of either rising or fall of the values when 15 new drillholes were made compared with the infill with 10 collars.

6.2.4.2 Real data set

The optimized infill made at the Capanema drilling data set contains 21 new drillholes optimized after 10000 iterations, this scenario was repeated 10 times for each objective function and the results were sampled using the rockdrill data. The comparisons made are the same used in the synthetic data, considering mean, standard deviation, and Kolmogorov-

Smirnov distance. In Figure P02-20 the comparisons are presented, with the values of mean and standard deviation being proportional to the populational values. The intervals and best values for each objective function are presented in Table P02-09. Considering the populational mean, the best value was obtained by the KVS minimization, and the second-best results are those of SBCVKV minimization. Both objective functions attained overall better results regarding the mean. Considering the spread of the results interval regarding the mean, SBV and KVS obtained the highest, with the smallest spread obtained by the SBCV. Although the KVS has the highest spread the maximum value was interpreted as the best objective function to represent the populational mean. The standard deviation results once again present a distinction, with the SBCVKV and KVS minimization with the best overall results, with the SBCVKV objective function being interpreted as the best result. The standard deviation results tend to be closer to the original sampling data when compared to the mean results. The SBCVKV minimization has the highest interval, indicating high uncertainty, while the smallest interval was presented by the SBV objective function. Even with the smallest spread of standard deviation, the best values obtained by the SBV minimization were worse than the best standard deviation obtained by the KVS minimization, in which the interval of results is rather small compared to the other objective functions. Therefore, the KVS has the best results regarding the population standard deviation. The last comparison made regards the Kolmogorov-Smirnov distance, which once again presents a distinction between the results, with both KVS and SBCVKV objective function, having the smallest overall results. The best result was obtained by the SBCVKV objective function, with the best KVS minimization result close. Although the KVS and SBCVKV objective functions obtained the best result, the interval for both is the highest among the objective functions considered. Regardless of that, worse results are mostly higher than the ones obtained by KVS and SBCVKV objective functions. Therefore, the best objective function regarding the Kolmogorov-Smirnov distance was the minimization of the KVS.

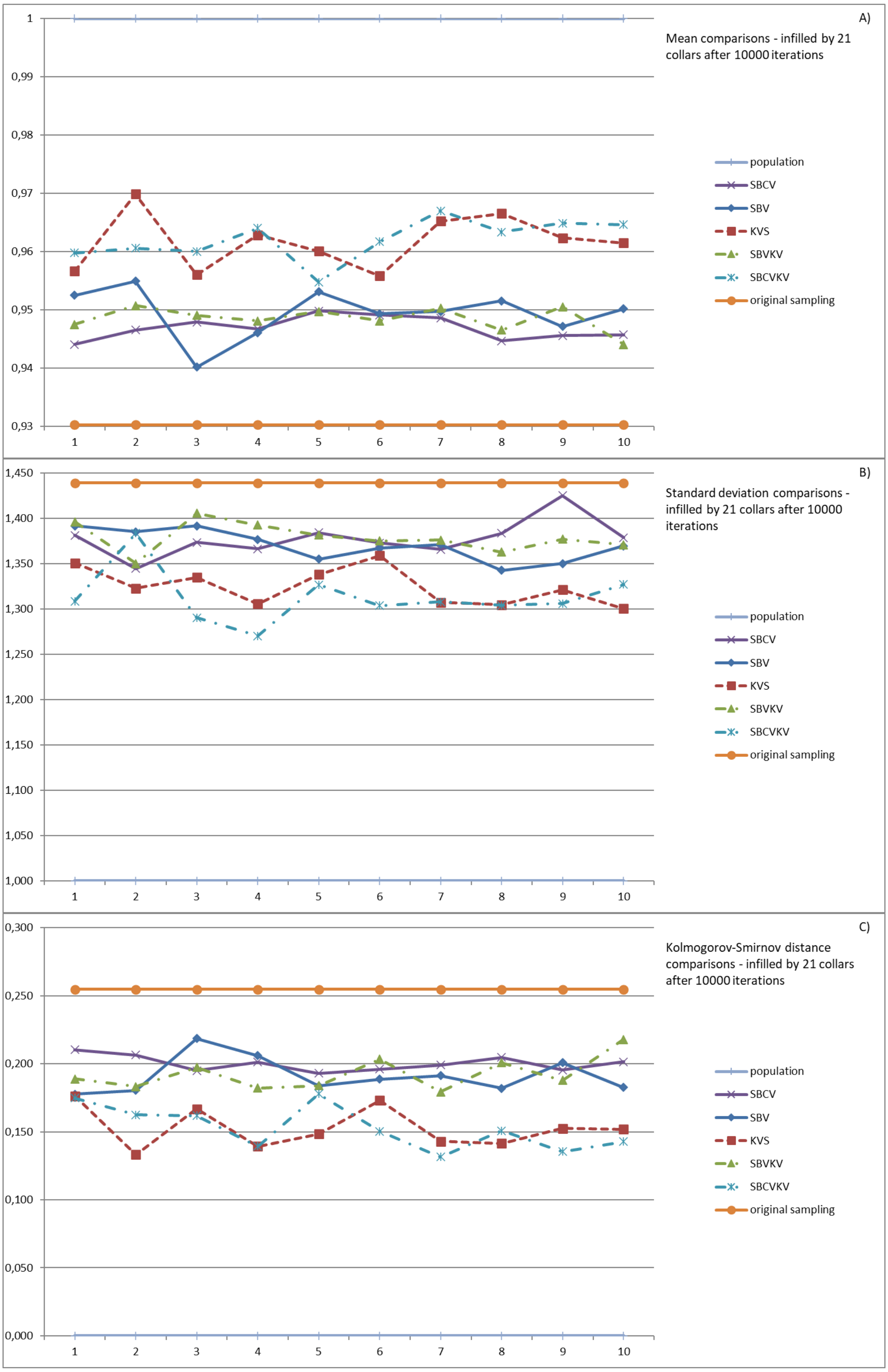


Figure P02-20 – Infill optimization of Capanema drillholes results comparisons. The values compared are mean (A), standard deviation (B), and Kolmogorov-Smirnov distance (C). The scenario considered is to locate 21 infill collars optimized by 10000 iterations. The Kolmogorov-Smirnov distance value is obtained by comparing each distribution to the population, hence the population being zero at (C). Each mean and standard deviation value are taken proportionally to the rockdrill data, hence the reason for the population values being equal to 1.

Table P02-10 – Collar infill distribution spread and best results for each objective function applied. The distribution spread was obtained considering the 10 optimizations applied with each objective function.

	SBV	SBCV	KVS	SBVKV	SBCVKV
Distribution spread (mean)	0.015	0.006	0.014	0.007	0.012
Distribution spread (std. dev.)	0.049	0.081	0.059	0.056	0.113
Distribution spread (K-S Distance)	0.041	0.017	0.043	0.039	0.046
Maximum mean	0.955	0.950	0.970	0.951	0.967
Minimum std. dev.	1.343	1.345	1.300	1.350	1.270
Minimum K-S distance	0.178	0.193	0.133	0.179	0.132

Those results differ from those obtained in synthetic dataset comparisons, considering the objective function competence. In the synthetic dataset comparisons, the overall best objective function was the SBV, while the KVS obtained the best results to infill the Capanema sampling, regarding the representativity. To better compare the Capanema and synthetic dataset infill it is important to consider the differences between variables. In Table P02-10 one can see the variance coefficient of the synthetic dataset initial sampling, the synthetic dataset population, sampling by drillholes of Capanema mining, and the assumed population of Capanema, a sampling made with rockdrill. Considering the coefficient of variation of data, it is remarkable that the synthetic dataset sampling has a higher variability, compared to the Capanema information. That is to be expected as the synthetic dataset data is a copper occurrence and the Capanema is an iron deposit. The fact that the Capanema data have lower variability can explain the reason that the KVS objective function obtained the best responses. Lower variability indicates that the data do not necessarily need a more complex approach to indicate the regions with lower sampling density. The fact that the model has a small degree of uncertainty would indicate that the simplest approach fared better.

Table P02-11 – Comparisons between the coefficient of variation of each data considered. The original sampling of synthetic dataset and Capanema that was infilled and the population data for each sampling.

Data	Coefficient of variation
Synthetic dataset original drillhole sampling	0.308
Synthetic dataset population	0.583
Capanema original drillhole sampling	0.137
Capanema rockdrill sampling (population)	0.088

The last consideration between the Capanema and synthetic dataset infill is with the computational time. In Table P02-11 the infilled data processing time is compared, considering the same number of iterations for both optimizations, 10000, and with 10 collars infilled at the synthetic dataset data. The average processing time to optimize the synthetic dataset sampling is lower than the Capanema infill. The reason for this difference is the size of each domain. To the synthetic dataset data, the North and East size are, respectively, equal to 600 and 300 meters, meanwhile, the same sizes to the Capanema are equal to 2850 and 2150 meters. The domain size is used to define the number of iterations, meaning that the higher its size is the more iterations are recommended. However, KVS tends to be the faster objective function, as kriging uses less computational time. The compost objective function considering the synthetic dataset infill has not changed the processing time in the case of the SBCVKV. The SBVKV objective function to infill the synthetic dataset data has returned a higher elapsed processing time because some of the optimizations have demanded more time, which indicates a computational problem. Considering Capanema at least one day is demanded to process an infill optimization, with at least more than 9 hours when using a compost objective function. However, KVS minimization requires lesser processing time even when infilling the Capanema sampling. The processing time is longer if the model size is higher, therefore, Capanema infill will demand more time.

Table P02-12 – Processing time of each objective function considering both sampling data to be infilled, synthetic dataset, and Capanema drillhole. The synthetic dataset infill was considered with 10 infilled collars and 10000 iterations.

	SBCV	SBV	KVS	SBCVKV	SBVKV
Synthetic dataset infill time	04:07:36	04:06:11	01:33:10	05:28:43	09:19:00
Capanema infill time	24:13:24	24:08:58	08:57:27	33:05:15	33:02:12

6.2.5 Final considerations and conclusions

The optimization of the infill collar is not trivial and easily applicable. The new proposed compost approach does not fare better than objective functions considering only the simulated or estimated models. Although, all function tends to obtain better results to represent the population distribution when applied. The best objective function to infill the synthetic data was minimizing the SBV, and obtaining the closest distribution to the populational data. The KVS minimization obtained the best result to optimize the infill drillhole location of Capanema mining. A high number of infilled collars and iterations tends to obtain better results; however, this does not mean that a smaller number of both cannot return satisfactory infill locations. Therefore, it is interesting to apply the proposed objective functions, specifically the non-compost type that fared better overall.

The better representativity achieved by the optimizations proposed does not mean the study and application of infill drillhole location should only be made utilizing the proposed methodology. The use of the proposed objective functions and optimization algorithm as a guide with other infill localization methodologies is interesting. The methodology should be integrated and applied with the exploratory team's expertise to indicate high-interest areas, considering the lithology limits and directions and different limitations imposed by the domain. As a new tool to guide further knowledge and representativity of the sampling data, the proposed methodology can be applied with a low computational and personal cost.

6.3 The effect of sample values while optimizing infill drillhole locations

The first comparison made was between the optimization of both objective functions, SBV and SBCV, by the same algorithm, SA with fast cooling, with the different values associated with the infill samples while searching for the optimum. The values chosen for this test are 4, mean, median, 10^{Th} percentile (P10), and 90^{Th} (P90) percentile of the nearest simulated node, to the infill sample to be considered. The results of the optimizations are shown in Figure P03-21. The first noticeable point is the fact that, systematically, the worse optimum objective function value found, for the same function, was obtained by associating the P90 to the infill samples. This could be explained by the fact that the P90 probably being an outliner in this distribution, therefore, when its value is considered as the value of the new sample there is a rise in the variance value over the domain, which affects the objective functions and rises then as well. Another systematic behavior is the fact that both the P10 and median obtained smaller optimum values when compared with the mean, with the P10 obtaining 11 of the 15 smallest objective function values when minimizing the SBV function, and 8 when minimizing the SBCV function.

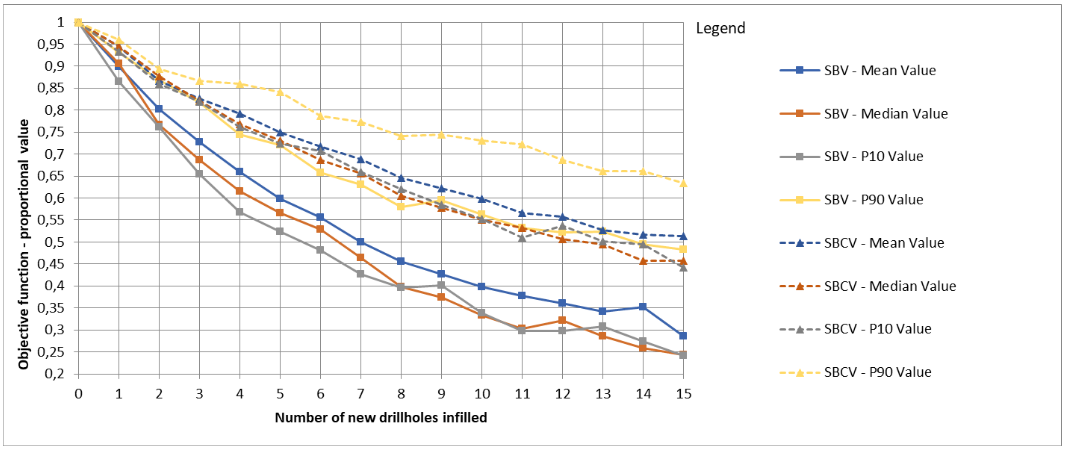
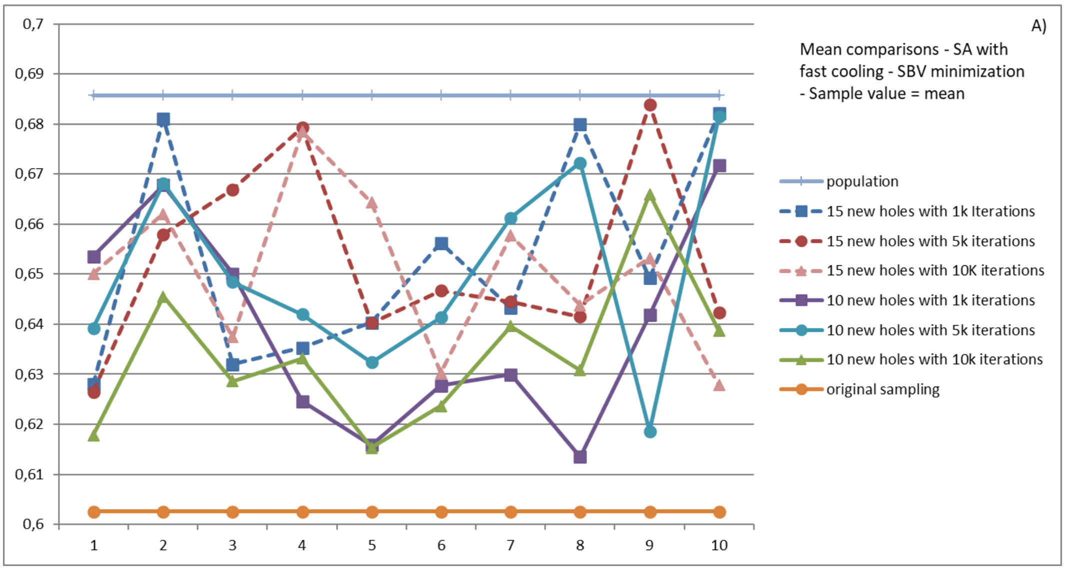
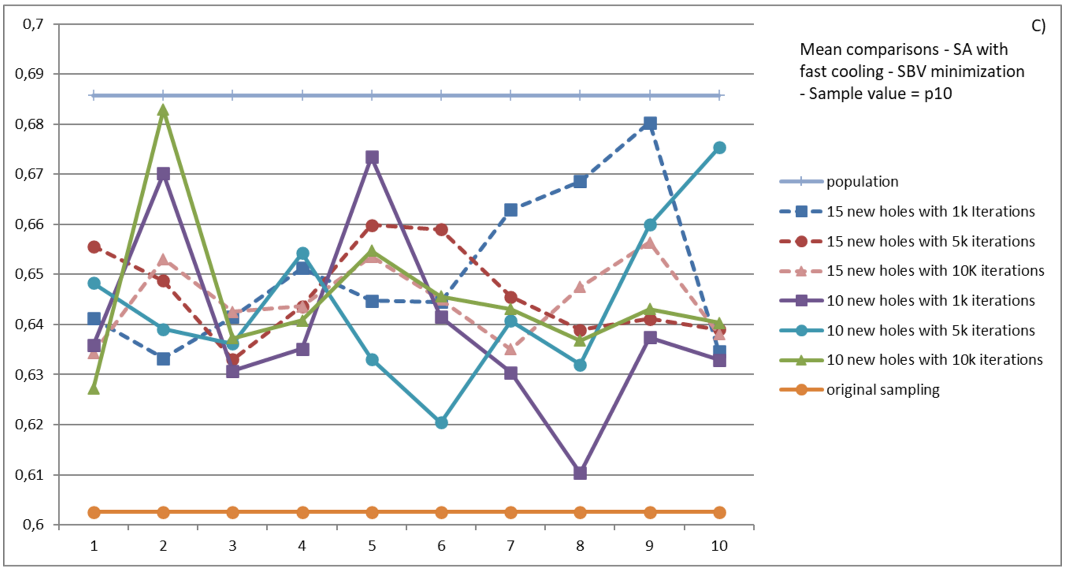


Figure P03-21 – Comparisons between each objective function optimization, obtained after 1000 iterations, under the same parametrizations while using different values associated with the infill in the searching for the optimum.

To fully assess the effect of each value in locating infill samples 10 optimizations were made with each possible value and under different parameterizations. The number of new drillholes to be located by the optimization were 10 and 15, while the number of iterations to complete the optimization were 1000, 5000, and 10000. With those six parametrizations, and with both SBV and SBCV objective functions, the optimum collar location obtained by each optimization was sampled in the synthetic body to compare the results, to then assess which sample value would return the infill that best represents the population distribution. To analyze the results the choice made was to use 3 parameters that would be compared with the population, being the mean, standard deviation, and Kolmogorov-Smirnov distance.

The results presented in Figure P03-22 are the comparison by the mean value for the 4 utilized values associated with the samples while minimizing the SBV function. The infill utilizing the median as the value of the samples while optimizing its location returned the values closest to the populational mean. The results considering the mean as the value are more erratic, with values almost equal to the population value while obtaining values close to the original sampling mean. The infill optimized while utilizing the P10 as a value is less erratic overall, with most of the values concentrating in the 0.64-0.65 interval. The optimizations considering the P90 as the sample values show a shift when compared to the others, where the means tend to the original sampling values, with one optimum collar configuration being different from even the original sampling. This outlier mean value could be related to the populational distribution, in the fact that 0.5 is the value with the most repetitions (and being the minimum), therefore the P90 in that optimization obtained most of the data with values with the minimum value of the population.





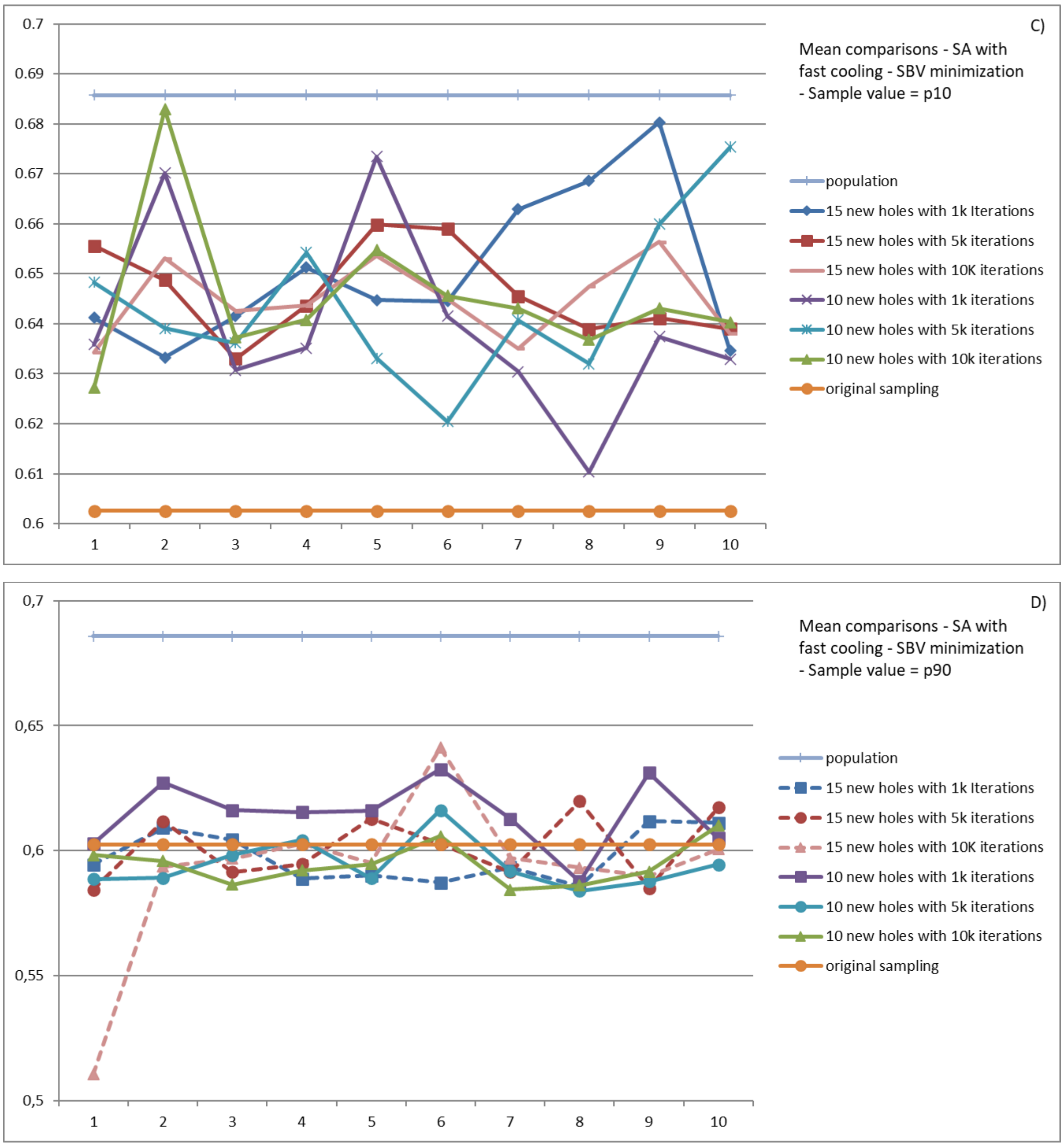
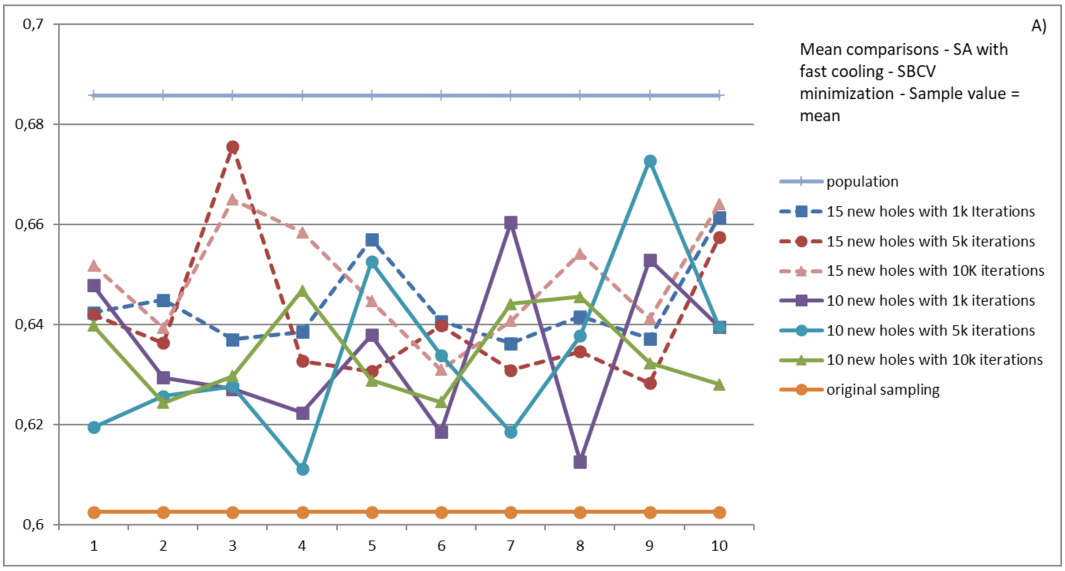
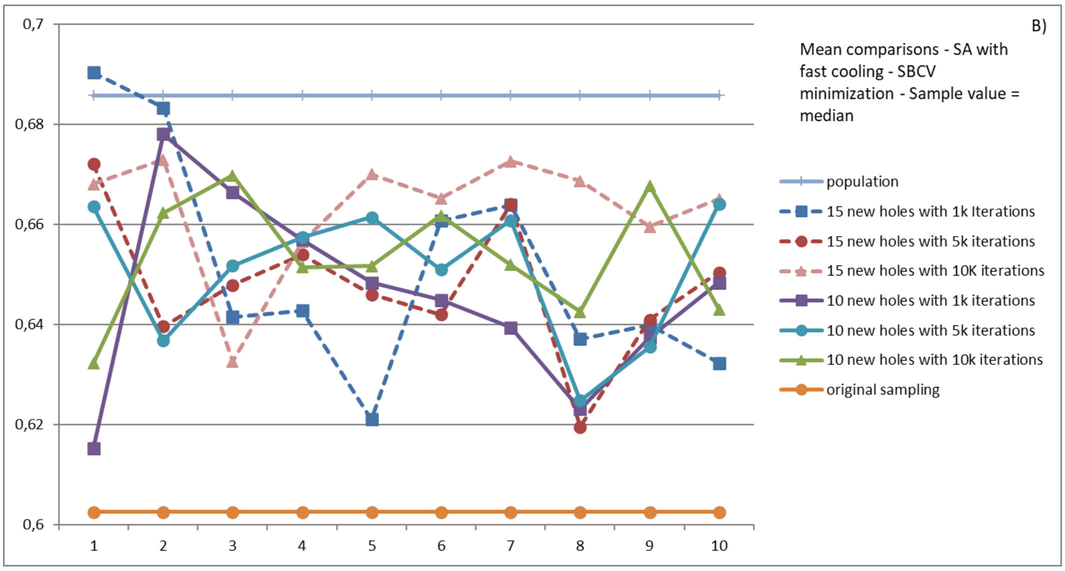
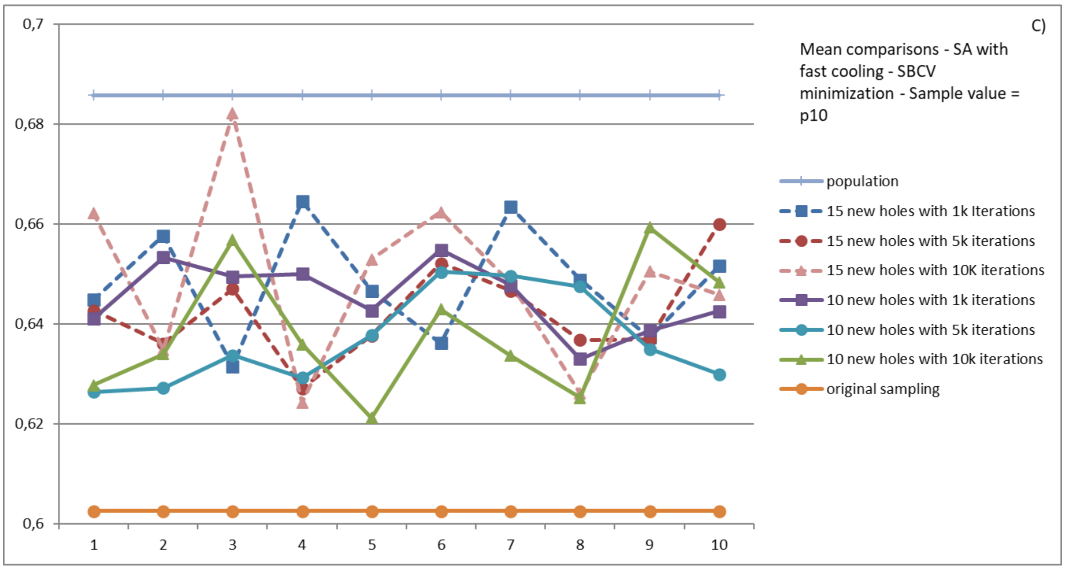


Figure P03-22 – Comparisons between the mean of 10 optimums infill locations for each search parametrization, number of new collars, and number of iterations, while minimizing the SBV function, and the populational and original sampling mean. The values associated with the infill samples while searching for the optimum were: mean (A); median (B); P10 (C); and P90 (D).

The same comparisons to the mean were made while minimizing the SBCV function and are presented in Figure P03-23. The use of the median as the value of the infill samples while minimizing the SBCV obtained the infilled data closer to the populational mean, the same as seen in the SBV minimization. The infill utilizing the mean and the P10 as sample values while optimizing obtained the infill with mean values around the same interval, with the P10 results once again occurring in a smaller interval, less erratic. The optimization considering the P90 as sample values once again obtained values closest to the original sampling, similar to the result obtained while minimizing SBV.


A)
Mean comparisons - SA with fast cooling - SBCV minimization - Sample value = mean
0,7
0,68
0,66
0,64
0,62
0,6
1
2
3
4
5
6
7
8
9
10
population
15 new holes with 1k Iterations
15 new holes with 5k iterations
15 new holes with 10K iterations
10 new holes with 1k iterations
10 new holes with 5k iterations
10 new holes with 10k iterations
original sampling


B)
Mean comparisons - SA with fast cooling - SBCV minimization - Sample value = median
0,7
0,68
0,66
0,64
0,62
0,6
1
2
3
4
5
6
7
8
9
10
population
15 new holes with 1k Iterations
15 new holes with 5k iterations
15 new holes with 10K iterations
10 new holes with 1k iterations
10 new holes with 5k iterations
10 new holes with 10k iterations
original sampling


C)
Mean comparisons - SA with fast cooling - SBCV minimization - Sample value = p10
0,7
0,68
0,66
0,64
0,62
0,6
1
2
3
4
5
6
7
8
9
10
population
15 new holes with 1k Iterations
15 new holes with 5k iterations
15 new holes with 10K iterations
10 new holes with 1k iterations
10 new holes with 5k iterations
10 new holes with 10k iterations
original sampling

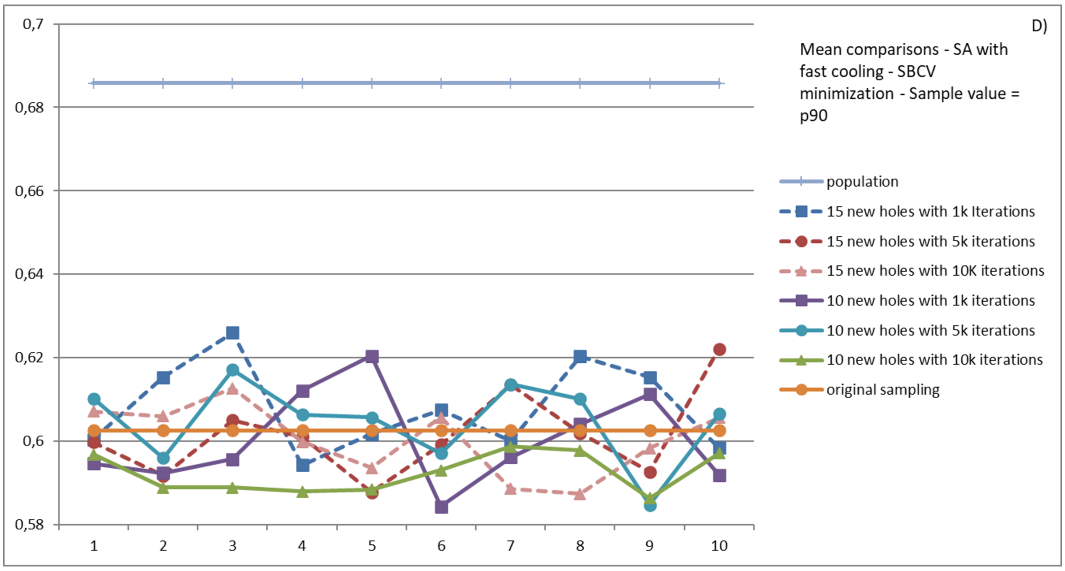


Figure P03-23 – Comparisons between the mean of 10 optimums infill locations for each search parametrization, number of new collars, and number of iterations, while minimizing the SBCV function, and the populational and original sampling mean. The values associated with the infill samples while searching for the optimum were: mean (A); median (B); P10 (C); and P90 (D).

In Table P03-12 the comparisons of the mean are presented focusing on the maximum and minimum, the interval, and median value of each parametrization. The minimization of SBV with 15 new collars and after 5000 iterations while utilizing the mean as the infill samples values obtained the mean value of the infilled distribution closest to the populational mean. The parametrization with the median of the 10 optimizations that came closer to the populational mean was the minimization of SBV with 15 new collars and 10000 iterations while utilizing the median as the infill samples. Both the utilization of P10 and P90 obtained the smallest intervals, considering the SBV and SBCV minimizations. The minimization of SBCV obtained the overall smaller intervals, but the maximum and median of the infilled distributions mean are, overall, closer to the populational mean when the SBV was minimized.

Table P03-13 – Values of minimum, maximum, interval, and median, of the infill sample distribution mean, obtained by the 10 optimizations for each different parametrizations considering the 4 sample values utilized while searching for the optimum, mean, median, P10, and P90.

SBV

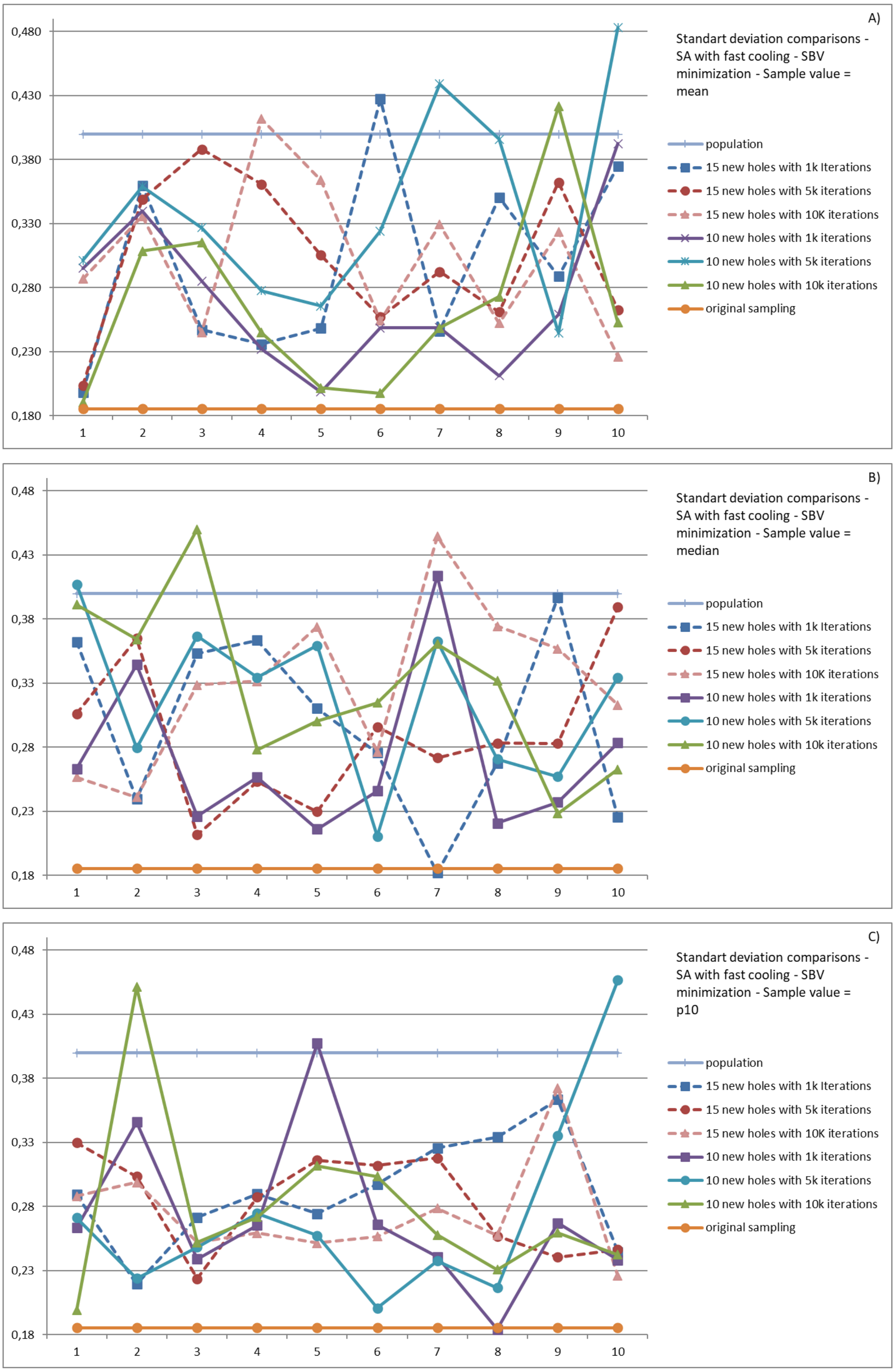
	collars	10	10	10	15	15	15
	iterations	1000	5000	10000	1000	5000	10000
Mean	maximum	0.672	0.682	0.666	0.682	0.684	0.679
	interval	0.058	0.063	0.051	0.054	0.057	0.051
Median	maximum	0.679	0.681	0.678	0.677	0.683	0.683
	interval	0.049	0.053	0.042	0.073	0.049	0.041
P10	maximum	0.673	0.675	0.683	0.680	0.660	0.656
	interval	0.063	0.055	0.056	0.047	0.027	0.022
P90	maximum	0.633	0.616	0.610	0.612	0.620	0.641
	interval	0.045	0.032	0.026	0.026	0.035	0.131

SBCV

	collars	10	10	10	15	15	15
	iterations	1000	5000	10000	1000	5000	10000
Mean	maximum	0.657	0.664	0.644	0.651	0.669	0.662
	interval	0.045	0.047	0.024	0.027	0.046	0.033
Median	maximum	0.678	0.664	0.670	0.690	0.672	0.673
	interval	0.063	0.039	0.038	0.069	0.052	0.040
P10	maximum	0.655	0.650	0.659	0.665	0.660	0.682
	interval	0.022	0.024	0.038	0.033	0.033	0.058
P90	maximum	0.620	0.617	0.599	0.626	0.622	0.613
	interval	0.036	0.032	0.012	0.032	0.034	0.025
Populational mean					0.686		

The second comparison was made regarding the standard deviation of the population data. In Figure P03-24 the infill optimization by minimizing the SBV for each of the 4 values associated with the samples is presented. The standard deviation of the optimum infill presents a more erratic behavior than the mean, but there are some similarities, such as the fact that the optimizations considering the P10 and P90 are less erratic than the ones utilizing the mean and median. The fact that the utilization of P90 as the sample values tends to shift the standard deviation of the optimums to the original sampling occurs in the comparisons considering the standard deviation too, likewise with the comparisons considering the mean. It is not clear as in the comparisons considering the mean, but in the case of the standard deviation, the median has a slightly better, i.e., standard deviation values of the infill samples distribution closer to the populational value. Although the optimums located by utilizing the

P10 return a less erratic standard deviation, the results are worse than those obtained by utilizing the mean or median as the sample values while searching for the optimum.



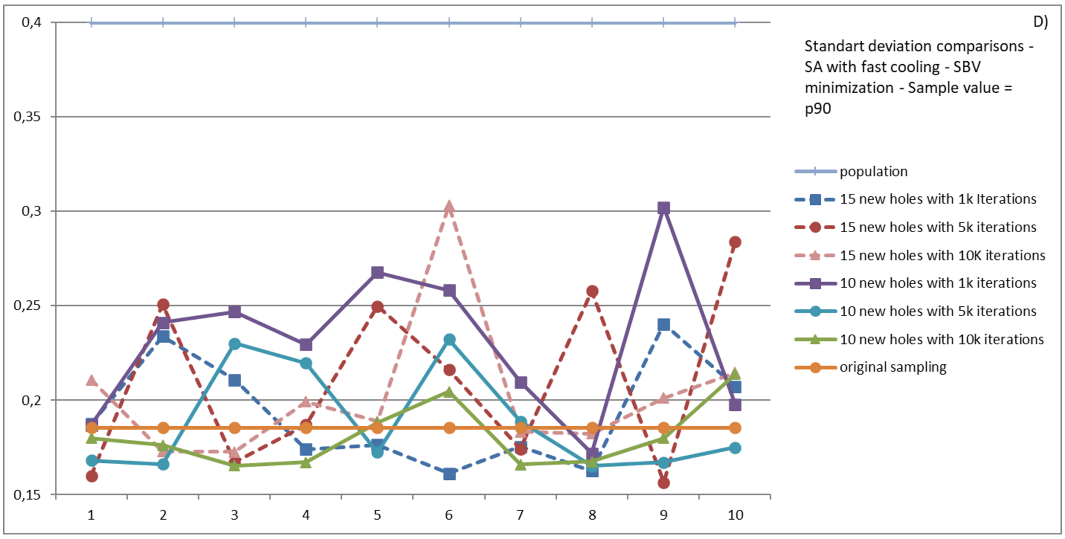
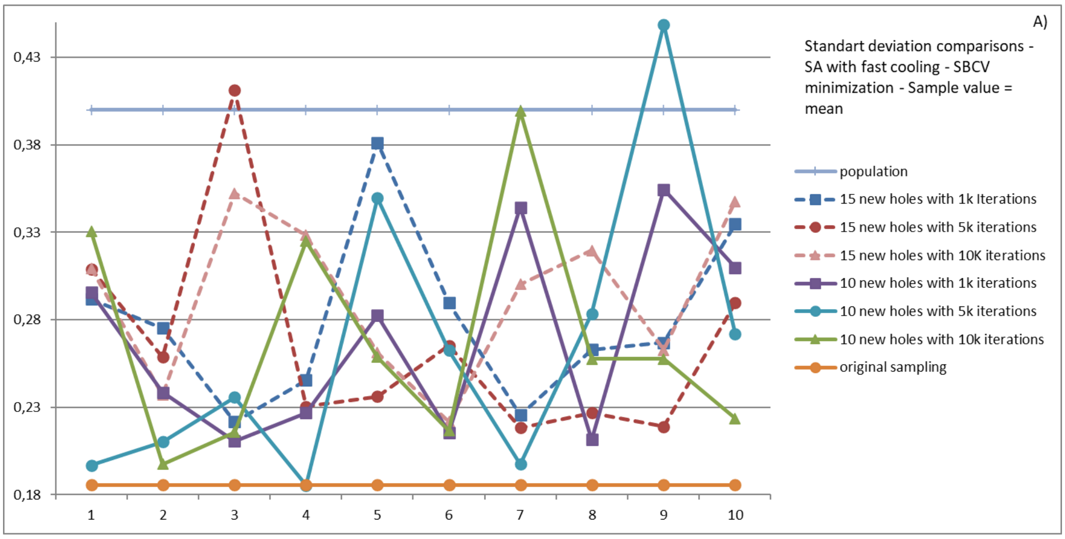


Figure P03-24 – Comparisons between the standard deviation of 10 optimums infill locations for each search parametrization, number of new collars, and number of iterations, while minimizing the SBV function, and the populational and original sampling standard deviation. The values associated with the infill samples while searching for the optimum were: mean (A); median (B); P10 (C); and P90 (D).

The comparisons made considering the optimum obtained while minimizing the SBCV are presented in Figure P03-25. The minimization of SBCV obtained a result less erratic than those of the SBV minimization, similar to that seen when comparing the mean. The optimizations utilizing the median as sample values while locating the infill obtained the results closer to the populational standard deviation. The optimizations utilizing the mean and the P10 obtained results around the same interval, with the mean being more erratic, as expected from the previous results. The results utilizing the P90 as sample value have shifted the results to the original sampling standard deviation again.



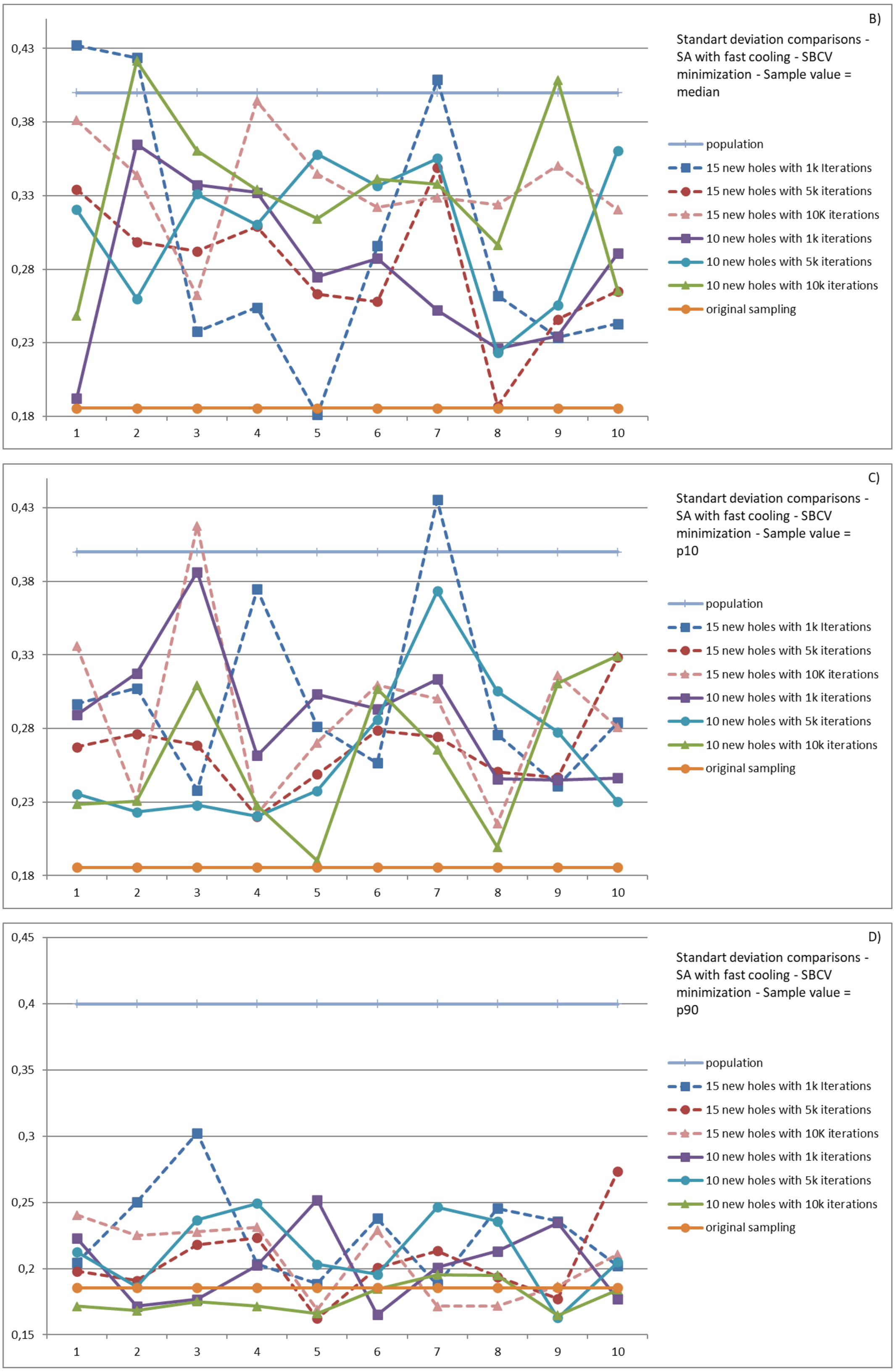


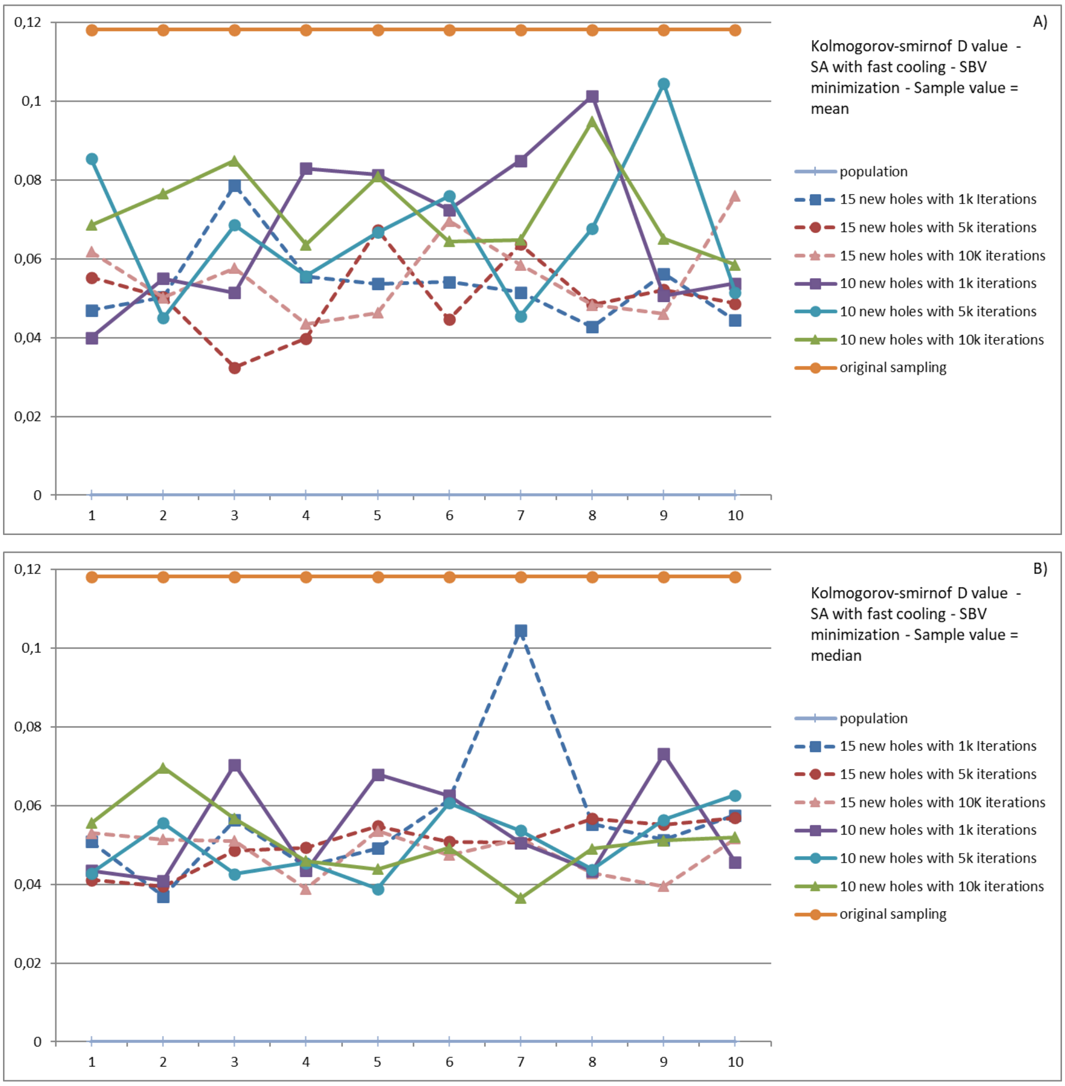
Figure P03-25 – Comparisons between the standard deviation of 10 optimums infill locations for each search parametrization, number of new collars, and number of iterations, while minimizing the SBCV function, and the populational and original sampling standard deviation. The values associated with the infill samples while searching for the optimum were: mean (A); median (B); P10 (C); and P90 (D).

In Table P03-13 the comparisons of the standard deviation are presented focusing on the maximum and minimum, the interval, and median value of each parametrization. The configuration that obtained the best standard deviation value, i.e., closest to the population value, was the minimization of the SBV locating 15 new drillholes with 1000 iterations and that used the median as the sample value while searching for the optimum. The optimizations utilizing the P90 as the sample value obtained the smaller intervals overall, but those results are the worse compared with the other values adopted. When considering the maximum value, the optimization minimizing SBV obtained the values closest to the populational standard deviation. The optimizations utilizing the median as the sample value obtained the closest values to the population standard deviation when considering the maximum and median values obtained by the overall optimization parameters adopted.

Table P03-14 – Values of minimum, maximum, interval, and median, of the infill sample distribution standard deviation, obtained by the 10 optimizations for each different parametrizations considering the 4 sample values utilized while searching for the optimum, mean, median, P10, and P90.

				SBV			
	collars	10	10	10	15	15	15
	iterations	1000	5000	10000	1000	5000	10000
Mean	maximum	0.393	0.483	0.422	0.427	0.388	0.412
	interval	0.194	0.238	0.231	0.229	0.184	0.186
Median	maximum	0.414	0.407	0.450	0.397	0.389	0.444
	interval	0.198	0.196	0.222	0.215	0.177	0.203
P10	maximum	0.407	0.457	0.451	0.364	0.330	0.373
	interval	0.223	0.256	0.252	0.144	0.106	0.147
P90	maximum	0.302	0.232	0.214	0.240	0.284	0.303
	interval	0.130	0.067	0.048	0.079	0.127	0.130
				SBCV			
	collars	10	10	10	15	15	15
	iterations	1000	5000	10000	1000	5000	10000
Mean	maximum	0.327	0.387	0.278	0.285	0.371	0.315
	interval	0.131	0.198	0.081	0.098	0.147	0.093
Median	maximum	0.365	0.361	0.422	0.432	0.349	0.394
	interval	0.173	0.137	0.173	0.251	0.162	0.132
P10	maximum	0.386	0.373	0.329	0.435	0.328	0.418
	interval	0.141	0.153	0.139	0.197	0.108	0.202
P90	maximum	0.252	0.249	0.195	0.302	0.273	0.241
	interval	0.087	0.086	0.031	0.114	0.111	0.072
Population standard deviation					0.400		

The last comparison made is regarding the Kolmogorov-Smirnov distance between the sample data infilled and the population, which returns the minimal distance between the two distributions. In Figure P03-26 the Kolmogorov-Smirnov distance of the infill optimization for each parametrization while minimizing the SBV function is presented. The results of the Kolmogorov-Smirnov distance are less erratic than the previous two comparisons, regarding the mean and standard deviation. The optimization utilizing the median as the sample values while searching for the optimum obtained the closer overall distributions to the population and are the ones with less erratic behavior. The distance results of the optimizations utilizing the P90 as sample values mostly rise the distance between the sampling and the population, with only 4 optimizations that minimized the original sampling distance to the population distribution.



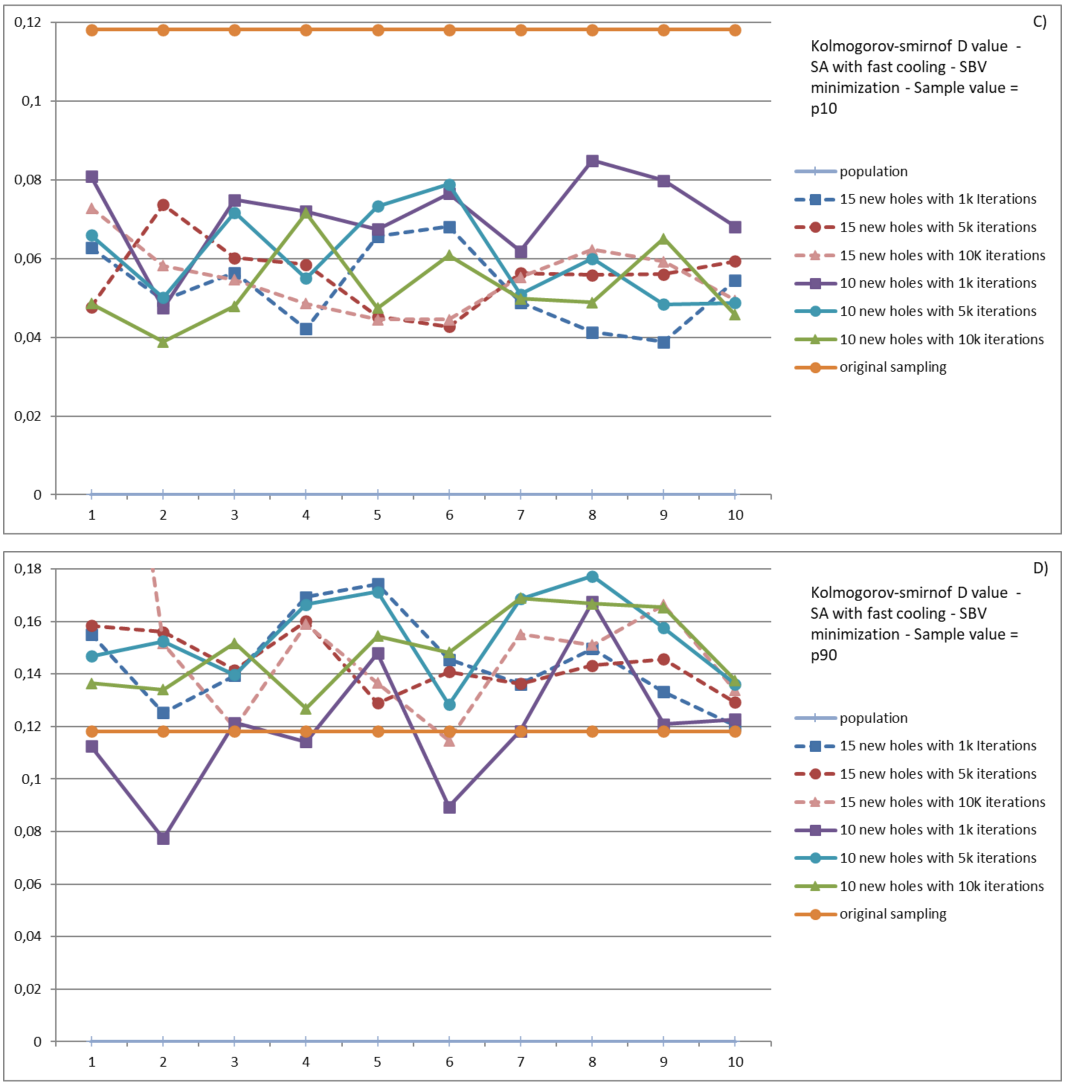
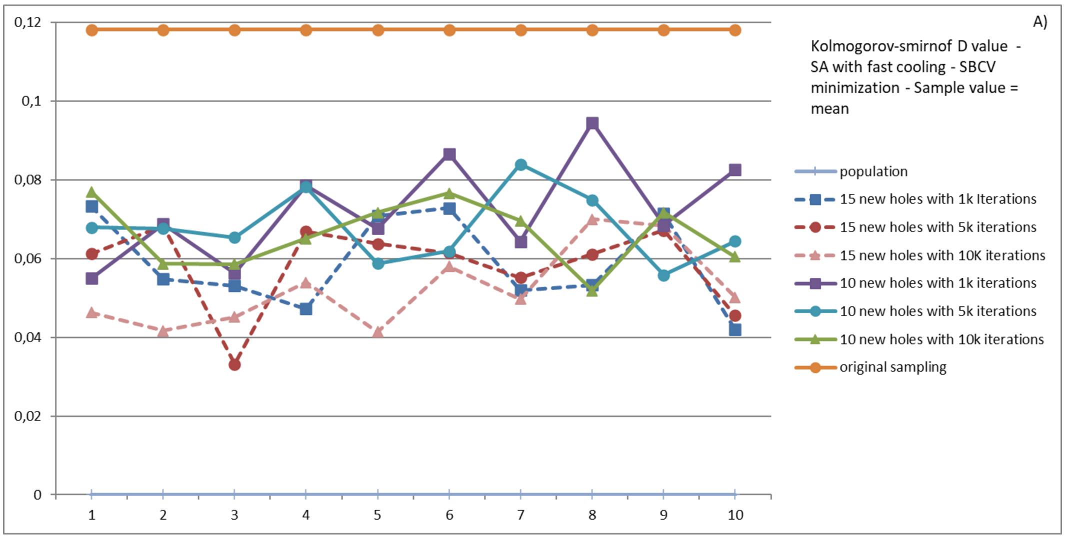
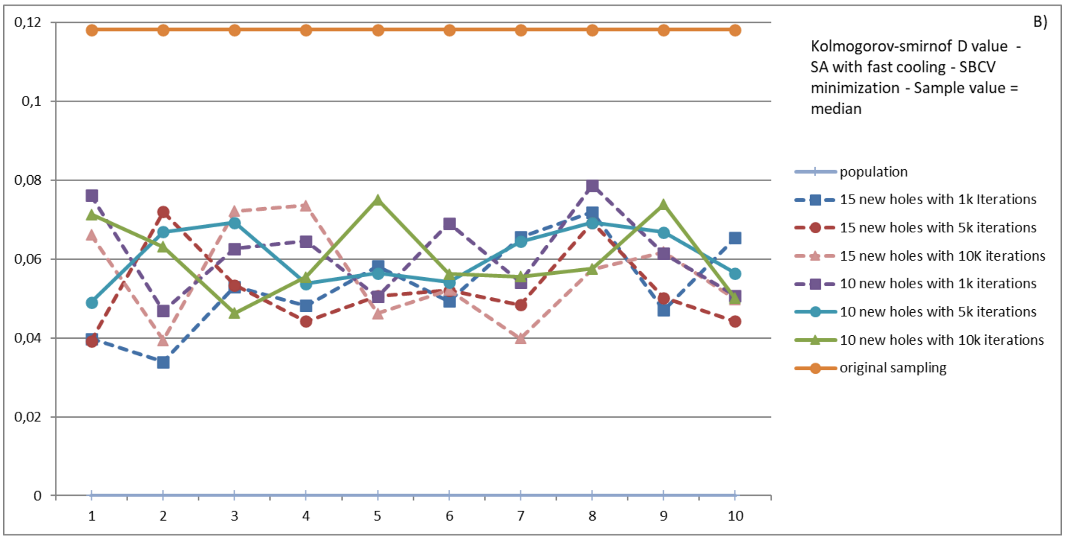
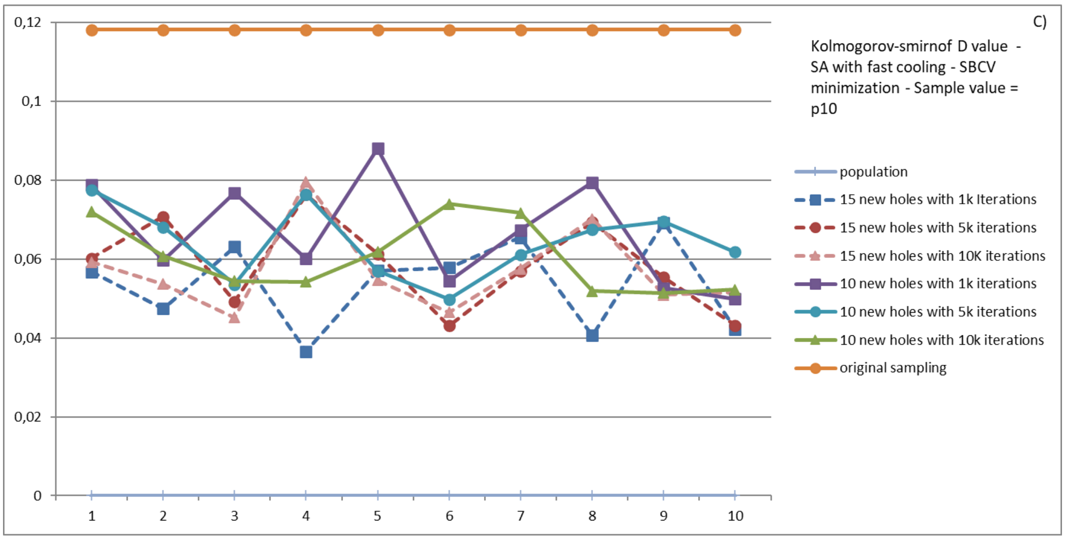


Figure P03-26 – Comparisons between the Kolmogorov-Smirnov distance of 10 optimums infill locations for each search parametrization, number of new collars, and number of iterations, while minimizing the SBV function, and the populational and original sampling standard deviation. The values associated with the infill samples while searching for the optimum were: mean (A); median (B); P10 (C); and P90 (D).

Figure P03-27 compares the Kolmogorov-Smirnov distance between the optimizations minimizing the SBCV function and the population distribution. The minimization of the SBCV returned worse distance values than the ones obtained by minimizing SBV, overall. Both the optimizations utilizing P10 and median fared better concerning the distance, with the median faring slightly better. The optimization considering P90 as sample values once again worsened the original sampling distance to the population, even worse than when the SBV was minimized.


A)
Kolmogorov-smirnof D value - SA with fast cooling - SBCV minimization - Sample value = mean
0,12
0,1
0,08
0,06
0,04
0,02
0
1
2
3
4
5
6
7
8
9
10
population
15 new holes with 1k Iterations
15 new holes with 5k iterations
15 new holes with 10K iterations
10 new holes with 1k iterations
10 new holes with 5k iterations
10 new holes with 10k iterations
original sampling


B)
Kolmogorov-smirnof D value - SA with fast cooling - SBCV minimization - Sample value = median
0,12
0,1
0,08
0,06
0,04
0,02
0
1
2
3
4
5
6
7
8
9
10
population
15 new holes with 1k Iterations
15 new holes with 5k iterations
15 new holes with 10K iterations
10 new holes with 1k iterations
10 new holes with 5k iterations
10 new holes with 10k iterations
original sampling


C)
Kolmogorov-smirnof D value - SA with fast cooling - SBCV minimization - Sample value = p10
0,12
0,1
0,08
0,06
0,04
0,02
0
1
2
3
4
5
6
7
8
9
10
population
15 new holes with 1k Iterations
15 new holes with 5k iterations
15 new holes with 10K iterations
10 new holes with 1k iterations
10 new holes with 5k iterations
10 new holes with 10k iterations
original sampling

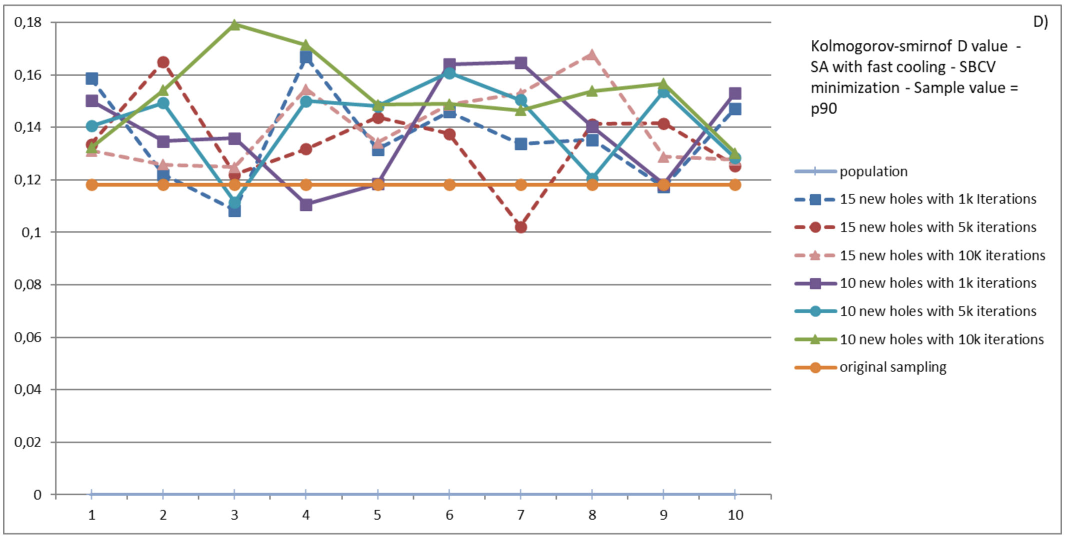


Figure P03-27 – Comparisons between the Kolmogorov-Smirnov distance of 10 optimums infill locations for each search parametrization, number of new collars, and number of iterations, while minimizing the SBCV function, and the populational and original sampling standard deviation. The values associated with the infill samples while searching for the optimum were: mean (A); median (B); P10 (C); and P90 (D).

Table P03-14 presents the comparisons of the Kolmogorov-Smirnov distance to the populational distribution focusing on the maximum and minimum, the interval, and median value of each parametrization. The parametrization with the smallest distance was obtained minimizing the SBV while infilling with 15 new drillholes and 5000 iterations that used the mean as sample values while optimizing. The intervals obtained by minimizing the SBV were higher, but the minimum tends to be smaller, therefore more similar to the population distribution. When considering the interval, minimum, and median, the optimizations that utilized the median as sample values fared better with both objective functions than utilizing the other values.

Table P03-15 – Values of minimum, maximum, interval, and median, of the infill sample distribution Kolmogorov-Smirnov distance, obtained by the 10 optimizations for each different parametrizations considering the 4 sample values utilized while searching for the optimum, mean, median, P10, and P90.

SBV							
	collars	10	10	10	15	15	15
	iterations	1000	5000	10000	1000	5000	10000
Mean	minimum	0.040	0.045	0.058	0.043	0.032	0.043
	interval	0.061	0.059	0.037	0.036	0.035	0.032
Median	minimum	0.041	0.039	0.037	0.037	0.040	0.039
	interval	0.032	0.024	0.033	0.067	0.017	0.015
P10	minimum	0.047	0.048	0.039	0.039	0.043	0.044
	interval	0.037	0.031	0.033	0.029	0.031	0.028
P90	minimum	0.077	0.128	0.127	0.120	0.129	0.115
	interval	0.090	0.049	0.042	0.054	0.031	0.230

SBCV							
	collars	10	10	10	15	15	15
	iterations	1000	5000	10000	1000	5000	10000
Mean	minimum	0.042	0.051	0.043	0.049	0.039	0.041
	interval	0.049	0.026	0.044	0.017	0.047	0.039
Media n	minimum	0.047	0.049	0.046	0.034	0.039	0.039
	interval	0.032	0.020	0.029	0.038	0.033	0.034
P10	minimum	0.050	0.050	0.051	0.037	0.043	0.045
	interval	0.038	0.028	0.023	0.033	0.033	0.034
P90	minimum	0.111	0.111	0.130	0.108	0.102	0.125
	interval	0.054	0.049	0.049	0.058	0.063	0.043

The optimizations utilizing the median as sample values while searching for the optimum obtained the best representation of the population when considering all the comparisons to the different parametrizations applied. Even in the case that the median has not obtained the smallest interval, the results tend to be closer to the populational values. The utilization of the mean as sample values while optimizing tends to return values close to the populational, but those results were the most erratic, with a higher interval overall, especially when the objective function is the SBV minimization. The explication of the worse results systematically obtained when utilizing the P90 as sample values while optimizing comes from the objective function. The original sampling has a negative skewness, with most of the values tending to the minimal of the distribution, 0.5; when the objective function deals with the minimization of the variance of the simulated node, if a higher value is obtained, there is a higher probability that the node variance would rise, not fall as the objective function pretends to do. Therefore, the new samples will tend to the locations with smaller variability possible, i.e., areas of the domain with smaller ore values. This does not occur when the mean or median are utilized as sample values, as these values could rise the overall variance. When an outlier occurs, the mean is affected by its value, but not the median, therefore the median is more robust and obtained better results overall. Regarding the parametrizations, the location of 15 new drillholes fared better than when utilizing 10, but not always the highest number of iterations returned the overall best results, nor the ones with less interval.

6.4 Optimization of infill drillhole while variating the azimuth and dip direction

The variation of direction, contemplating azimuth, and dip was implemented in the algorithm to observe the impact of the direction on the competence in locating infill and to better the population representability of the data. The variation of the azimuth and dip was implemented in the SA-based algorithm in the same procedure as the coordinate variation while searching for optimums. When either the azimuth or the dip is randomly picked to variate in the iterations it shall be obtained in the interval, between a maximum and minimum that the user defines. Therefore, the SA algorithm now can variate 4 collar parameters in the iteration, furthering the possible outcomes of the optimization and possible results poll. To compare both methods, fixed or variate directions, the same parametrization was applied while searching for the optimum infill localization while minimizing the SBV and SBCV functions. Figure P04-28 has presented the comparisons for each objective function and considerations of the direction, with each new drillhole infill optimum being obtained after 1000 iterations. Considering the same objective function, the optimum obtained when variating the drillhole direction systematically obtained the smaller objective function value. This result is expected, as there is a higher possibility that while variating the direction of the drillhole, some of the new drillholes could have more information than only maintaining the drillholes perpendicular to the orebody, then further raising the number of samples. The possibility of a new drillhole economy is visible when one compares the results of the SBV minimization by both procedures, in the instance of 6 drillholes located while variating the drillhole direction fared almost equally as 10 drillholes with fixed direction were located.

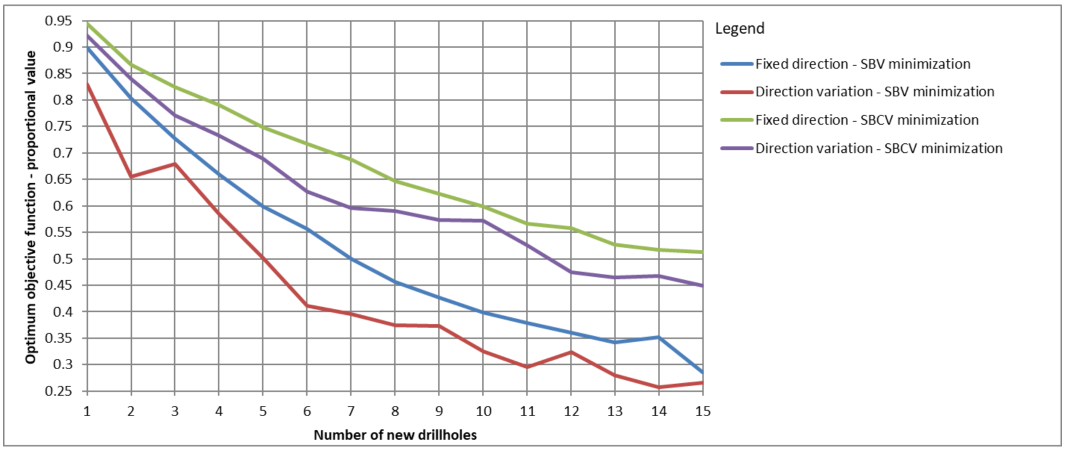


Figure P04-28 – Comparisons between each objective function optimization, obtained after 1000 iterations, under the same parametrizations while utilizing fixed and variable directions of the drillhole while searching for the optimum infill location.

6.5 Final considerations

The synthetic and the real data sets are different, in the sense of the statistical distribution, and also sampling representativity. Therefore, a unique parametrization could not return the best results for both data. Considering the synthetic data set, the SBV objective function was the best of all functions in obtaining more representativity with the optimized infill. For the Capanema data set, the KVS objective function performed best in terms of representativity. This could be explained by the fact that both data sets have different statistical distributions, with the synthetic one having a higher variability. In terms of the composed objective function, neither of the two obtained a better representativity when compared to the other functions.

The value associated with the samples while optimizing changes the results representativity. The tests made in the synthetic data set show that the best value to use was the median of the nearest simulated block. The use of the mean and P10 have satisfactory results too, but with less representativity than the results that utilized the median. The optimizations utilizing the P90 as a value show that in some cases this methodology tends to worsen the original sampling representativity, with most cases of the Kolmogorov-Smirnov distance values being higher than the original sampling. The other statistics compared to show that most optimization results tend to the original sampling value, therefore, do not further the representativity. This could be explained by the fact that the P90 is an outlier on the data set, utilizing it as a sample value would return a higher variance, as opposed to the desired result, and the minimization of the SBV objective function would minimize the variance. There were no tests made in the real data.

Lastly, the optimization considering the collar direction while searching for the optimum shows that the results tend to obtain a higher minimization of the objective function than utilizing a fixed direction while performing the optimization. The optimization test was not made to appoint the competence of this parametrization in furthering the representativity.

7 CONCLUSIONS

The optimization of infill location by the methodologies presented is functional. The SBV minimization is the best at furthering the representativity of the synthetic data while the KBV fared better with the Capanema Mine information. This is probably related to the type of distribution each of those bodies has, the synthetic being a high variability, positive skewness copper ore, and the Capanema iron ore that has a smaller variability, with negative skewness. More tests must be made to check if a distribution with lower variability can be better infilled with the KBV minimization than the others, with the help of synthetic data with the populational value at hand. The compost objective function, which considers the kriged and simulated models to locate the infill fared worse regarding population representativity than the direct functions under the same constraints.

The use of different values associated with the samples while searching for the optimum infill location demonstrates that the median fared better compared to the other values, for the synthetic data. Meanwhile, the P90 systematically fared worse in all comparisons made. The mean as sample value is functional but tends to show more chaotic results, with a higher uncertainty of better representing the populational distribution than the median. Optimization that variates the direction of the drillholes tends to better minimize the objective function, as more data could be found if different directions are made. But further tests must be made to assess the competence of the direction variation while locating infill.

The optimization methods proposed are functional in furthering the representativity regarding the population, but this does not mean that should be used alone. The consideration of different information must be taken while the infill is located. One point of contempt to the methods presented is the economic factor of making new drillholes, i.e., the cost of the new information compensates when considering the profitability that said information returns. This and other information, such as geological information of the ore body, tonnage of ore, and others, should be considered with the methods proposed to better assess the locations that would return the information of more interest to the mining. Therefore, this work proves that the proposed methods are competent in at least helping find the portions of the domain that would provide more representation to the population, and less uncertainty, while the infill location is considered.

REFERÊNCIAS BIBLIOGRÁFICAS

Al-Mudhafer, W. J., 2013, A practical economic optimization approach with reservoir flow simulation for infill drilling in a mature oil field, in North Africa Technical Conference & Exhibition, Cairo, Egypt, Society of Petroleum Engineers14 p., doi:10.2118/164612-MS

Armstrong, M., 1994. Is research in mining geostats as dead as a dodo?; in Geostatistics for the Next Century, Springer, Netherlands, p. 303-312.

Armstrong M., 1998. Basic linear geostatistics. Springer Science & Business Media; New York154pp.

Bianchi L., Dorigo M., Gambardella L. M., Gutjahr, W.J., 2008. A survey on metaheuristics for stochastic combinatorial optimization. Natural Computing, v. 8, no. 2, p. 239-287.

Blum, C., Roli, A., 2003, Metaheuristics in combinatorial optimization: overview and conceptual comparison: ACM Computing Surveys, v. 35, no. 3, p. 268-308.

Boucher, A., Dimitrakopoulos, R., Vargas-Guzman, J. A., 2004, Joint simulations, optimal drillhole spacing and the role of the stockpile, in International Geostatistics Congress, 7th, Banff, Alberta, Canada, p. 35-44.

Boucher, A., Dimitrakopoulos, R. and Vargas-Guzmán, J.A., 2005. Joint simulations, optimal drillhole spacing and the role of the stockpile. In Geostatistics Banff 2004 (pp. 35-44). Springer, Dordrecht.

Chiles, J.P., Delfiner, P., 2009. Geostatistics: modeling spatial uncertainty. John Wiley & Sons. New York, 695pp.

Dantzig, G. B., 1981, Reminiscences about the origins of linear programming, Technical Report SOL 81-5, Systems Optimization Lab, Stanford University, California, 11 p.

Delmelle, E.M. and Goovaerts, P., 2009. Second-phase sampling designs for non-stationary spatial variables. Geoderma, 153(1-2), pp.205-216.

Deutsch, C.V., Journel, A.G., 1992. Geostatistical software library and user’s guide. New York. Oxford University Press, New York, 340pp.

Deutsch D. V. & Journel A. G., 1998. GSLiB – Geostatistical Software Library and User’s Guide. Oxford University Press, New York, 369 p.

Deutsch D. V., 2002. Geostatistical Reservoir Modeling. Oxford University Press, New York, 376pp.

Dirkx, R., Dimitrakopoulos, R., 2018, Optimizing infill drilling decisions using multi-armed bandits: application in a long-term, multi-element stockpile, Mathematical Geosciences, v. 50, no. 1, p. 35-52.

Dutaut, R. and Marcotte, D., 2020. A New Semi-greedy Approach to Enhance Drillhole Planning. Natural Resources Research, 29(6), pp.3599-3612.

Fatehi, M., Asadi Haroni, H. and Hossein Morshedy, A., 2017. Designing infill directional drilling in mineral exploration by using particle swarm optimization algorithm. Arabian Journal of Geosciences, 10(22), pp.1-14.

Fonseca, M.A., 1990. O sinclinal de Ouro Fino: análise descritiva e cinemática de um segmento do sistema Fundão, Quadrilátero Ferrífero, Minas Gerais. Master dissertation, Departamento de Geologia da Escola de Minas, Universidade Federal de Ouro Preto, 120pp.

Gall J. 2014. Simulated Annealing; in Computer Vision: A Reference Guide. Springer, New York, P. 737-741.

Gershon, M., 1987, Comparisons of geostatistical approaches for drillhole site selection, in APCOM 87, Proceedings of the Twentieth International Symposium on the Application of Computers and Mathematics in the Mineral Industries, Volume 3: Geostatistics, Johannesburg, p. 93-100.

Goovaerts P., 1997. Geostatistics for natural resources evaluation. Oxford University Press on Demand. New York, 483pp.

Goria S, Galli A, Armstrong M., 2001. Quantifying the impact of additional drilling on an open pit gold project. In2001 annual conference of the International Association for Mathematical Geology. Cancun, Mexico.

Isaaks, E.H., Srivastava, M.R., 1983. Applied geostatistics. Oxford University Press, New York, 561pp.

Kirkpatrick, S., Gellat, C. D., Vecchi, M. P., 1983, Optimization by simulated annealing, Science, v. 220, no. 4598, p. 671-680.

Luenberger, D. G., Ye, Y., 2008. Linear and nonlinear programming, vol. 2, Springer, New York,549 p.

Lloyd, C.D., Atkinson, P.M., 1999, Designing optimal sampling configurations with ordinary and indicator kriging. InProceedings of the 4th International Conference on GeoComputation

Martínez-Vargas, A., 2017, Optimizing grade-control drillhole spacing with conditional simulation, Minería Y Geología, v. 33, no. 1, p. 1-12.

Massahud, J. S., and Viveiros, J. F. M., 1983, Geology of Capanema iron ore deposit: a review, in Simpósio sobre minério de ferro da Fundação Carl Duisberg. Belo Horizonte, 453 p.

Matheron, G., 1963. Principles of geostatistics. Economic geology, 58(8), pp.1246-1266.

Matheron G., 1971. The theory of regionalised variables and its applications. Les Cahiers du Centre de Morphologie Mathématique. 5, 212pp.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., Teller, E., 1953, Equation of state calculations by fast computing machines, The Journal of Chemical Physics, v. 21, no. 6, p. 1087-1092.

Mohammadi, S. S., Hezarkhani, A., Tercan, A. E., 2012, Optimally locating additional drill holes in three dimensions using grade and simulates annealing, Journal of the Geological Society of India, v. 80, no. 5, p. 700-706.

Nowak, M., Leuangthong, O., 2019, Optimal drill hole spacing for resource classification. In: Mining goes Digital: Proceedings of the 39th International Symposium'Application of Computers and Operations Research in the Mineral Industry (APCOM 2019).

Olea, R.A., 2012. Geostatistics for engineers and earth scientists. Springer Science & Business Media. New York 303pp.

Pardalos, P. M., Mavridou, T. D., 2009, Simulated annealing: AS, in Floudas, C. A., Pardalos, P. M., eds., Encyclopedia of optimization (second edition), Springer, p. 3591-3593.

Pearl, J., 1984, Heuristics: intelligent search strategies for computer problem solving, Artificial Intelligence Series, Addison-Wesley Publishing Company, 382 p.

Pilger, G.G., Costa, J.F., Koppe, J.C., 2001. Additional samples: where they should be located. Natural resources research. Sep;10(3):197-207p.

Pinheiro, M., Emery, X., Rocha, A.M., Miranda, T., Lamas, L., 2017. Boreholes plan optimization methodology combining geostatistical simulation and simulated annealing. Tunneling and Underground Space Technology. 70, 65-75. https://doi.org/10.1016/j.tust.2017.07.003

Pintér, J D., 2009, Global optimization in the analysis and management of environmental systems, in Floudas, C. A., Pardalos, P. M., eds., Encyclopedia of optimization (second edition), Springer, p. 1282-1286.

Ramos, G. Z., 2016, Otimização do infill para redução das incertezas em um depósito sintético de cobre, Dissertação de Mestrado, Instituto de Geociências, Universidade de São Paulo, 86 p.

Rocha, M.M.D., 1999. Estudo comparativo entre os metodos computacionais de avaliacao de reservas e controle de teores da Mina de Capanema, Santa Barbara, MG Master dissertation, Instituto de Geociências, Universidade de São Paulo, 134pp.

Romeijn H. E. 2009. Random Search Methods; in Encyclopedia of Optimization. Springer, New York, p. 3245-3251.

Safa, M., Soltani-Mohammadi, S., 2018, Distance function modeling in optimally locating additional boreholes, Spatial Statistics, v. 23, p. 17-35.

Schäffer, S., 2012, Global optimization: a stochastic approach, Springer series in operations research and financial engineering, 157 p.

Schoen, F., 1991, Stochastic techniques for global optimization: a survey of recent advances, Journal of Global Optimization, v. 1, no. 3, p. 207-228.

Schoen, F., 2009, Stochastic global optimization: stopping rules, in Floudas, C. A., Pardalos, P. M., eds., Encyclopedia of optimization (second edition), Springer, p. 3743-3746.

Silva, D. S. F., Boisvert, J. B., 2013. Infill Drilling Optimization for Maximizing Resource Tonnage. CGC Annual Report, 15, 314/1-314/14

Snyman, J. A., Wilke, D. N., 2018, Practical mathematical optimization: basic optimization theory and gradient-based algorithms, v. 133, Springer, 388 p.

Soltani, S., Hezarkhani, A., Tercan, A. E., 2011, Use of genetic algorithm in optimally locating additional drill holes, Journal of Mining Science, v. 47, no. 1, p. 62-72.

Soltani-Mohammadi, S., Hezarkhani, A., 2013. A simulated annealing-based algorithm to locate additional drillholes for maximizing the realistic value of information. Natural Resources Research, 22(3), pp.229-237.

Szidarovszky, F., 1983. Multiobjective observation network design for regionalized variables. International Journal of Mining Engineering, 1(4), pp.331-342.

Takafuji, E. H. de M., 2015. Estudo comparativo entre a simulação sequencial gaussiana e a simulação baseada em wavelets aplicado a quantificação de minério de CU em um depósito sintético. Master dissertation, Instituto de Geociências, Universidade de São Paulo, 97pp.

Takafuji, E.H. de M., da Rocha, M.M., Ramos, G.Z., Urtiga, L.A., 2017. Simulação estocástica aplicada a um banco de dados sintético de cobre. Geologia USP. Série Científica, 17(2), 247-256.

Van Groenigen, J.W., Siderius, W. and Stein, A., 1999. Constrained optimization of soil sampling for minimization of the kriging variance. Geoderma, 87(3-4), pp.239-259.

Voss S. 2009. Metaheuristics; in Encyclopedia of Optimization. Springer, New York, p. 2061-2075.

Wilde, B., 2009, Minimizing error variance in estimates by optimum placement of samples – a comparison of optimization techniques, Centre for Computational Geostatistics Annual Report 11, p. 406-1 – 406-10.

Yamamoto J. K. 2001. Avaliação e Classificação de Reservas Minerais. EDUSP, FAPESP, São Paulo, 226 p.

Zagré, G.E., Marcotte, D., Gamache, M. and Guibault, F., 2019. New tabu algorithm for positioning mining drillholes with blocks uncertainty. Natural Resources Research, 28(3), pp.609-629.