A Conditional Finite Domain Implementation with LU Simulation

Martha E. Villalba and Clayton V. Deutsch

The bootstrap and spatial bootstrap are traditional tools to evaluate the uncertainty in statistical parameters; however, those techniques do not consider the domain limits and are not conditional to the available data. Conditional Finite Domain base on LU simulation is proposed to evaluate the uncertainty in the mean of the input univariate distribution. The simulation is performed only at the points to be sampled conditioned to the data, the configuration of those sampled points follows the original strategy of sampling; the sampling of many simulated configurations gives different mean values that define uncertainty. A program cfdlu.for was developed to implement this technique and incorporate the conventional and spatial bootstrap algorithms.

1. Introduction

The Conditional Finite Domain (CFD) technique quantifies the uncertainty in the mean. This technique samples from translate and/or rotated configurations of the data to obtain different mean values and assembles a distribution of them. The CFD technique may be considered as an extension of the conventional bootstrap and spatial bootstrap techniques. CFD starts by creating new configurations by random translation and/or random rotation of the data locations throughout the domain. An order k_i, $i = 1,..,K$ of the CFD approach represents sets of realizations or configurations that use reference distributions from the previous order (k_{i-1}). The configurations are simulated with individual reference distributions and the realizations are conditioned to the original data. The sampling of many simulated configurations gives different mean values that define a mean and uncertainty for each order. The uncertainty is stabilized after many configurations and orders are performed.

The total number of realizations is equal to the number of orders required multiplied by the number of configurations. Each realization has n conditioning data and n locations to simulate. The original CFD technique was implemented using sequential Gaussian simulation (sgsim) by (Babak & Deutsch, 2008). The use of sgsim is relatively inefficient because the search and covariance lookup table use the full grid. The LU algorithm permits simulation only at the n locations to be sampled which is more efficient than sgsim.

Simulation is done in Gaussian units, then, a back transformation is performed. The normal scores transformation of the data becomes sensitive to the extrapolation options in the lower and upper tail as the order of simulation increases. Reasonable values must be chosen in the implementation.

2. LU Simulation in CFD

Simulation through LU decomposition of the covariance matrix is a well established multiGaussian simulation technique (Goovaerts, 1997). The LU decomposition provides a fast solution provided there are not too many locations. The CFD technique simulates only at the locations to be sampled, therefore, the CFD technique based on LU decomposition is reasonable. The simulation of a continuous attribute z at N nodes $\mathbf{u}^{(j)}$ conditioned to the data set $\{z(\mathbf{u}_i), i = 1,\ldots,n\}$. The LU simulation approach is strictly made in Gaussian space. The original z values must be transformed to normal scores and the variogram must be calculated with these normal score values.

The covariance matrix between n input data values and N nodes to simulate is as follow:

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} C(\mathbf{u}_1-\mathbf{u}_1) & \cdots & C(\mathbf{u}_1-\mathbf{u}_n) \\ \vdots & \ddots & \vdots \\ C(\mathbf{u}_n-\mathbf{u}_1) & \cdots & C(\mathbf{u}_n-\mathbf{u}_n) \end{bmatrix} & \begin{bmatrix} C(\mathbf{u}_1-\mathbf{u}^{(1)}) & \cdots & C(\mathbf{u}_1-\mathbf{u}^{(N)}) \\ \vdots & \ddots & \vdots \\ C(\mathbf{u}_n-\mathbf{u}^{(1)}) & \cdots & C(\mathbf{u}_n-\mathbf{u}^{(N)}) \end{bmatrix} \\ \begin{bmatrix} C(\mathbf{u}^{(1)}-\mathbf{u}_1) & \cdots & C(\mathbf{u}^{(1)}-\mathbf{u}_n) \\ \vdots & \ddots & \vdots \\ C(\mathbf{u}^{(N)}-\mathbf{u}_1) & \cdots & C(\mathbf{u}^{(N)}-\mathbf{u}_n) \end{bmatrix} & \begin{bmatrix} C(\mathbf{u}^{(1)}-\mathbf{u}^{(1)}) & \cdots & C(\mathbf{u}^{(1)}-\mathbf{u}^{(N)}) \\ \vdots & \ddots & \vdots \\ C(\mathbf{u}^{(N)}-\mathbf{u}^{(1)}) & \cdots & C(\mathbf{u}^{(N)}-\mathbf{u}^{(N)}) \end{bmatrix} \end{bmatrix}$$

The covariance $\mathbf{C}$ is symmetric since $\mathbf{C}_{21} = \mathbf{C}_{12}^{T}$. Also $\mathbf{C}$ is positive definite. A symmetric positive definite matrix has a Cholesky $\mathbf{LU}$ decomposition $\mathbf{C} = \mathbf{L}\,\mathbf{U} = \mathbf{L}\,\mathbf{L}^{T}$. Where the lower triangular matrix has all the elements above the

diagonal as null and the upper triangular matrix has all the elements below the diagonal as null. Then, the decomposition of the large covariance matrix **C** is defined as follow:

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{L}_{11} & 0 \\ \mathbf{A}_{21} & \mathbf{L}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{U}_{11} & \mathbf{B}_{12} \\ 0 & \mathbf{U}_{22} \end{bmatrix}$$

Both $\mathbf{L}_{11}$ and $\mathbf{L}_{22}$ are lower matrices; while, the upper triangular matrices are $\mathbf{U}_{11}$ and $\mathbf{U}_{22;}$ otherwise, the matrices $\mathbf{A}_{21}$ and $\mathbf{B}_{12}$ are not triangular matrices. From the previous expression, the covariance matrix between the data locations is determined by LU decomposition.

$$\mathbf{C}_{11} = \mathbf{L}_{11}\mathbf{U}_{11}$$

Both matrices $\mathbf{B}_{12}$ and $\mathbf{A}_{21}$ are expressed in function of their covariances.

$$\mathbf{C}_{12} = \mathbf{L}_{11}\mathbf{B}_{12} \Rightarrow \mathbf{B}_{12} = \mathbf{L}_{11}^{-1}\mathbf{C}_{12}$$

One may notice that $(\mathbf{U}^{-1})^{T} = (\mathbf{L}^{-1})$ or $(\mathbf{L}^{-1})^{T} = (\mathbf{U}^{-1})$ because they are symmetric matrices. Then, the transpose of $\mathbf{A}_{21}$ matrix is equal to the $\mathbf{B}_{12}$ matrix. These matrices are equal to the product of triangular matrices $\mathbf{L}_{11}^{-1}$ and $\mathbf{C}_{12}$.

$$\mathbf{C}_{21} = \mathbf{A}_{21}\mathbf{U}_{11} \Rightarrow \mathbf{A}_{21} = \mathbf{C}_{21}\mathbf{U}_{11}^{-1} \Rightarrow \mathbf{A}_{21}^{T} = \left(\mathbf{U}_{11}^{-1}\right)^{T}\mathbf{C}_{12} = \mathbf{L}_{11}^{-1}\mathbf{C}_{12}$$
$$\therefore \mathbf{A}_{21}^{T} = \mathbf{B}_{12} = \mathbf{L}_{11}^{-1}\mathbf{C}_{12}$$

The covariance matrix between estimated node locations $\mathbf{C}_{22}$ is equal to the **LU** decomposition $\mathbf{L}_{22}\mathbf{U}_{22}$ only if the covariance matrix between nodes and data $\mathbf{C}_{21}$ is equal to zero. Otherwise, the covariance $\mathbf{C}_{22}$ is derived from the following:

$$\mathbf{C}_{22} = \mathbf{A}_{21}\mathbf{B}_{12} + \mathbf{L}_{22}\mathbf{U}_{22}$$

The product of $\mathbf{A}_{21}$ and $\mathbf{B}_{12}$ is replaced by their equivalent equations derived from the large covariance matrix **C**. Then, the product of the lower triangular matrix $\mathbf{L}_{22}$ and the upper triangular matrix $\mathbf{U}_{22}$ is expressed as follow:

$$if\ \mathbf{A}_{21}\mathbf{B}_{12} = \mathbf{C}_{21}\mathbf{U}_{11}^{-1}\mathbf{L}_{11}^{-1}\mathbf{C}_{12} = \mathbf{C}_{21}\mathbf{C}_{11}^{-1}\mathbf{C}_{12}$$
$$\Rightarrow \mathbf{L}_{22}\mathbf{U}_{22} = \mathbf{C}_{22} - \mathbf{C}_{21}\mathbf{C}_{11}^{-1}\mathbf{C}_{12}$$

The equations for simple kriging and the kriging variance are used to demonstrate that $\mathbf{L}_{22}\mathbf{U}_{22}$ corresponds to matrix of error covariances $\mathbf{K}_{22}$ (Deutsch, 2000). Simple kriging is written as a function of covariances as:

$$Z^{*} = \mathbf{C}_{21}\mathbf{C}_{11}^{-1}Z_{\alpha}$$
$$if\ sk\ system\ \lambda = \mathbf{C}_{12}\mathbf{C}_{11}^{-1} \Rightarrow \lambda^{T} = \mathbf{C}_{21}\mathbf{C}_{11}^{-1}$$
$$\therefore Z^{*} = \lambda^{T}Z_{\alpha}$$

The matrix of error covariances $\mathbf{K}_{22}$ is deduced as follow:

$$E\left\{\left[Z\left(\mathbf{u}^{(i)}\right) - Z^{*}\left(\mathbf{u}^{(i)}\right)\right]\left[Z\left(\mathbf{u}^{(j)}\right) - Z^{*}\left(\mathbf{u}^{(j)}\right)\right]\right\} \forall (i), (j) = 1, \ldots, N$$
$$E\left\{\left[Z\left(\mathbf{u}^{(i)}\right)Z\left(\mathbf{u}^{(j)}\right) - Z\left(\mathbf{u}^{(i)}\right)Z^{*}\left(\mathbf{u}^{(j)}\right) - Z^{*}\left(\mathbf{u}^{(i)}\right)Z\left(\mathbf{u}^{(j)}\right) + Z^{*}\left(\mathbf{u}^{(i)}\right)Z^{*}\left(\mathbf{u}^{(j)}\right)\right]\right\}$$

The simple kriging system $Z^{*} = \lambda^{T}Z_{\alpha}$ is replaced in the previous extended expression.

$$E\left\{\left[Z\left(\mathbf{u}^{(i)}\right)Z\left(\mathbf{u}^{(j)}\right) - \lambda^{T}Z_{\alpha}Z\left(\mathbf{u}^{(i)}\right) - \lambda^{T}Z_{\alpha}Z\left(\mathbf{u}^{(j)}\right) + \lambda^{T}Z_{\alpha}\lambda^{T}Z_{\alpha}\right]\right\}$$

The extended equation of the covariance is split in four terms resulting in the covariance between estimated node locations minus the transpose of the weights matrix times the covariance between data locations and estimated node locations. This result corresponds to $\mathbf{L}_{22}\mathbf{U}_{22}$ and is the decomposition of the $\mathbf{K}_{22}$ matrix of error covariances. The lower triangular matrix $\mathbf{L}_{22}$ is used in algorithms of conditional or non conditional simulation.

$$E\left\{Z\left(\mathbf{u}^{(i)}\right)Z\left(\mathbf{u}^{(j)}\right)\right\} = \mathbf{C}_{22}$$
$$-\lambda^T E\left\{Z_\alpha Z\left(\mathbf{u}^{(i)}\right)\right\} = -\lambda^T \mathbf{C}_{12}$$
$$-\lambda^T E\left\{Z_\alpha Z\left(\mathbf{u}^{(j)}\right)\right\} = -\lambda^T \mathbf{C}_{12}$$
$$+\lambda^T E\left\{Z_\alpha Z_\alpha^T\right\}\lambda = +\lambda^T \mathbf{C}_{11}\lambda$$
$$\Rightarrow \mathbf{C}_{22} - 2\lambda^T \mathbf{C}_{12} + \lambda^T \mathbf{C}_{11}\lambda = \mathbf{C}_{22} - 2\lambda^T \mathbf{C}_{12} + \lambda^T \mathbf{C}_{11}\mathbf{C}_{12}\mathbf{C}_{11}^{-1} \therefore \mathbf{K}_{22} = \mathbf{C}_{22} - \lambda^T \mathbf{C}_{12}$$

Then, a conditional simulation is performed by the product of the **L** matrix by **w** matrix of normal deviates. Also, Y_1 is the column matrix of *n* normal score conditioning data and Y_2 represents the column matrix of the *N* conditional simulated values (Deutsch & Journel, 1998).

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \mathbf{L}\mathbf{w} = \begin{bmatrix} \mathbf{L}_{11} & 0 \\ \mathbf{A}_{21} & \mathbf{L}_{22} \end{bmatrix}\begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \end{bmatrix}$$
$$Y_1 = \mathbf{L}_{11}\mathbf{w}_1 \Rightarrow \mathbf{w}_1 = \mathbf{L}_{11}^{-1}Y_1$$

The vector of the conditional simulation Y_2 is shown as follow. Where, $\mathbf{w}_1$ is replaced by its equivalent expression. The $\mathbf{L}_{22}\mathbf{w}_2$ term symbolizes the error vector and represents an unconditional simulation. Otherwise, the first term is linked with the expression of kriging.

$$Y_2 = \mathbf{A}_{21}\mathbf{w}_1 + \mathbf{L}_{22}\mathbf{w}_2 = \mathbf{A}_{21}\mathbf{L}_{11}^{-1}Y_1 + \mathbf{L}_{22}\mathbf{w}_2$$

Conditional simulations are generated by drawing set of error vectors $\mathbf{L}_{22}\mathbf{w}_2$, that is, by drawing different random number vectors $\mathbf{w}_2$ (Deutsch, 2000). These vectors $\mathbf{w}_2$ are independent normal (0,1) random values that are correlated through the product with $\mathbf{L}_{22}$. As explained before, $\mathbf{L}_{22}$ is the lower triangular matrix of the decomposition of $\mathbf{K}_{22} = \mathbf{L}_{22}\mathbf{U}_{22} = \mathbf{C}_{22} - \mathbf{A}_{21}\mathbf{B}_{12}$, that is, $\mathbf{K}_{22}$ matrix is positive definite and symmetric and their **LU** decomposition is always possible.

3. Methodology

The uncertainty in the mean converges to a constant value after many orders of simulation are performed. Uncertainty in the mean relies on the series of simulations and stochastic algorithms. Thus, similar to the Markov Chain Monte Carlo approach, a period of "burn-in is followed by a period of stabilization where similar fluctuations of the uncertainty around some constant value are observed. The algorithm of CFD is summarized as follows:

- Assemble the representative histogram of the input data; consider the use of declustering method for irregular grid and debiasing method if the scenario requires.
- Set *L* number of configurations through translation and rotation of the data locations.
- k_i, $i = 0$, apply **LU** conditional simulation with the distribution of the original data as reference. Simulation of l_j, $j = 1,...,L$ equiprobable realizations of the variable of interest in *L* new configurations.
- Sample simulated values of each l_j configuration. *L* reference distributions are assembled for the next step.
- k_i, $i = 1$, apply LU conditional simulation with reference distributions established in k_{i-1} to create l_j, $j = 1,...,L$ equiprobable realizations of the variable of interest in *L* configurations and conditioned to the original data.
- Sample simulated values of each l_j configuration, calculate and store the mean of the configuration, the mean of the order and standard deviation of the means.
- Assemble new l_j, $j = 1,...,L$ reference distributions with sampled values of the configurations.
- The last three steps are repeated as a loop until the desired number of *K* orders is reached.

Through the whole process, the mean and variance are stored for every l_j realization and k_i order. The uncertainty in the mean is evaluated from the distribution of the possible means in *L* configurations and *K* orders. Also, the uncertainty of *K* orders could be illustrated in a plot of the standard deviation against the order number to verify convergence. The process of sampling is done from *L* configurations within the study area; hence the Finite Domain part of the name.

Figure 1 shows a small example that illustrates the methodology, where the original distribution in blue is used like reference distributions only in the order zero. Four samples $z(\mathbf{u}_i)$ for i = 1,...,4 are located in a domain of four by four units. The reference distributions that are updated after every order are illustrated in orange.

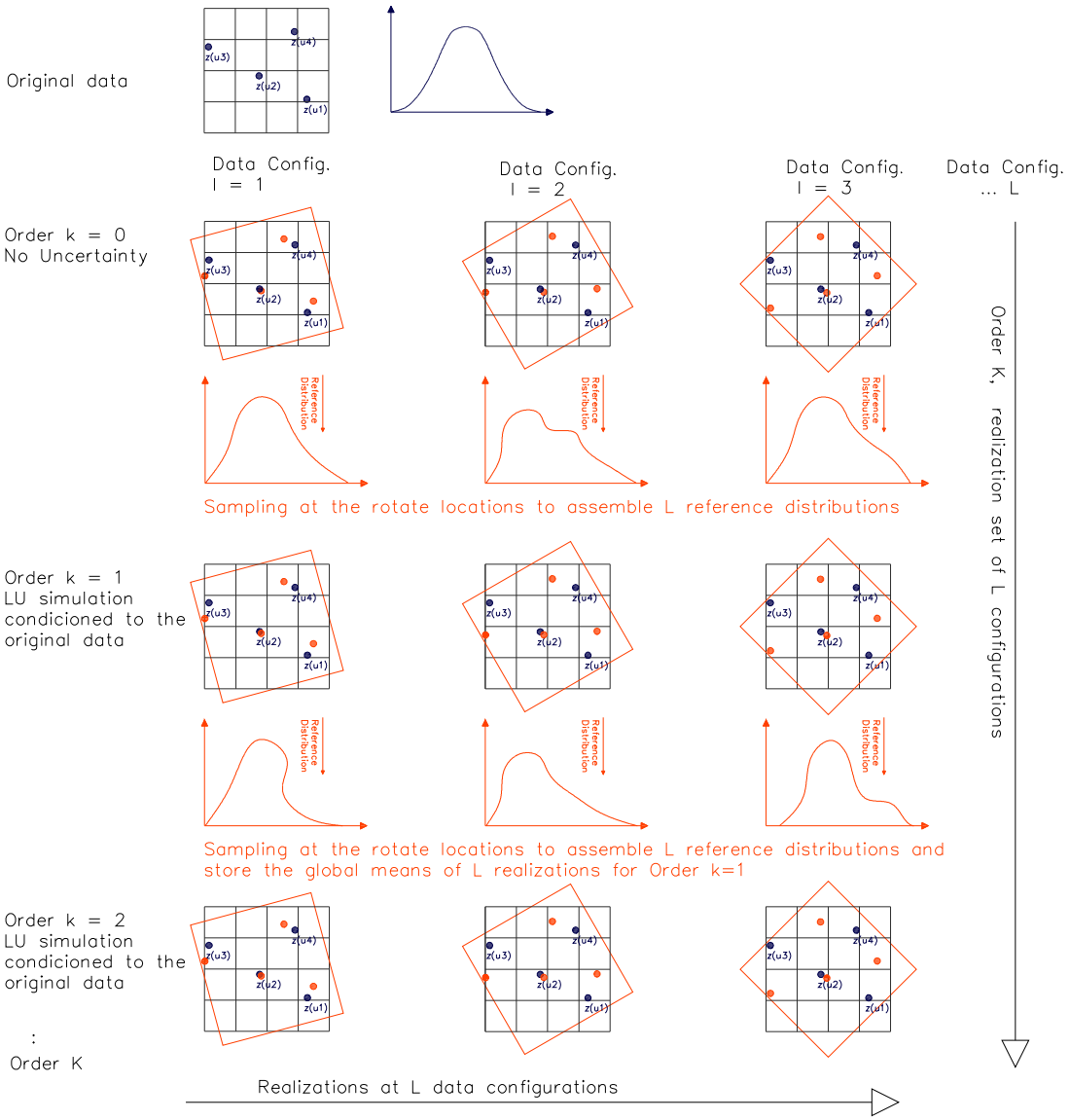


Figure 1: Sketch of the conditional finite domain process, blue points are locations of the conditioning data, orange points are set of new configurations and reference distribution is illustrated by each realization.

The simulations are done at l_j, j = 1,...,L configurations locations conditioned to the original data. The L configurations are the same for every order and CFD technique sample and evaluate uncertainty of the mean in base on the global means of $L \times K$ conditional simulations.

4. Implementation

The data are not equally spaced and preferential sampling is often observed, that is, the histogram should use declustering weights. The discretization of the domain during evaluation ensures a representative global mean; however, the simulation in CFD is performed only at the points to be sampled and conditioned to the original data.

Then, the declustering weights are relevant in the process. The theory about declustering methods is found in (Isaaks & Srivastava, 1989, pp. 235-248) among others.

Normal transformation of the original values is performed $y_i = G^{-1}(p_i)$, $i = 1,...,n$. Specification of the distribution tails are important to obtain a reasonable uncertainty in the input parameter.

The rotated and translated configurations of the original data should sample the whole domain. A representative sampling of the domain is relevant in the process to obtain realistic uncertainty in the mean. Then, this result is transferred to the simulation to improve a posterior evaluation of the global uncertainty and the decision making. The configuration of the data versus the shape of the domain must be checked before implementing CFD. The shape of the domain should permit representative re-sampling. Specific strategies of re-sampling should be done in some scenarios. For instance, sometimes irregular domain shape is observed in some complex deposits. Then, the translation and rotation of the sampling configurations could be restricted to a narrow space between sample locations and the boundary of the domain. **Error! Reference source not found.** shows a graphic of an irregular domain, where the samples $z(\mathbf{u}_i)$, $i = 1,...,6$ are located close to the boundary of the domain.

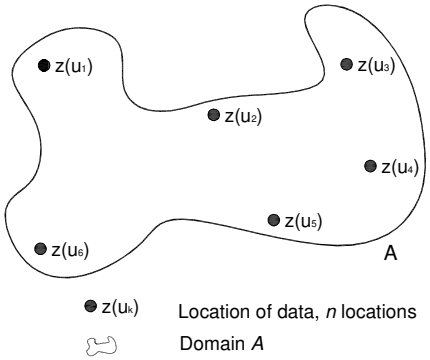


Figure 2: Example of domain that could truncate rotated and translated configurations.

The new configurations are truncated to the domain *A*. Then, all the new strategies of sampling will be close to one another. The uncertainty of the mean would be very small and not realistic.

The uncertainty in the mean using CFD often is greater than conventional bootstrap and spatial bootstrap; however, sometimes conditioning leads to less uncertainty than the SB. This is reasonable since CFD results are more relevant because of the conditioning. An example of this is observed in a domain that shows low variability or the values do not show strong anisotropy. Also, the size of the domain influences on the uncertainty in the mean. Applications of the methodology show that the uncertainty in the mean increases as the size of the domain is expanded with the same available data.

5. Application and Challenges

The methodology is applied and comparison with traditional techniques is evaluated. The influence of the size of domain on the uncertainty in the mean is verified. The data set that we use for this evaluation is *red data*; this file is available in the CCG network. The data have 68 samples of gold, silver, copper and zinc and are illustrated in Figure 3. The thickness of the samples is between 0.13 meter and 18.86 meters. Eight samples with no metal no thickness values were removed. The spatial distance between samples are about 30 meters, this value is used to define the cell size of the declustering process.

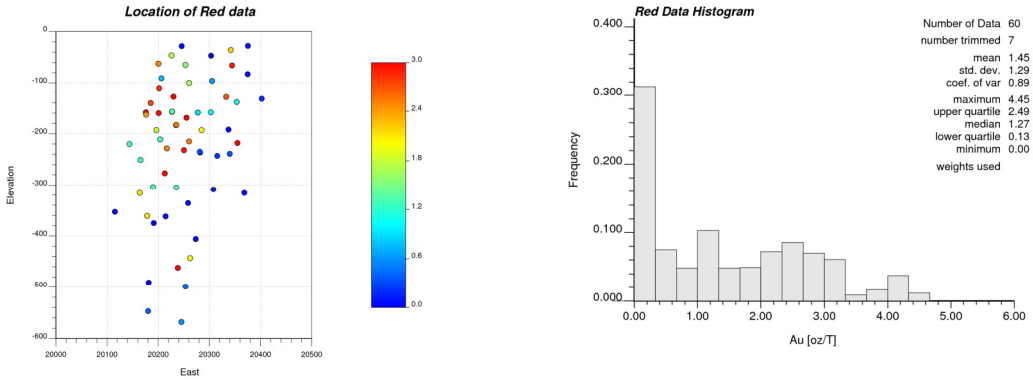


Figure 3: The spatial location of the Red data on the left side and the distribution of the data on the right side.

A variogram model of gold values in normal score units is required because the spatial bootstrap and conditional finite domain work in Gaussian space. Figure 4 shows the experimental and variogram model.

$$\gamma(\mathbf{h}) = 0.44 Exp_{\substack{ah1=100 \\ ah2=90}}(\mathbf{h}) + 0.56 Exp_{\substack{ah1=250 \\ ah2=95}}(\mathbf{h})$$

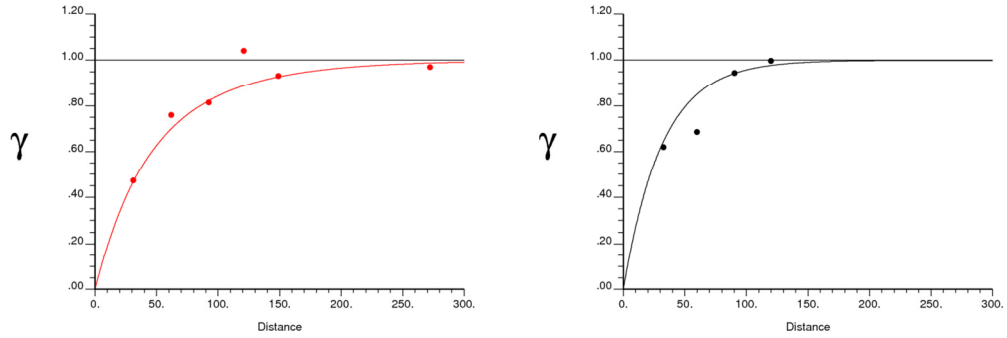


Figure 4: The left exponential variogram model corresponds to the major axis and the right one corresponds to the minor axis.

Two structures were required to model the experimental variogram, *ah1* is the major range in the 15° azimuth and *ah2* is the minor range in perpendicular direction to *ah1*. The program *cfdlu* was developed to apply the methodology. One hundred *L* new configurations are used to sampling the domain. The domain could be defined by polygons in each plane or by specified the space where the centroid of the data spatial location could be displaced.

One hundred orders are performed for this application. For each k_i order l_j, j = 1,...,100 realizations are performed at 60 points conditioned to original data. One thousand realizations are performed, each one provides a global mean and the deviation of these means is the uncertainty in the mean of the univariate input distribution. Conventional Bootstrap and Spatial Bootstrap are set with 1000 realizations as well. **Error! Reference source not found.** shows the shape of the uncertainty in the mean generated by traditional techniques and CFD against the univariate input distribution shape.

The weighted mean of the input distribution is 1.45 g/t and the standard deviation is 1.29, conventional bootstrap obtain a uncertainty in this mean 0.16 equivalent to 10% of the standard deviation of the data, the spatial bootstrap that consider the spatial correlation between the 60 values obtain an uncertainty in the mean 0.32 equivalent to 20% of the standard deviation of the data and finally the methodology of the chapter obtain 0.36 equivalent to the 30% of the standard deviation of the data. The means of 1000 realizations of both CB and SB are almost the same to the mean of the univariate input distribution; however, the mean of 1000 realizations in CFD is a little less because the domain contain almost 30% of the data between the range 0 g/t - 0.25 g/t and those values are located in the limits of the configuration of the data. Extrapolations of these values are made until the limit of the domain is reached.

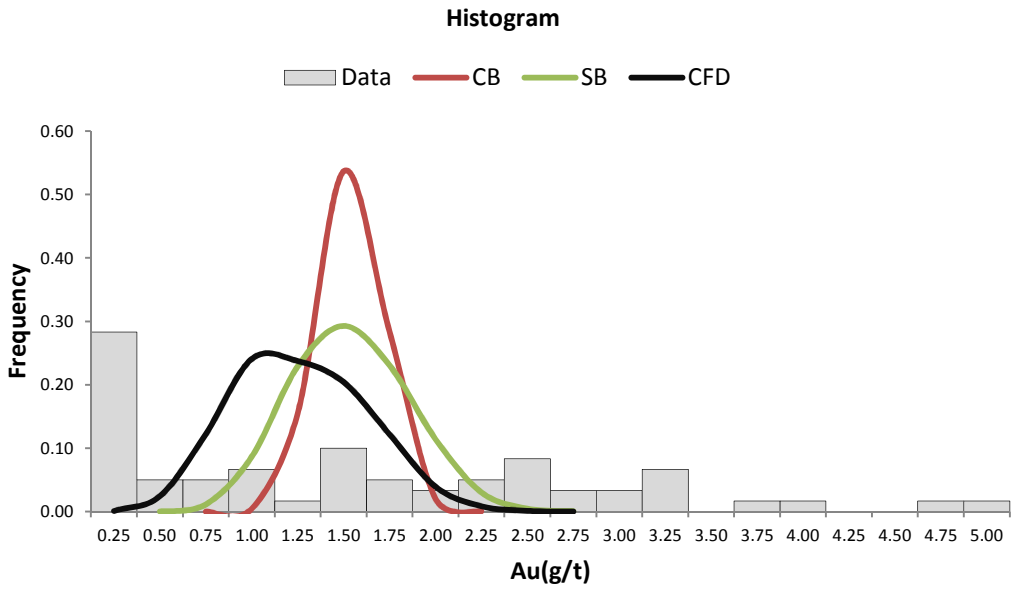
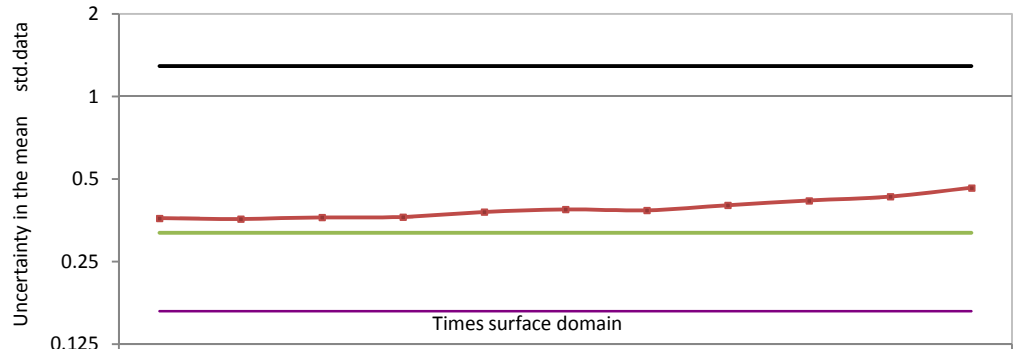


Figure 5: The Histogram of the original Red data is overlapped with three distributions of 1000 means that define the uncertainty in its mean. Distribution in red line is the results of CB, distribution in green line is the results of SB and distribution in black line is the results of CFD.

For the example, we assume the domain *A* of the Red data as a bounded polygon around the data. This domain is expanded 10 times, that is, the 100 configurations will have more area to sampling simulated points conditioned to the original data. Since the sampling configuration is more distant from the conditioning data, an increase of uncertainty is expected because the sampled points are less correlated. Figure 6 shows the impact of change the size of the domain *A* on the evaluation of uncertainty in the mean.



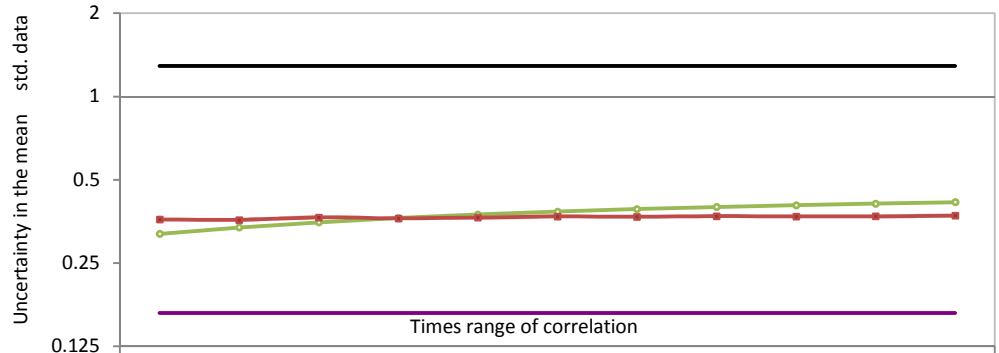
	1	1.1	1.2	1.4	1.6	1.8	2.0	2.2	2.5	2.7	3.0
Data	1.29	1.29	1.29	1.29	1.29	1.29	1.29	1.29	1.29	1.29	1.29
cb	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16
sb	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32
cfdlu	0.36	0.36	0.36	0.36	0.38	0.39	0.38	0.40	0.42	0.43	0.47

Figure 6: Sensitivity analysis of the uncertainty (evaluated by different techniques) with respect to the change of domain size. The standard deviation of the data is illustrated to compare to the uncertainty of the means by CB, SB and CFD techniques.

The uncertainty in the mean (calculated by CFD methodology) increases from 0.36 to 0.47 when the domain is expanded three times. As expected, the CB and SB do not change as the size of the domain is increased. The

sampling both assuming independence and accounting for the spatial correlation among samples are not enough when the domain is expanded without support of new data. All the scenarios require a realistic evaluation of uncertainty and the CFD technique is more robust in domains where the data support only a small percentage of the domain. In practice, zones without support of data are known like potential resources.

CFD techniques looks more robust to the sensitivity analysis of size of domain and range of correlation, the range of correlation should impact the uncertainty in the mean, more correlated samples should give more less uncertainty; however, the SB shows opposite results because as the range of correlation become similar to and bigger than the size of the domain, ergodic fluctuations take place. The relation between size of the domain and range of correlation should be at least more than 10 times to avoid ergodic fluctuations (Leuangthong, McLennan, & Deutsch, 2005) . Figure 7 shows the influence of the range of correlation on the uncertainty in the mean.

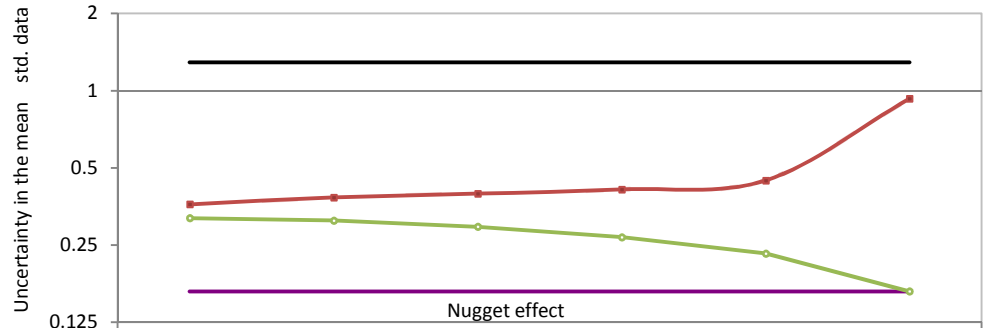


	1	1.1	1.2	1.4	1.6	1.8	2.0	2.2	2.5	2.7	3.0
Data	1.29	1.29	1.29	1.29	1.29	1.29	1.29	1.29	1.29	1.29	1.29
cb	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16
sb	0.32	0.34	0.35	0.36	0.37	0.38	0.39	0.40	0.40	0.41	0.41
cfdlu	0.36	0.36	0.37	0.36	0.36	0.37	0.37	0.37	0.37	0.37	0.37

Figure 7: Sensitivity analysis of the uncertainty with respect to the change of the correlation range. The evaluation of uncertainty (std) in the mean of the input univariate distribution is made by different techniques.

The conventional bootstrap does not change because only the data locations are used to sampling without taking account the spatial correlation among data, their value of uncertainty in the mean 0.16 is the same for all the evaluation of sensitivity analysis showed before. By the other hand, the SB execute 1000 realizations, sample 60 values from the cdf of the input distribution and use the covariance matrix between data to correlated the sampled values. The increase of range of correlation in SB leads to increase in uncertainty from 0.32 to 0.41 because ergodic fluctuations make the results no representatives. The uncertainty with CFD is almost the same from 0.36 to 0.37 when the range of correlation increase until it reaches 3 times the original range.

Sometimes, the variogram will not tend to zero as **h** does. It means that in small distance **h** the result of sampling could be different. Poor analytical precision, poor sample preparation or highly erratic mineralization at low scale could lead to nugget effect C_o (David, 1977). This C_o component of the variogram measure the independence of the data, for instance, when the nugget effect is equal to the variance or to one in variogram with normal score units, the spatial bootstrap results 0.16 equal to conventional bootstrap because the data become independent. The increase of nugget effect indicates less correlation between samples, and then an increase of uncertainty in the mean is expected. The conditional finite domain gives an uncertainty that rises as the samples become less correlated. Figure 8 shows the sensitivity analysis of the uncertainty in the mean with respect to the change of nugget effect.



	0.0	0.2	0.4	0.6	0.8	1.0
Data	1.29	1.29	1.29	1.29	1.29	1.29
cb	0.16	0.16	0.16	0.16	0.16	0.16
sb	0.32	0.31	0.29	0.27	0.23	0.16
cfdlu	0.36	0.38	0.40	0.41	0.45	0.93

Figure 8: Sensitivity analysis of the uncertainty with respect to the change of the nugget effect. The evaluation of uncertainty (std) in the mean of the input univariate distribution is made by different techniques.

The CFD shows more robustness in many scenarios; however, many parameters are required to set the methodology. Also, many realizations at different configurations and orders could be computationally intensive with big set of data even if CFD simulate only in the points to be sampled. An evaluation of uncertainty in the mean should be simple as CB and SB and robust as CFD.

6. Conclusions

The Conditional Finite Domain (CFD) technique samples from rotated and translated configurations of the data to obtain different mean values. The configurations are simulated with different reference distributions, but are conditioned to the same original data. The sampling of many simulated configurations gives different mean values that define uncertainty for every iteration level or order. The uncertainty is stabilized after many configurations and orders are performed.

The original conditional finite domain technique was implemented using sequential gaussian simulation (sgsim) by (Babak & Deutsch, 2008). The use of sgsim is relatively inefficient because the search and covariance lookup table use a full grid. This paper proposes the use of LU simulation. The LU algorithm simulates only at the *n* locations to be sampled which is more efficient than sgsim.

The *n* locations correspond to configurations that are created by random translation and rotation of the data locations limited to some domain. The *order* in CFD is defined as the series of simulation for each of the configurations. LU simulation is executed for each configuration and the reference distribution is conditioned to the original data. The reference distribution is taken from the previous order for each configuration. The uncertainty in the mean is taken from the final realizations.

A sensitivity analysis demonstrated robustness and reasonableness of this approach in scenarios where other techniques struggle with unrealistic uncertainty. An increase in the nugget effect, a reduction of the range of correlation or an increase in the size of the domain leads to more uncertainty, which is reasonable.

7. References

Babak, O., & Deutsch, C. V. (2008). Reserves Uncertainty Calculation Accounting for Parameter Uncertainty. *Canadian Petroleum Technology, 8*, 41-49.

Chilès, J.-P., & Delfinier, P. (1999). *Geostatistics: Modeling Spatial Uncertainty.* New York: John Wiley & Sons.

David, M. (1977). *Geostatistical Ore Reserve Estimation.* Netherland: Elsevier Scientific Publishing Company.

Deutsch, C. V. (2000). *Special Topics in Geostatistics.* Alberta: University of Alberta.

Deutsch, C. V. (2002). *Geostatistical Reservoir Modeling.* New York: Oxford University Press, Inc.

Deutsch, C. V. (2004). A statistical Resampling Program for Correlated Data, Spatial Boostrap. *Six Annual Report of the Centre for Computational Geostatistics.* Edmonton: Department of Civil & Environmental Engineering University of Alberta.

Deutsch, C. V., & Journel, A. (1998). *GSLIB: Geostatistical Softare Library and User's Guide* (2nd Edition ed.). New York: Oxford University Press.

Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics, 7*, 1-26.

Goovaerts, P. (1997). *Geostatistics for Natural Resources Evaluation.* New York: Oxford University Press.

Isaaks, E. H., & Srivastava, R. M. (1989). *An introduction to Applied Geostatistics.* New York: Oxford University Press.

Journel, A. G. (1994). Resampling from Stochastic Simulation. *Environmental and Ecological Statistics, 1*, 63-91.

Leuangthong, O., Khan, K. D., & Deutsch, C. V. (2008). *Solved Problem in Geostatistics.* Hoboken, New Jersey.: John Wiley & Sons, Inc.

Leuangthong, O., McLennan, J., & Deutsch, C. V. (2005). Acceptable Ergodic Fluctuations and Simulation of Skewed Distributions. *Seventh Annual Report of the Centre for Computational Geostatistics.* Edmonton: Department of Civil&Environmental Engineering-University of Alberta.

Montgomery, D. C. (2000). *Design and Analysis of Experiments Fifth Edition.* New York: John Wiley & Sons, Inc.

Pyrcz, M., & Deutsch, C. (2002). Two artifacts of probability field simulation. *Mathematical Geology, 33*, 775-800.

Strunk Jr., W., & White, E. (2009). *The Elements of Style* (Fiftieth Anniversary ed.). New York, United States: Pearson Education, Inc.

Wang, F., & Wall, M. M. (2003). Incorporating Parameter Uncertainty into Prediction Intervals for Spatial Data Modeled via Parametric Variogram. *Agricultural, Biological and Environmental Statistics.*, 296-309.