REDUCING ERROR IN SHORT TERM GRADE PREDICTION INCLUDING IMPRECISE AND INACCURATE DATA

*C.P.Araújo [1], J. F.C.L. Costa[1], V.C.Koppe [1] and A.Soares[2]

[1] *FEDERAL UNIVERSITY OF RIO GRANDE DO SUL-UFRGS*

9500, Bento Gonçalves Avenue

Porto Alegre, Brasil, 91.509-900

*(*Corresponding author: cristinapaixaoaraujo@yahoo.com.br)*

[2] *INSTITUTO TECNICO DE LISBOA-IST*

1, Rovisco Pais Avenue

Lisboa, Portugal, 1049-001

24th World Mining Congress
MINING IN A WORLD OF INNOVATION
October 18-21, 2016 • Rio de Janeiro /RJ • Brazil

REDUCING ERROR IN SHORT TERM GRADE PREDICTION INCLUDING IMPRECISE AND INACCURATE DATA

ABSTRACT

Proper mining requires correct decisions on each block destination, i.e. an extracted block should go to the mill or to the waste dump. To help this decision, short term mine planning requires further sampling to reduce prediction uncertainty. Commonly the practice of estimation keeps only hard data, i.e., data considered precise and accurate verified by a quality control program. In a worst case scenario practitioners combine these hard data with soft data that are imprecise and in some circumstances biased. In this paper, direct sequential cosimulation is used to integrate secondary imprecise, biased data (soft data) into short term mine planning to update grades block model. Direct sequential cosimulation models are used to assess the uncertainty and using hard data against the model derived from using both hard and soft data, the last standardized to filter the bias. The results show the benefit of incorporating soft data after its bias correction. A case study illustrates the method.

KEY WORDS

Direct sequential cosimulation; integration data, update reserves

1- INTRODUCTION

Sampling is a continuous process along the life of the mine. Uncertainty in mine planning can be reduced by adding more samples. Frequently, sampling methods produce samples derived from distinct methods with different quality and quantity of samples. During exploration sampling is carried out by diamond drill cores (more common), which are expensive and known to produce accurate and precise results. At this stage, there are few data with high quality and they are herein referred as hard data.

Later during quasi mining more samples are available. Frequently these samples may have been collected along several campaigns, prepared using different protocols or analyzed at different laboratories. In General these informations do not have the same quality, therefore, they constitute different statistical populations. At this late stage there are many data with poor quality if compared to diamond drill cores. These samples are named soft data.

The idea here is evaluate uncertainty reduction and the benefits in enhancing resource classification of resources compared to the results using only hard data, accessing error by stochastic simulation and cosimulations. The results from different methodologies were checked considering precision for the different produced model.

2- METHODOLOGY

Three methods were chosen to evaluate the uncertainty of block grades given imprecise and biased soft data combined with hard data, namely: Direct sequential simulation (Soares, 2001) (DSS) using only hard data; Direct sequential simulation (DSS) with hard and soft data with bias and imprecision correction; Direct Sequential Cosimulation (CoDSS) with hard and soft data.

The main advantage of the proposed algorithm (DSS) is that it allows the simulaton/cosimulation without calling for any transformation of the original variables. The idea is to use the simple kriging estimated local mean and variance, not to define the local cdf but to sample from the global cdf.

2.1- Direct Sequential Simulation

Let us consider the continuous variable $Z(x)$ with a global cdf $F_z(z) = \text{prob}\{Z(x) < z\}$ and stationary variogram $\gamma(h)$.The intention is to reproduce both $F_z(z)$ and γ (h) in the final simulated maps.

The direct sequential simulation algorithm of a continuous variable follows the classical methodological sequence:

1. Define a random path over the entire grid of nodes x_u, $u = 1, N_s$, to be simulated.
2. Estimate the local mean and variance of $z(x_u)$, identified, respectively, with the simple kriging estimate $z(x_u)^*$ and estimation variance $\boldsymbol{\sigma^2_{SK}}$ (x_u) conditioned to the experimental data $z(x_i)$ and previous simulated values $z^s(x_i)$.
3. Define the interval of $F_z(z)$ to be sampled, by using the Gaussian cdf:

$$G\,(y\,(x_u)^*, \boldsymbol{\sigma^2_{SK}}\,(x_u)), \text{ where } (\boldsymbol{\varphi_1}(z\,(x_u)^*).$$

4. Draw a value $z^s(x_u)$ from the cdf $F_z(z)$.
• Generate a value p from a uniform distribution $U(0, 1)$;
• Generate a value y^s from $G\,(y\,(x_u)^*, \sigma^2_{SK}\,(x_u))$: $y^s = G^{-1}\,(y\,(x_u)^*, \sigma^2_{SK}\,(x_u), p)$;
• Return the simulated value $z^s_l(x_u) = \varphi_1^{-1}\,(y^s)$.
5. Loop until all Ns nodes have been visited and simulated.

2.2- Direct Sequential Cosimulation

Instead of simulating N_v variables simultaneously, each variable is simulated in turn conditioned to the previously simulated variable (Gomez-Hernandez, Jaime and Journel, 1993; Goovaerts, 1997). Suppose just two variables, $Z_1(x)$ and $Z_2(x)$. Choosing the primary variable, say $Z_1(x)$, as the most important or with a more evident spatial continuity (Almeida & Journel, 1994), the joint simulation algorithm is described in detail as follows:

1. Define a random path visiting each node of a regular grid of nodes.

2. At each node x_u Simulate the value $x_l^s(x_u)$ using the DSS algorithm described in step 2 above:
• Identify the local mean and variance of z_1 (x_u) as the SK estimate and estimation variance z_1 $(x_u)^*$ and σ_{SK}^2 (x_u); calculate $y\,(x_u)^* = \varphi_1(z\,(x_u)^*)$, φ_1 being the normal score transform of the primary variable $z_1\,(x)$;
• Generate a value p from a uniform distribution $U\,(0,\,1)$;
• Generate a value y^s from $G\,(y\,(x_u)^*, \sigma_{SK}^2\,(x_u))$: $y^s = G^{-1}\,(y\,(x_u)^*, \sigma_{SK}^2\,(x_u), p)$;
• Return the simulated value $z_l^s(x_u) = \varphi_1^{-1}\,(y^s)$ of the primary variable.

The same DSS algorithm is applied to simulate $Z_2(x)$ assuming the previously simulated $Z_1(x)$ as the secondary variable. Colocated simple cokriging is used to calculate $z_2(x)^*$ and σ_{SK}^2 (x_u) conditioned to neighborhood data $z_2(x_\alpha)^*$ and the colocated datum z_1 (x_u) (Goovaerts, 1997):

$$Z_2(x_u)_{SCK=}^* \sum_{\alpha=1}^{N} \lambda_\alpha\,(x_u)[z_2(x_\alpha) - m_2] + \lambda_\beta(x_u)[z_1(x_u) - m_1] + m_2 \qquad \textbf{(1)}$$

• Transform $y\,(x_u)^* = \varphi_2(z_2(x_u)^*)$. φ_2 is the normal score transform of the $Z_2(x)$ variable.
• Generate a value p from a uniform distribution $U\,(0,\,1)$;
• Generate a value y^s from G $(y_2\,(x_u)^*, \sigma_{SK}^2\,(x_u))$: $y^s = G^{-1}\,(y_2\,(x_u)^*, \sigma_{SK}^2\,(x_u), p)$;
• Return the simulated value $z_2^s\,(x_u) = \varphi_2^{-1}(y^s\,)$ of the secondary variable.
3. Loop until all nodes are simulated.

2.3- Standardized data: Proposal for filtering the bias and imprecision error in soft data

From geostatistical view this difference in precision and accuracy data has to be considered for integrating the two data types.

For building the model, where the hard and soft data were pooled together, it was used a correction factor to correct the mentioned bias in secondary (soft) data. This workflow can be used for situations where hard and soft data are strongly correlated. Innitally, soft data ($Z_2(u_{\alpha 2})$) are standardized (Equation 2) using the mean (m_2) and the standard deviation (σ_2) of the soft data. The transformation using Equation 2 leads to a zero mean and an unity standard deviation in the transformed data.(Minnitt & Deutsch, 2014)

$$Z_2(u_{\alpha 2})^* = \frac{[Z2(u_{\alpha 2}) - m_2]}{\sigma_2} \qquad \textbf{(2)}$$

Next, the soft standardized data $Z_2(u_{\alpha 2})^*$ are rescaled to match the hard data statistics (Equation 3) using their mean (m_1) and standard deviation (σ_1). Thus, the mean for the hard and soft data would now match.

$$Z_2(u_{\alpha 2})_T^* = Z_2(u_{\alpha 2})^{**}\sigma_1 + m_1 \qquad \textbf{(3)}$$

It is important to evaluate the data, when was created the variable "soft data with bias correction" (global correction), does not guarantee that the corrected values are within the limits "realistic"(positive values). For example, if the distribution content starts at zero, the corrected values may theoretically be negative.

3- CASE STUDY

3.1- Data Presentation

This study uses the exhaustive Walker Lake dataset (Isaaks & Srivastava, 1989) with 78 000 point support samples distributed regularly at 1 × 1 m (V_Ref_Points),. The variable V was used and the original unit was rescaled so that it resembled grades from a copper mineral deposit. To obtain the reference block grade distribution (V_Ref_blocks), the exhaustive point support dataset was averaged into 3120 blocks of size 5 × 5 m. These blocks represent the true block grades and were used for comparison.

In this case study, the data set was adapted (Araujo, 2015) from the original ones. Two types of data were considered. First, point samples were obtained regularly spaced at 20 × 20 m (V_20x20). These samples were precise and accurate and mimick diamond drillhole samples (hard data). Next soft data (secondary samples) were obtained sampling the exhaustive dataset at a of 5 × 5 m regular grid where imprecision and bias were added. Note that the direct cosimulation was performed, assuming soft data are known every node of the grid (collocated cosimulation) For this, ordinary kriging was carried out in order to obtain exhaustive soft data (V_5X5_+25%_Exhaustive). Figure 1 shows the regularly spaced samples used in this case study.

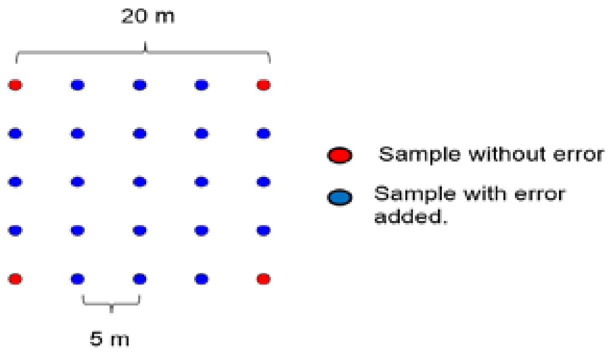


Figure 1- Data set with regular spaced samples

Table 1 shows the summary statistics for the reference point support dataset (V_Ref_points), the reference block grade distribution (V_Ref_blocks), and the sample dataset with accurate and precise data (V_20×20). The sample datasets have their mean values very close to the true mean, which indicates that there were no biases or imprecision. The data with bias and imprecision (V_5×5_+25% and V_5X5_+25%_Exhaustive) has a mean and standard deviation 25% greater than those of the reference point distribution (V_Real_points) to mimic the situation in which poor-quality data induce biases that are subsequently transferred to the grade estimation process.

Table 1- Summary statistics for the original reference and for the biased and imprecise soft data

Data	Nº of samples	Mean	Standard Deviation	CV	Minimum	Maximum
V_Ref_points	78000	2.78	2.50	0.90	0.00	16.31
V_Ref_blocks	3120	2.78	2.49	0.89	0.00	15.68
V_20X20	195	2.73	2.43	0.89	0.00	10.13
V_5X5_+25%	2925	3.44	3.12	0.90	0.00	18.30
V_5X5_+25%_Exhaustive	78000	3.44	3.12	0.90	0.00	18.30

Figure 2 shows the cross correlogram between hard data and soft data. The samples present moderated correlation (0.60). This moderate correlation can be caused by different sampling techniques or by distinct preparation protocols, which lead to possible measurement errors formed by laboratory analytical error plus the sum of all other sampling preparation errors.

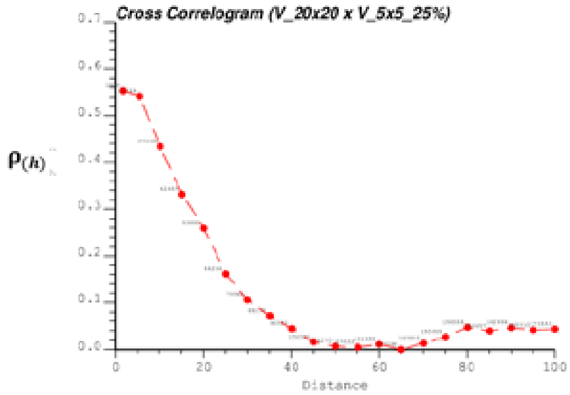


Figure 2- Omnidirectional cross correlogram for hard data and soft data

Equation 4 shows the variogram used for simulating using the different methodologies: Direct sequential simulation and Direct sequential cosimulation. In the cosimulation, for simplicity, the collocated cokriging was applied with the Markov-type approximation (Goovaerts, 1997), i.e., only the hard data variogram and the correlation coefficient between hard and soft data is needed. For the spatial continuity, the major direction was defined as 157.5º, the minor direction as 67.5º and a spherical (Sph) variogram model. For each methodology, 50 realizations were simulated.

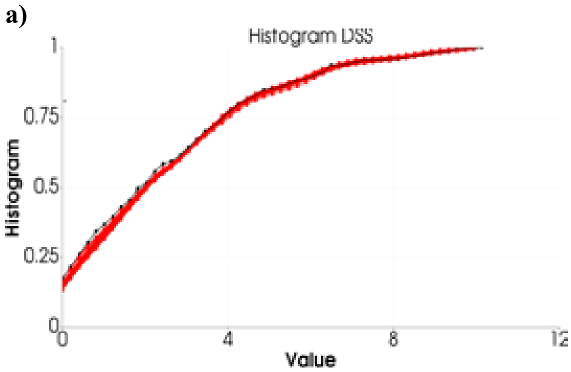
$$\gamma_V(\mathbf{h}) = \mathbf{1.0 + 2.0 \cdot Sph(1) \cdot \left(\frac{N157.5E}{36\ m}, \frac{N67.5E}{16\ m}\right) + 2.92 \cdot Sph(2) \cdot \left(\frac{N157.5E}{84\ m}, \frac{N67.5E}{40m}\right)} \quad \mathbf{(4)}$$

4- RESULTS AND DISCUSSION

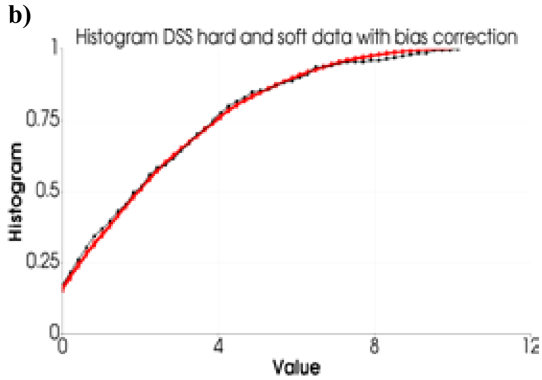
4.1 Validation

In Figure 3 the black line represents the declustered hard data (V_20X20) cumulative histogram and the red lines are the realization histograms. Figure 3a shows the histograms of Direct sequential simulation with only hard data whilst figure 3b the histograms of Direct sequential simulation with hard and soft data with bias correction. The plots show good statistics reproduction. Figure 3c shows histogram reproduction for Direct sequential cosimulation with hard and soft data. The results depart from the hard data statistics when soft data was used with their error embedded. The difference is more evident around the upper quartile (Q3), where cumulative probabilities are overestimated when compared against situations a and b. In this last case, the bias and imprecision were transferred to realizations.

a)



b)



c)

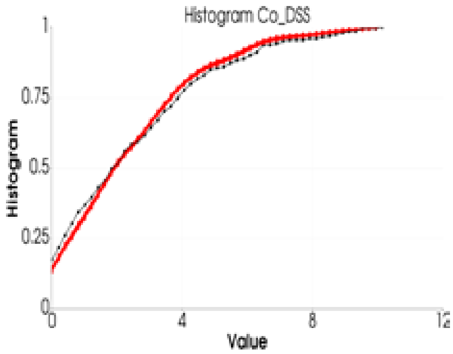
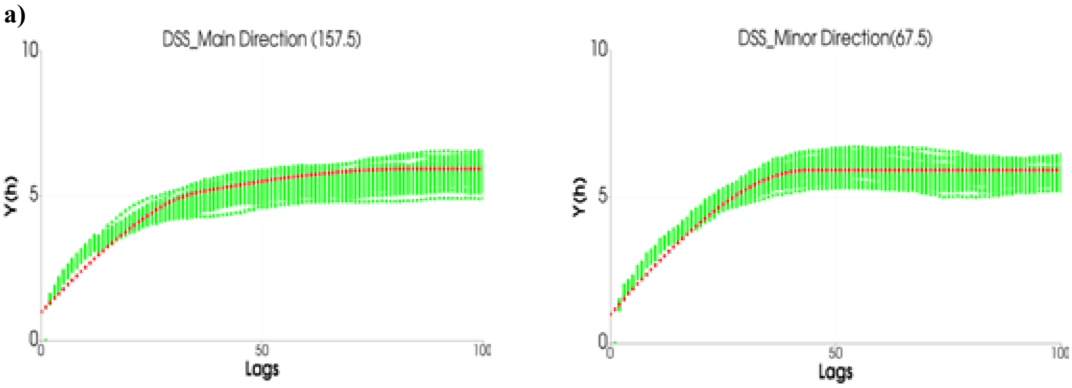


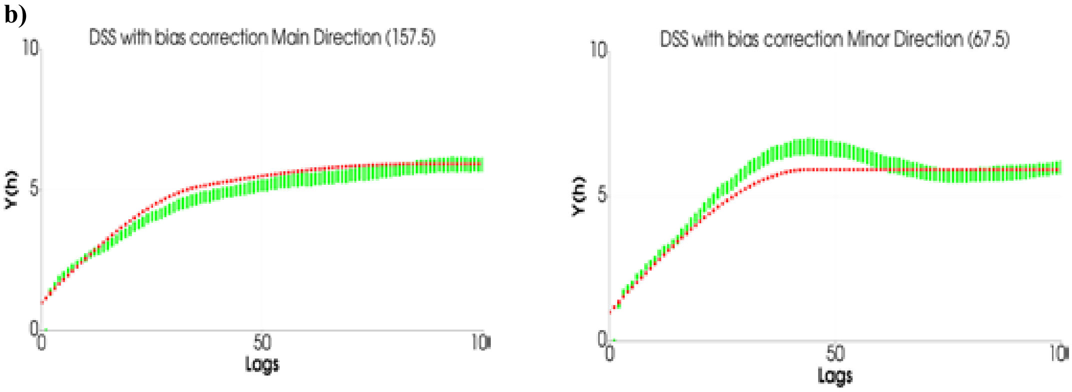
Figure 3- Histogram of simulated models compared against the hard data histogram a) Direct Sequential Simulation with only hard data b) Direct Sequential Simulation with hard and soft data with bias correction c) Direct Sequential Cosimulation with hard and soft data

Figure 4 shows the variograms reproduction by the models. Red line represent the input variogram model (V_20X20) and green lines the ones derived from the realizations at the major and minor directions.

a)



b)



c)

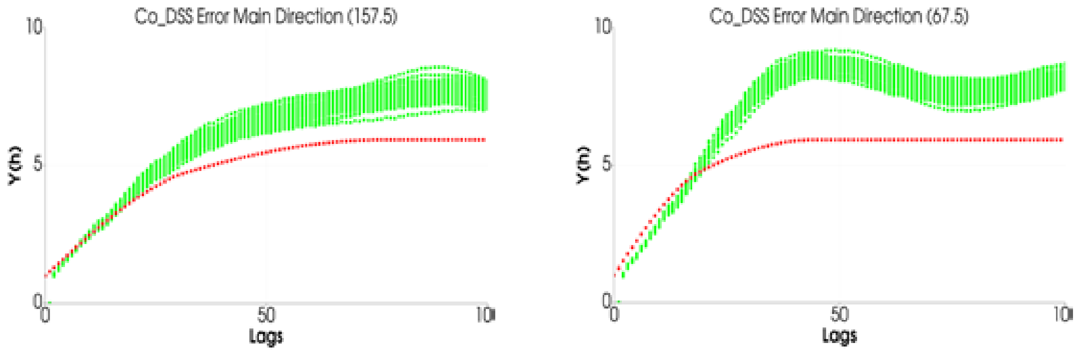


Figure 4- Variograms reproduction by the various models) Direct Sequential Simulation with only hard data b) Direct Sequential Simulation with hard and soft data with bias correction c) Direct Sequential Cosimulation with hard and soft data

For evaluating the correlation reproduction between hard and soft data by cosimulated models figure 5 shows the scatterplot between realization_4 and realization_20 obtained by direct cosimulation using both hard and soft data versus the reference block model, the correlation coefficient is 0.58, which is close to correlation between hard and soft data.

a) **b)**

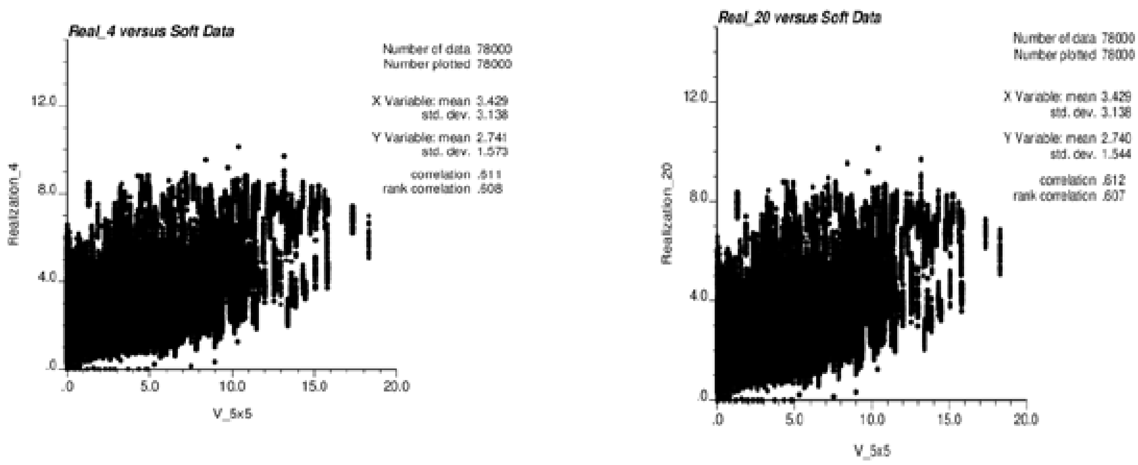


Figure 5-Scatterplot of the point simulated grid in Direct sequential cosimulation with hard and soft data a) Realization_4 and b) Realization_20 versus point references of soft data (ρ = 0.60). This is plotted point simulated grid versus point references

. **4.2-Reducing block misclassification**

The simulations and reference model were reblocked into 5x5 m blocks (Table 1). To calculate the real error (equation 5) it was used the E-type model and the reference block model [V_ (Ref_blocks)] which represents the true block values.

$$Error_{Real} = \frac{[E\_type - V_{\text{Ref\_blocks}}]}{E - type} \quad (5)$$

Figure 6 shows real error when the methodologies are compared. For Direct cosimulation with hard and soft data (brow line) the error is lower, it means,more precise when compared against Direct simulation using only hard data (blue line). Direct sequential simulation model with hard and soft data with bias correction (yellow line), has more blocks with less error and it is more precise. In this methodology, when the soft data was corrected using the bias and imprecise error, , more data was used to conditioning the realizations and the results tend to be closer to the reference block model, increasing influence of the soft data in the realizations.

Figure 6- Real error in block using the reference block model for comparison between the methodologies a) Direct Sequential Simulation with only hard data b) Direct Sequential Simulation with hard and soft data with bias correction c) Direct Sequential Cosimulation with hard and soft data

Equation 6 shows how error was obtained using the interquartile range and E-type from the simulations at blocks 5x5 meters (other approach of the error in Li, Dimitrakopoulos, Scott, & Dunn, 2004, for instance):

$$Error_{Calculated} = \frac{(Q_{95} - Q_5)/2}{E\text{-}type} \quad (6)$$

where the interquartile range $Q_R=(Q_{95} - Q_5)/2$, measures spread of the block values and E-type approximates the estimated value for the block.

Figure 7 shows the calculated block error used as a criterion to classify the blocks with uncertainty within 10% to 90% interval. In the Direct Sequential Simulation considering only hard data, all the blocks had uncertainty larger than 30%, due to the few and sparse data. In the Cosimulation, the results were better if compared to Direct Sequential Simulation using only hard data, as the correlation between hard and soft data is moderate. Using Direct sequential simulation with hard and soft data with bias correction, and considered as hard data afterwards, improved the results and decreased the uncertainty at every block.

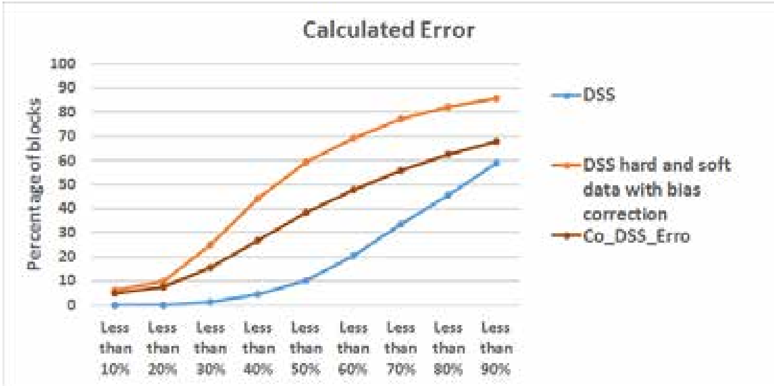


Figure 7- Calculated error of block using the reference block model for comparison between the methodologies a) Direct Sequential Simulation with only hard data b) Direct Sequential Simulation with hard and soft data with bias correction c) Direct Sequential Cosimulation with hard and soft data

5- CONCLUSION

Direct sequential simulation and Cosimulation presented in this case study use the original variable without requiring any priory or posterior transformation. In general, these algorithms showed a good reproduction of univariate and bivariate statistics and spatial continuity model of the data.

In the case of using few hard data (precise and accurate) withr Direct sequential simulation there is a clear loss of precision in the derived model.

For using soft (imprecise and inaccurate) data integrated with hard data it was proposed two methodologies: Direct sequential collocated cosimulation combining either hard and soft data. Hard and Soft data exhibit moderate correlation which means no significant weights will be assigned to the second ones when estimating/simulating the primary variable. An alternative chosen used using Direct Sequential Simulation with hard and soft data after bias correction. This led to the best results showing that the soft data may improve short term geological modelling whether an appropriate methodology is used to include these data (after correction). Thus, when the bias and imprecise error was filtered from the soft data, more data was used in the simulations and realizations were closer to the reference block model. Consequently, the real error and calculated error in the block values were lower.Therefore, it is advantageous to incorporate imprecise measurements for stochastic simulations using the adequate methodology.

6- ACKNOWLEDGMENTS

The authors would like to thank the Coordination for Higher Education Staff Development (CAPES) for supporting the Mineral Exploration and Mining Planning Research Unit (LPM) at the University Federal do Rio Grande do Sul.

7- REFERENCES

Almeida, A. S., & Journel, A. G. (1994). Joint simulation of multiple variables with a Markov-type coregionalization model. *Mathematical Geology*, *26*(5), 565–588. http://doi.org/10.1007/BF02089242

Araújo, C. P & Costa, J.F.C.L. (2015). Integration of different-quality data in short-term mining planning. *Rem: Revista Escola de Minas, 68(*2), 221–227.http:// doi.org/10.1590/0370-44672015680212

Gomez-Hernandez, Jaime and Journel, A. . (1993). Joint Sequential Simulation of MultiGaussian Fields. In A. Soares (Ed.), *Geostatistics Tróia`92* (Vol. 1, pp. 85–94). Kluwer Academic Publishers.

http://doi.org/10.1007/978-94-011-1739-5

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

Isaaks, E. H., & Srivastava, R. M. (1989). *Applied Geostatistics*. New York: Oxford University Press, Inc. http://doi.org/10.1016/0040-1951(74)90006-7

Li, S., Dimitrakopoulos, R., Scott, J., & Dunn, D. (2004). Quantification of Geological Uncertainty and Risk Using Stochastic Simulation and Applications in the Coal Mining Industry. In *Orebody Modelling and Strategic Mine Planning* (pp. 22–24). Perth.

Minnitt, R. C. A., & Deutsch, C. V. (2014). Cokriging for optimal mineral resource estimates in mining operations, *114*(March), 189–203.

Soares, A. (2001). Direct Sequential Simulation and Cosimulation. *Mathematical Geology*, *33*(8), 911–926. http://doi.org/10.1023/A:1012246006212