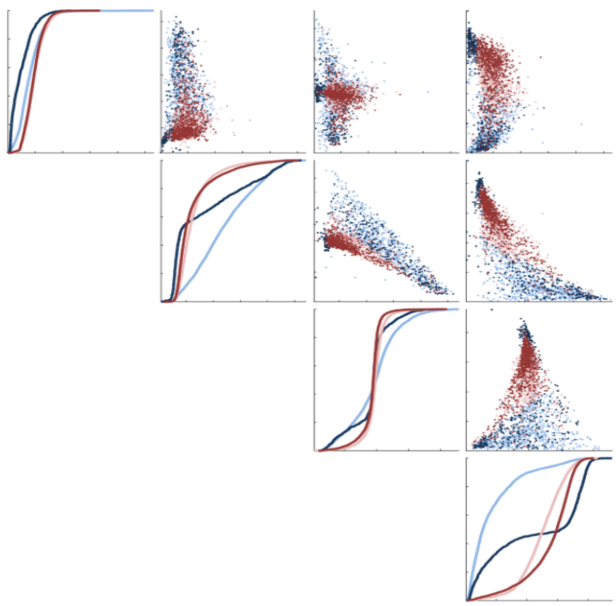
CCG Centre for Computational Geostatistics Solutions

Multivariate Modeling for Resources and Geometallurgy



March 18, 2016

Prepared by Ryan M. Barnett

Summary

- Technical decision making and process evaluation for mine projects is often dependent on multivariate relationships that exist between geometallurgical variables. It is therefore important for geostatistical models to reproduce those relationships.
- Conventional geostatistical algorithms (cosimulation, min./max auto-correlation factors, etc.) are based on the assumption that the covariance matrix fully characterizes multivariate relationships. Multivariate complexities that exist between geometallurgical variables cause this assumption to fail, meaning that the multivariate relationships will not be reproduced.
- To address this problem, the CCG has developed the Projection Pursuit Multivariate Transformation (PPMT), which transforms data of virtually any size and form to be multivariate Gaussian and uncorrelated. The transformed variables can then be independently simulated, before the back-transform restores the original multivariate relationships, including any complexity that may exist.
- Using a mining case study, the PPMT workflow is demonstrated to be simpler than conventional multivariate simulation workflows, while yielding superior results that can be expected to improve mine technical decision making and process evaluation.

1 The Setting

Evaluating the process performance of mining operations requires numerical models of many related geological variables, such as resource variables, contaminant variables, processability variables, etc. Taken together, they provide a characterization of the geologic deposit that forms the basis for engineering design and decision making. As an example, consider a Ni laterite deposit, which is frequently processed with an electro-arc furnace. Several auxilary variables must be modeled in addition to the Ni resource, including Fe, SiO2 and MgO. As the joint values of these variables dictates plant performance, effective geostatistical models should reproduce their multivariate relationships. For example consider the SiO2-MgO relationship in Figure 1.

It is critical that this complex relationship is reproduced, as the SiO2-MgO ratio (SMR) of the furnace feed impacts the process heat. Exceeding a critical SMR threshold will likely damage the furnace lining.

2 The Problem

Conventional multivariate simulation assumes that the data is multivariate Gaussian (multiGaussian). A multiGaussian distribution is schematically represented in the right panel of Figure 2, where the relationship follows elliptical density contours that are fully characterized by covariance. Unfortunately, geological data is rarely multiGaussian. Instead, geological data will likely contain multivariate complexities, such as those shown in the remaning panels. Note that all of the complexities that are schematically illustrated are present in the SiO2-MgO relationship.

Common transformations such as the normal score transform (Verly, 1983), principal component analysis (PCA) (Davis & Greenes, 1983), and min./max auto-correlation factors (MAF) (Desbarats & Dimitrakopoulos, 2000) do not remove these multivariate complexities. As a result, conventional multivariate simulation algorithms will not generate realizations that match the data distribution. Consider the normal score transform of SiO2 and MgO in Figure 3, which yields variables that are univariate Gaussian, but not multiGaussian. Using cosimulation generates multiGaussian realizations (blue) that match the correlation of the data, but not the distribution. After back-tranforming (bottom right panel), the original data distribution is not reproduced.

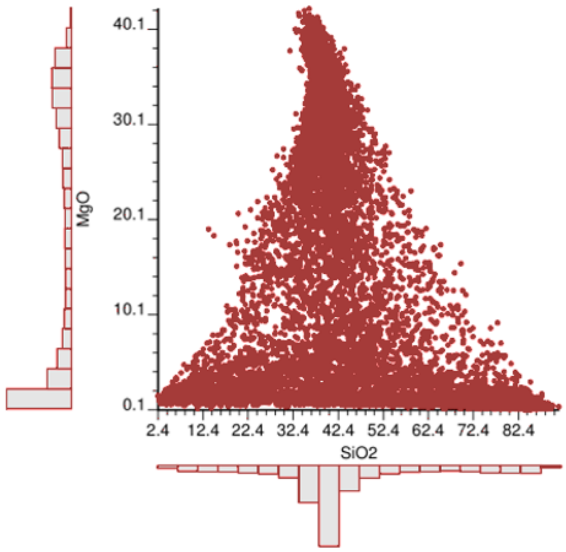


Figure 1: Cross-plot of SiO2 and Mgo.

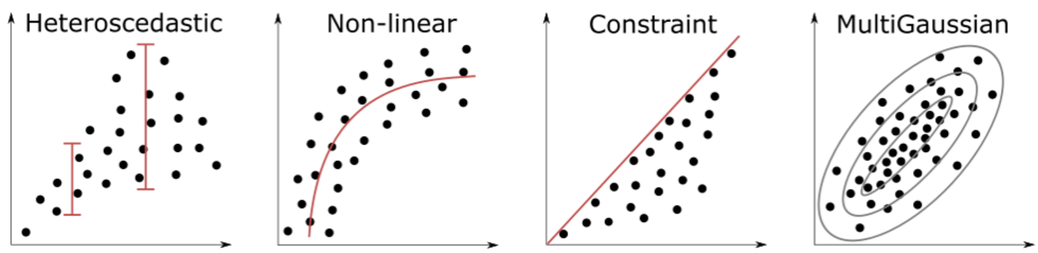


Figure 2: Schematic illustrations of multivariate complexities and a multiGaussian distribution.

3 The Solution

To address the described problem issue, a variety of techniques have beed developed by the CCG for transforming variables to be multiGaussian (Barnett et al., 2014; Leuangthong & Deutsch, 2003). Conventional Gaussian simulation algorithms may then be used with data that matches their assumptions, before using the associated back-transformations to return the original complexity to simulated realizations. Many of these transformations will also decorrelate the variables so that modeling is simplified to independent simulation. Associated back-transformations are then used to return the original correlation to simulated realizations. This workflow is illustrated using the Ni laterite variables in Figure 4.

Observe that the complex data is transformed to be bivariate Gaussian and uncorrelated (top right). As a result, independently simulation yields realizations that match the distribution of the transformed data (bottom right). Back-transformation of the realizations reintroduces the original complexity and correlation (bottom left). Since the SiO2-MgO is critical to reproduce, results of the above workflow would lead to improved operational planning (relative to the conventional covariance based workflow). As will be demonstrated, this result is achieved using a workflow that is also easier to implement than the conventional workflow.

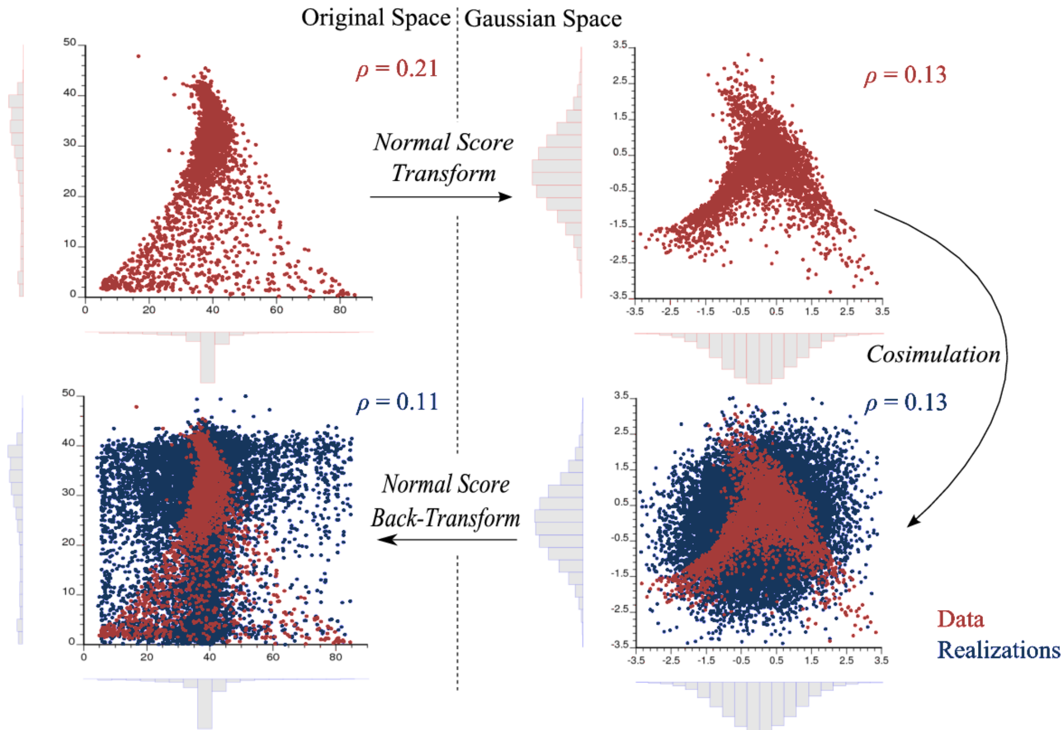


Figure 3: Cosimulation results with complex multivariate data.

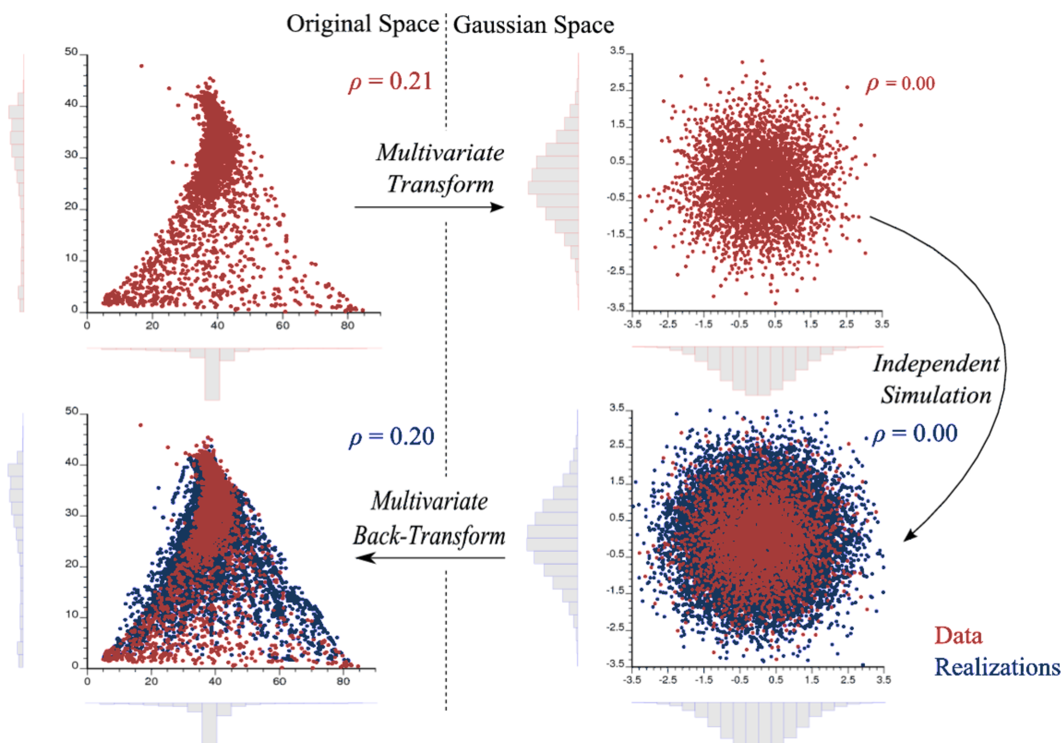


Figure 4: Multivariate transformation results with complex multivariate data.

4 Projection Pursuit Multivariate Transform

The Projection Pursuit Multivariate Transform (PPMT) (Barnett et al., 2014; Friedman & Tukey, 1974) was recently developed for transforming data of virtually any size and form to be uncorrelated and multiGaussian. Details of the PPMT will now be summarized. It should be emphasized that although the PPMT involves several steps, they are all accomplished within a single program.

Consider the data as a matrix $\mathbf{Z} : z_{\alpha i}, \alpha = 1, ..., n, i = 1, ..., K$. The univariate and multivariate properties of the Z_1 and Z_2 variables that are used for demonstration appear in Figure 5; the spatial properties appear in Figure 6. Note the complex multivariate features of the data and differing spatial continuity of each variable; it will be important to reproduce these features in geostatistical modeling.

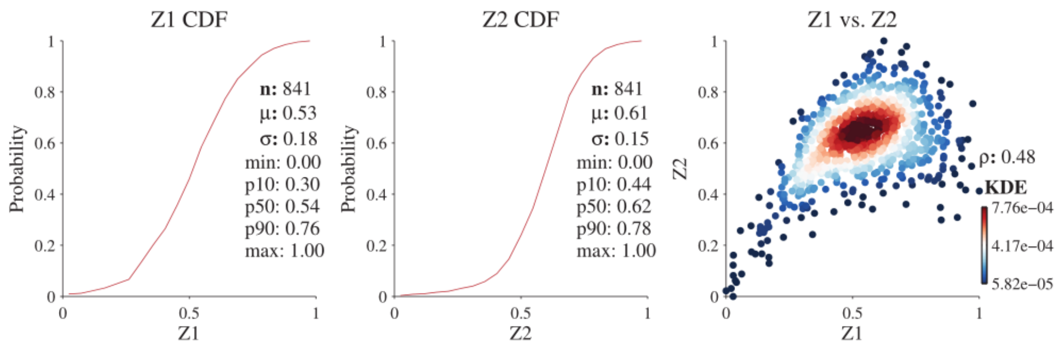


Figure 5: CDFs and KDE scatterplot of the variables.

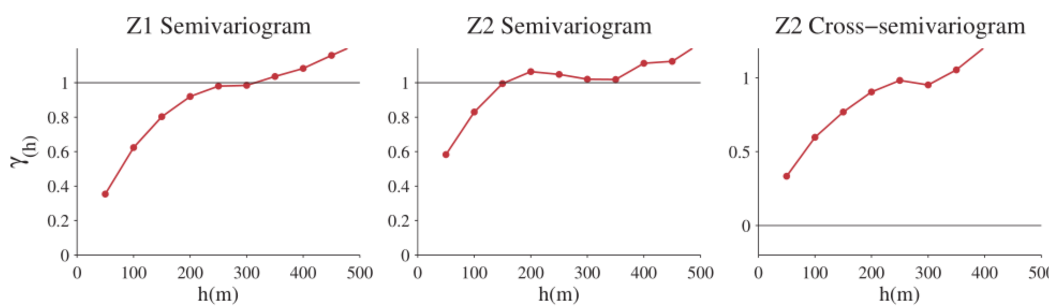


Figure 6: Semivariograms and cross-semivariogram of the variables.

Pre-processing

The first step of the PPMT applies the normal score transform to all of the variables according to:

$$y_{\alpha i} = G^{-1}\left(F_i(z_{\alpha i})\right), \text{ for } \alpha = 1, ..., n, i = 1, ..., K$$

This transforms the variables to univariate standard Gaussian, providing properties that will benefits subsequent steps. The 2-D data are normal scored in Figure 7, as is evident from their univariate statistics.

The second step of the PPMT applies data sphereing (Fukunaga, 1972) to all of the variables according to:

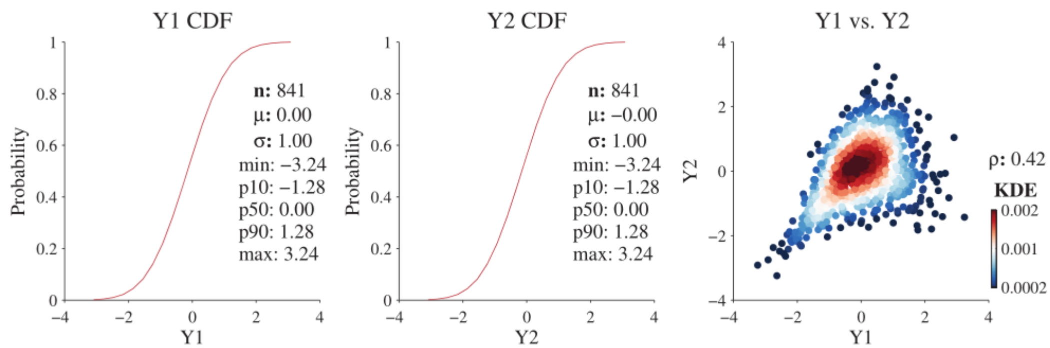


Figure 7: CDFs and KDE scatterplot of the normal score transformed variables.

$$\mathbf{X} = (\mathbf{Y} - E\{\mathbf{Y}\})\mathbf{S}^{-1/2},\ \mathbf{S}^{-1/2} = \mathbf{V}\mathbf{D}^{-1/2}\mathbf{V}^{\mathsf{T}}$$

This transforms the variables to be uncorrelated with unit variance, which is required for applying the next step of the PPMT. The 2-D data are sphered in Figure 8, as is evident from their correlation of zero and variance of one.

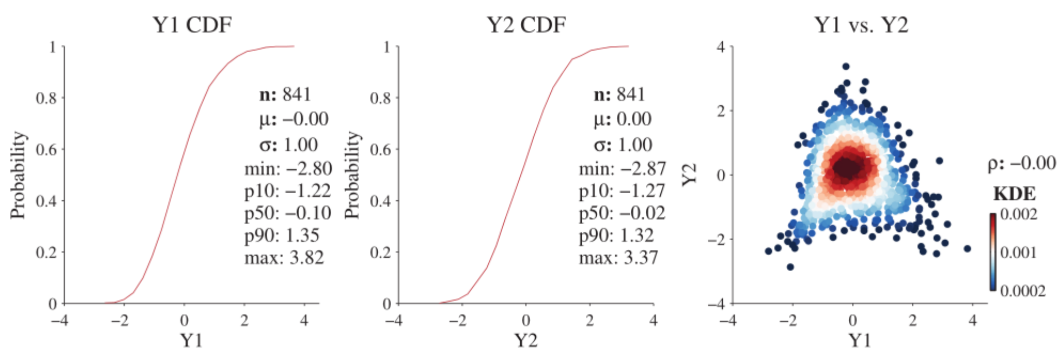


Figure 8: CDFs and KDE scatterplot of the sphere variables.

Projection Pursuit

Consider a Kx1 unit length vector, $\boldsymbol{\theta}$, and the associated projection of the data upon it, $\mathbf{p} = \mathbf{X}\boldsymbol{\theta}$. Any $\boldsymbol{\theta}$ should yield a $\mathbf{p}$ that is univariate Gaussian if $\mathbf{X}$ is multiGaussian. With this in mind, define the projection index, $I(\boldsymbol{\theta})$, as a test statistic that measures univariate non-Gaussianity. For any $\boldsymbol{\theta}$ where the associated $\mathbf{p}$ is perfectly Gaussian, $I(\boldsymbol{\theta})$ is zero.

The PPMT performs an optimized search to find the $\boldsymbol{\theta}$ that maximizes $I(\boldsymbol{\theta})$. Once found, the multivariate data, $\mathbf{X}$, is transformed so that its associated projection, $\mathbf{p} = \mathbf{X}\boldsymbol{\theta}$, is made standard normal Gaussian. This search and normalize procedure is repeated until $\mathbf{X}$ is made multiGaussian. This process is referred to as projection pursuit, and is demonstrated in Figure 9.

Following the 25th projection pursuit iteration, the data are transformed to be uncorrelated and multiGaussian. According to Gaussian model definition, this also renders the data independent. The variables can therefore be simulated independently, before inverting the described transforms to return the Gaussian realizations to original space.

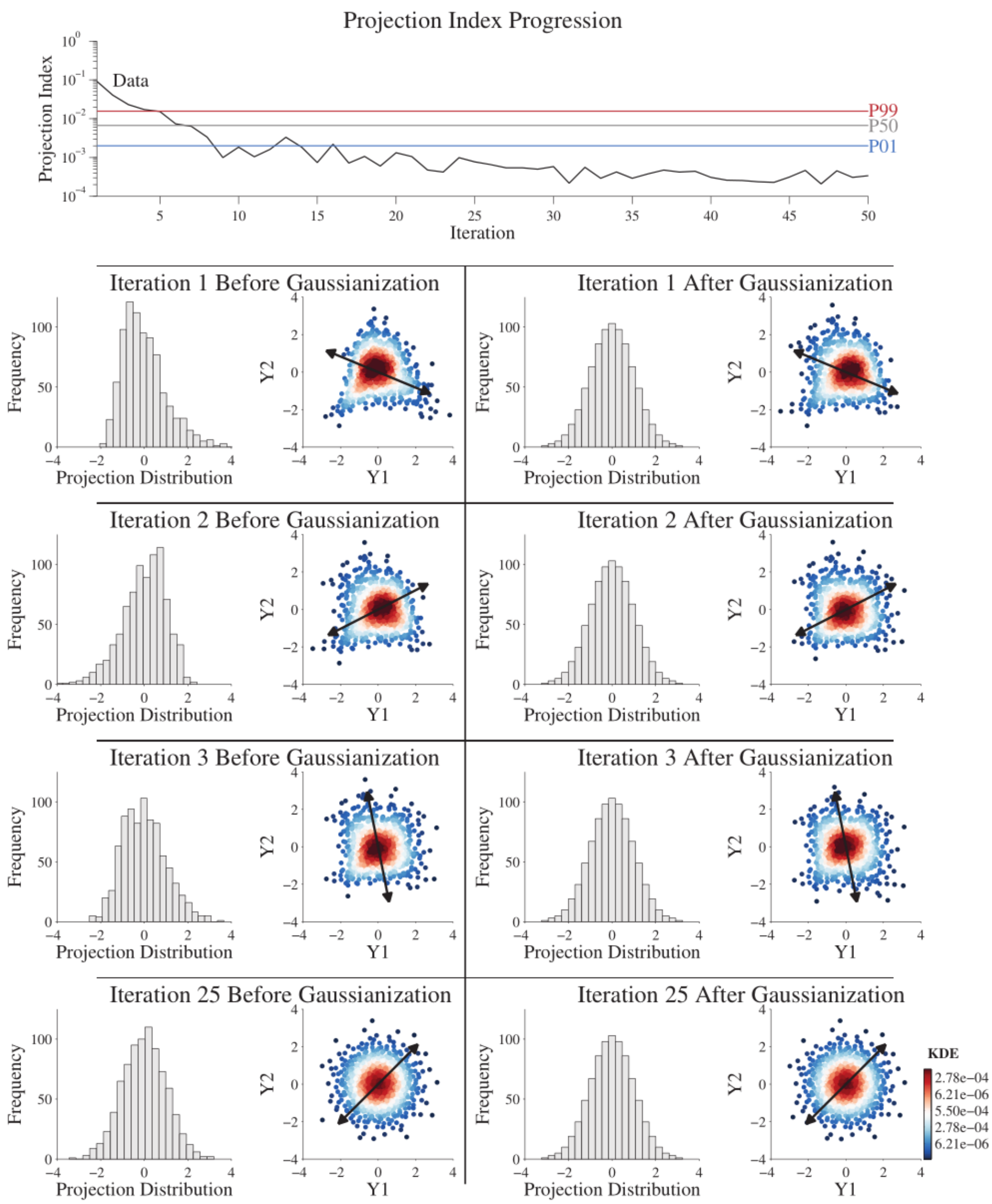


Figure 9: Progression of the data through the projection pursuit algorithm.

Simulation Results

Multivariate reproduction of the back-transformed realizations is evaluated in Figure 10; spatial reproduction is evaluated in Figure 11. Observe that the multivariate complexities are reproduced, as well as the spatial variability. The one potential concern with the PPMT workflow results is the cross-variogram reproduction, which relates to the fact that data are made independent at $h = 0$ lag (zero lag), but not necesarrily at $h > 0$ lag distances (spatial lags). To address this, the PPMT program has the option of performing a subsequent MAF rotation, which generally removes any remaining spatial cross-correlation. Note that the PPMT/MAF results have better reproduction of the semi-variograms and (in particular) the cross-variogram when using a subsequent MAF rotation.

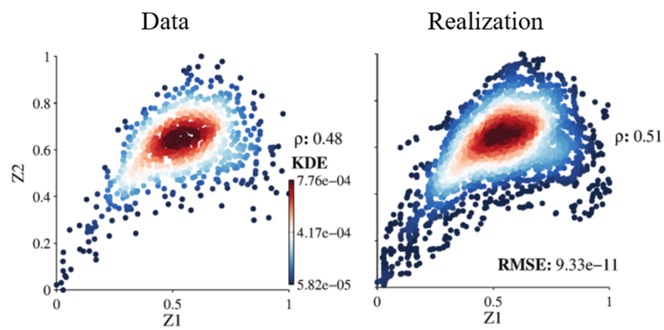


Figure 10: KDE scatterplot for a simulated realization following the PPMT workflow.

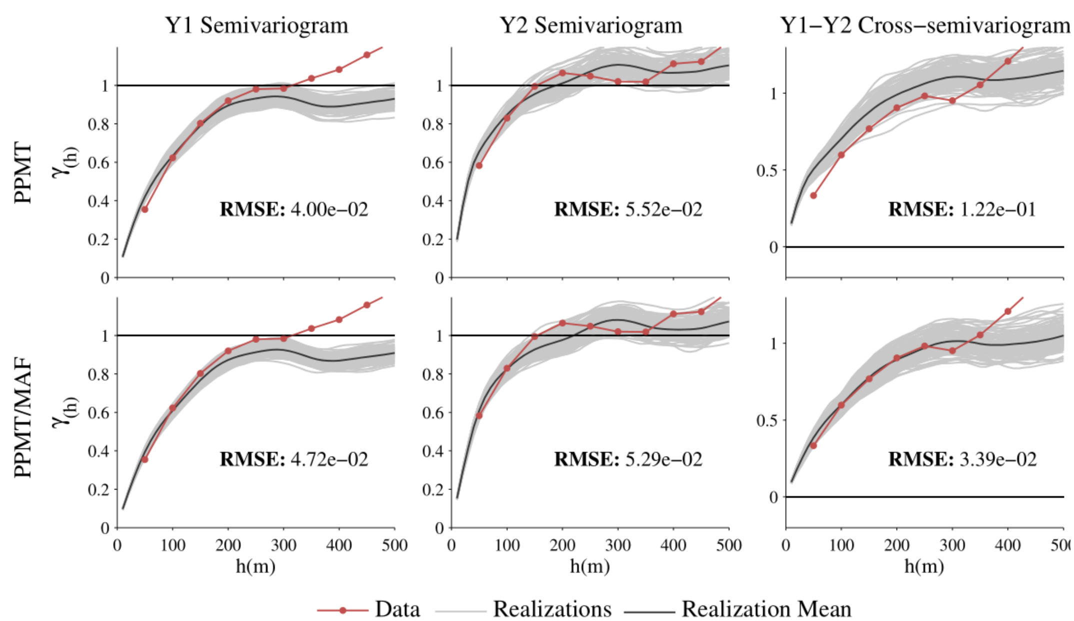


Figure 11: Semivariograms and cross-variograms with and without the use of a chained MAF transformation.

To provide a relative comparison of the PPMT results in terms of multivariate reproduction, several other workflows were executed using more conventional geostatistical modeling techniques. These include:

- Applying the normal score transformation to each of the variables before using colocated cosimulation;
- Applying the normal score transformation to each of the variables before using MAF to decorrelated the. Independent simulation is then performed with the decorrelated variables;
- Applying the stepwise conditional transformation (Leuangthong & Deutsch, 2003) to the data to transform them to a multiGaussian distribution. Independent simulation is then performed with the decorrelated variables.

All factors about the multivariate workflow are held constant, such as the variogram modeling approach, simulation grid, and simulation engine. Figure 12 displays the multivariate reproduction of each conventional workflow; the PPMT/MAF workflow result is also displayed for comparison. Observe that that the PPMT workflow leads to better reproduction of the multivariate features in terms of visual validation, correlation error, and root mean squared error (RMSE) of the kernel density estimation. As a result, the PPMT workflow would lead to superior process performance evaluation with transfer functions that are dependent on the joint distribution of Z_1 and Z_2.

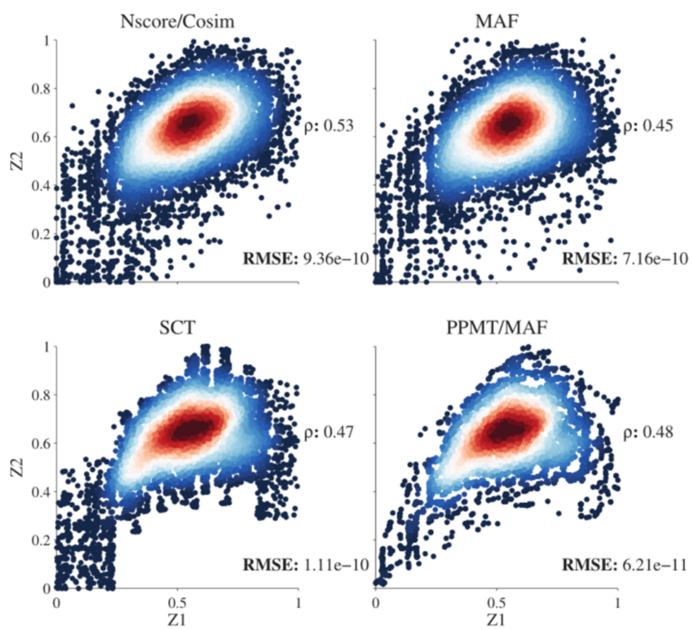


Figure 12: KDE scatterplots of the various transformation and simulation workflows.

5 Geometallurgical Case Study

The PPMT workflow will now be demonstrated with data from an operating Ni laterite mine. At least four variables have to be modeled for properly informing the blend planning of this mine. As already discussed, SiO2 and MgO are modeled because their ratio impacts the operating temperature of the electro-arc furnace. Fe is also modeled because high values can impact the resource recovery rate, as increasing Fe makes it difficult to seperate Ni from the slag. Finally, the Ni resource must be modeled for predicting recovery. Reproducing the multivariate relationships of all four variables is important because blend planning will segregate the material in stockpiles based on their joint values.

As with almost any hierarchical modeling workflow, the first step for modeling this Ni laterite data is to subdivide the data into stationary populations. Figure 13 displays the spatial configuration of geologic rocktypes that will be used for stationary subsetting; Figure 14 displays the univariate and multivariate properties of the rocktypes. The described PPMT workflow will be applied to each of these rocktypes in parallel.

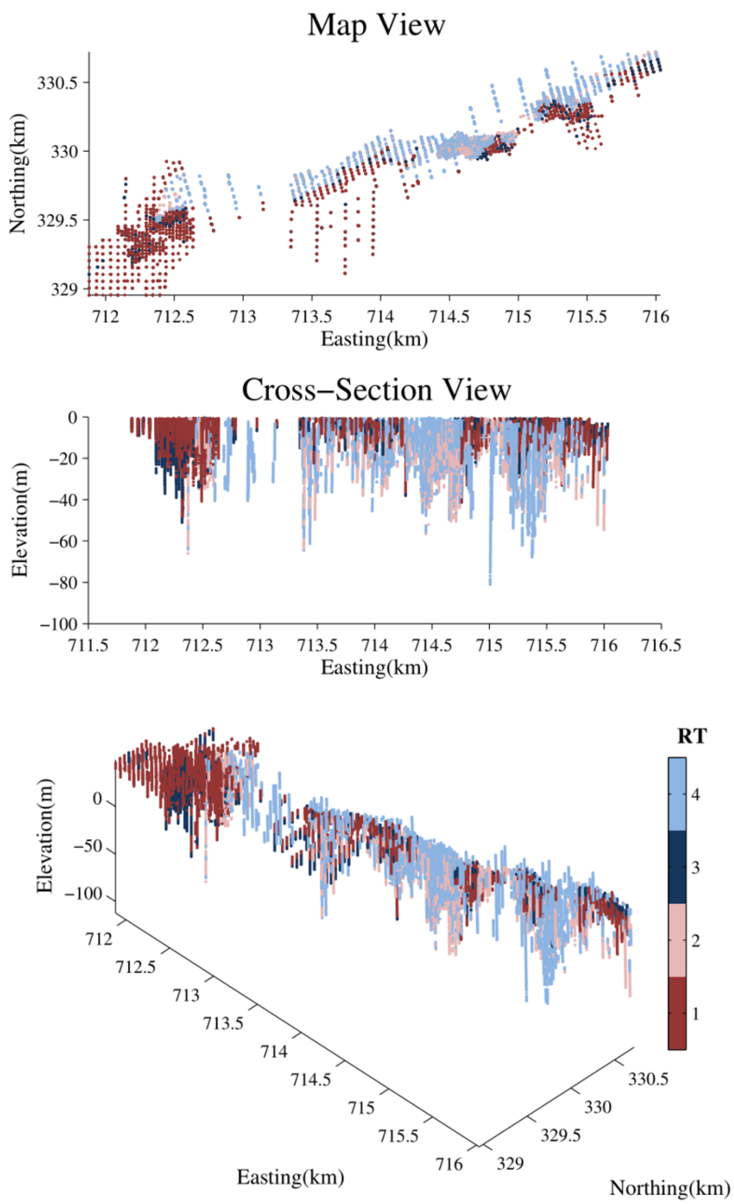


Figure 13: Various perspectives of the data locations, colored by rock type.

Note from the rocktype scatterplots that some multivariate complexitiy will be removed through rocktype subsetting. Nevertheless, Figure 15 (scatterplot of Rocktype 1) shows that multivariate complexity clearly remains within the rocktypes, motivating the use of the PPMT workflow.

The first step of the workflow applies the PPMT to the data, transforming the four variables to be multiGaussian and uncorrelated. Figure 16 displays the transformed data of Rocktype 1. As the variables are now independent, it is appropriate to simulate them independently before back-transforming to restore the original complexity.

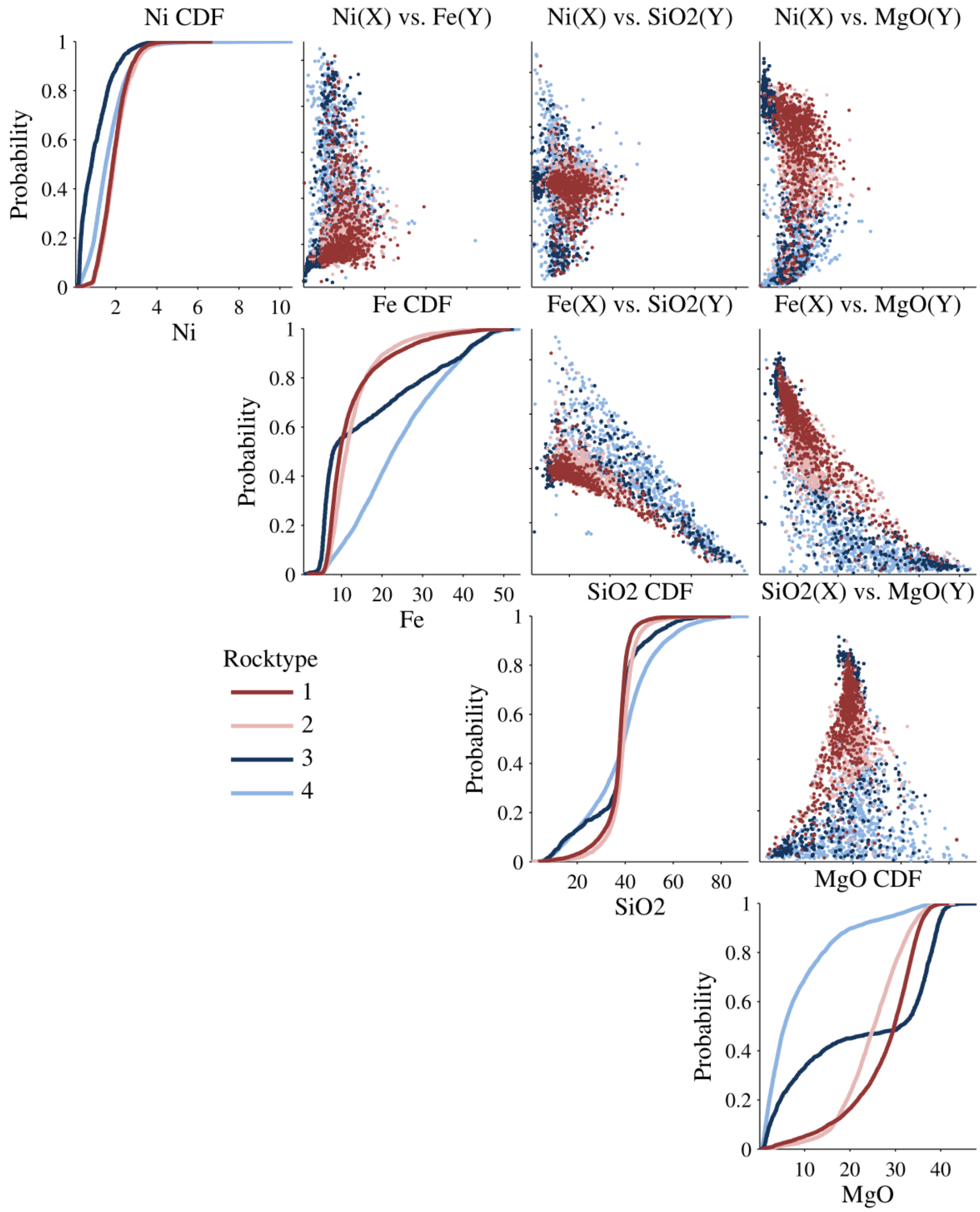


Figure 14: CDFs and scatterplots of the data, colored by rock type.

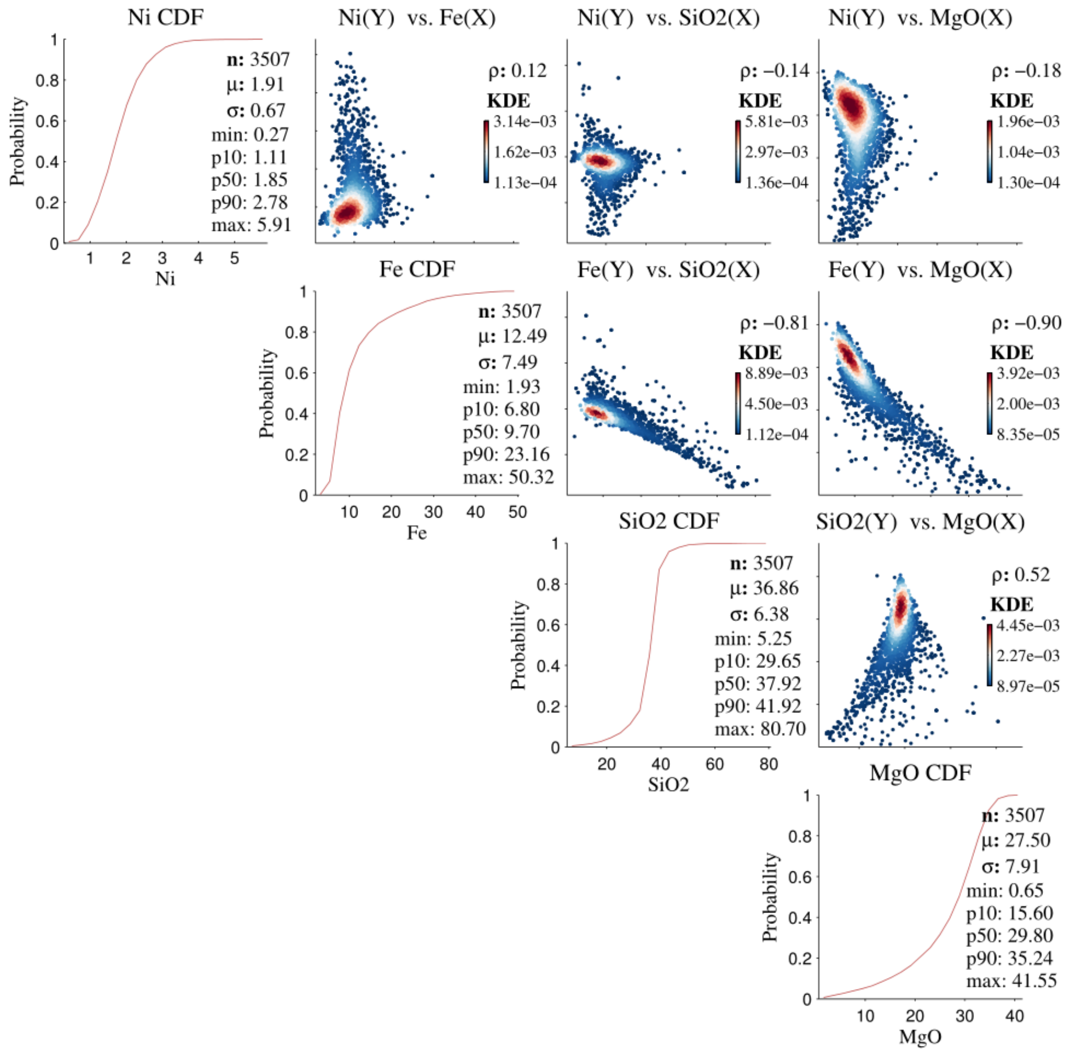


Figure 15: CDFs and scatterplots of Rocktype 1.

After recombining the simulated variables within each rocktype, Figure 17 compares scatterplots of a simulated realization with the original data (termed True). Observe that the multivariate complexities are well reproduced. As a result, the simple PPMT workflow has yielded simulated realizations that will be effective for blend planning and other technical decision making of this Ni laterite mine.

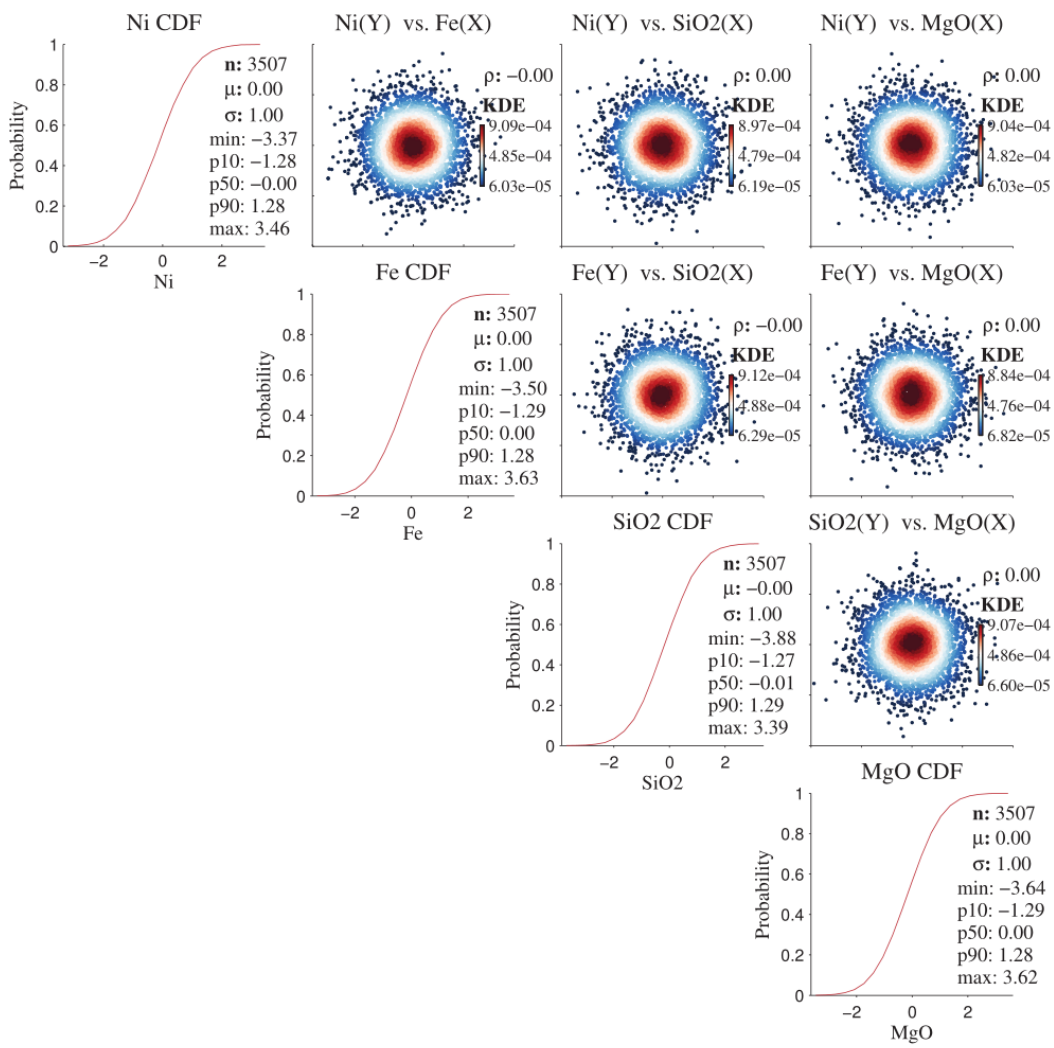


Figure 16: CDFs and scatterplots of the PPMT data.

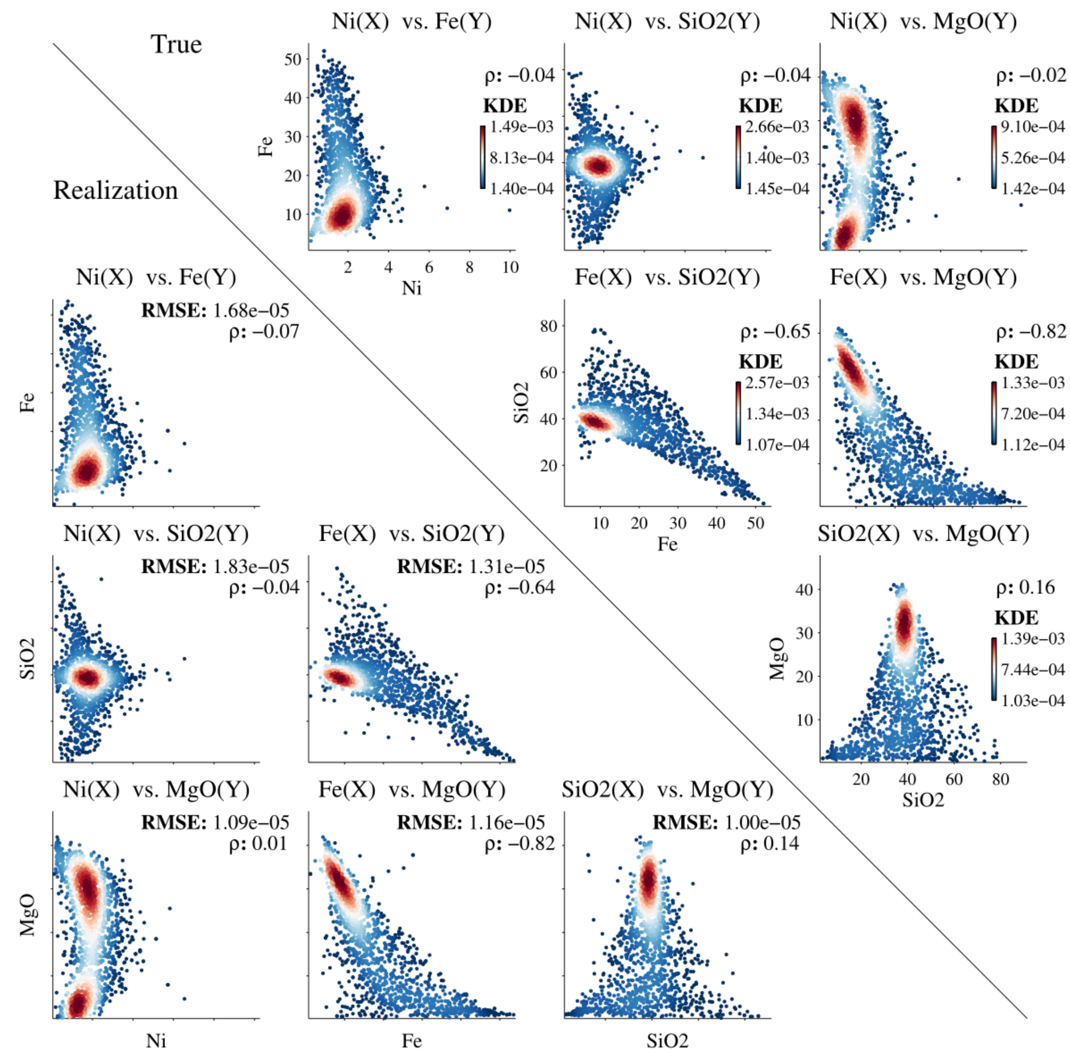


Figure 17: KDE scatterplots of a realization, with the true values shown for comparison.

References

Barnett, R. M., Manchuk, J. G., & Deutsch, C. V. (2014). Projection pursuit multivariate transformation. *Mathematical Geosciences*, *46*, 337--359.

Davis, B. M., & Greenes, K. A. (1983). Estimating using spatiallly distributed multivariate data: An example with with coal quality. *Mathematical Geology*, *15*, 287--300.

Desbarats, A. J., & Dimitrakopoulos, R. (2000). Geostatistical simulation of regionalized poresize distributions using min/max autocorrelations factors. *Mathematical Geology*, *32*, 919--942.

Friedman, J. H., & Tukey, J. W. (1974). A projection pursuit algorithm for exploratory data analysis. *IEEE Transactions on Computers*, *C-23*, 881--890.

Fukunaga, K. (1972). *Introduction to statistical pattern recognition.* (p. 386). Academic Press.

Leuangthong, O., & Deutsch, C. V. (2003). Stepwise conditional transformation for simulation of multiple variables. *Mathematical Geology*, *35*, 155--173.

Verly, G. (1983). The multiGaussian approach and its applications to the estimation of local reserves. *Mathematical Geology*, *15*, 249--286.

Who We Are

The CCG was launched by Professor Clayton V. Deutsch with the vision of becoming a leader in the education of geostatisticians and the delivery of geostatistical tools for modeling heterogeneity and uncertainty. The main objective of the CCG is to support the mutual needs of industry and academia in research and education. The benefits to industry include the opportunity to influence geostatistical research and education, interaction with students as potential employees, early access to publications and access to faculty members for discussions and presentations. The CCG provides a mechanism for industry to contribute to and sustain geostatistical research and teaching, which is of long term interest to many companies.

Contact Us

For more information regarding the demonstrated Solution or to discuss another problem that your project presents, please contact Professor Clayton V. Deutsch at: <cdeutsch@ualberta.ca>
Or drop by our offices at:

Centre for Computational Geostatistics
6-247 Donadeo Innovation Centre For Engineering
9211-116 Street, University of Alberta
Edmonton, Alberta, Canada T6G 1H9

Citation

Barnett, Ryan M. (2016) CCG Solutions - Multivariate Modeling for Resources and Geometallurgy. Centre for Computational Geostatistics, University of Alberta. Retrieved from http://www.ccgalberta.com/solutions/