Nonlinear data normalisation for reliable spatial inference with variational Gaussian processes

by I.G. Gonçalves[1], G.T. Nwaila[2]

Affiliation:
[1] Geological Modeling Laboratory, Federal University of Pampa, Brazil
[2] African Research Centre for Ore Systems Science (CORES), School of Geosciences, University of the Witwatersrand, South Africa

Correspondence to:
I.G Gonçalves

Email:
italogoncalves@unipampa.edu.br

Dates:
Received: 12 Jan. 2026
Revised: 28 Sep. 2026
Accepted: 02 Oct. 2026
Published: Oct. 2026

How to cite:
Gonçalves I.G., Nwaila G.T. 2026. Nonlinear data normalisation for reliable spatial inference with variational Gaussian processes. *Journal of the Southern African Institute of Mining and Metallurgy*, vol. 126, no. 10, pp. 585 – 596

DOI ID:
https://doi.org/10.17159/2411-9717/3880/2026

ORCiD:
I.G. Gonçalves
https://orcid.org/0000-0002-5546-6102
G.T. Nwaila
https://orcid.org/0000-0002-3974-9890

Abstract

Machine learning methods for grade estimation have been rising in popularity in recent years. Among these, the variational Gaussian process model possesses all the beneficial characteristics of the kriging estimator: conditioning on data, declustering properties, and an uncertainty measure that increases with distance from the data points. However, its use requires a transformation of the data to a Gaussian distribution, especially if the data is compositional, as is the case with drillhole assays. This work explores some transformations that were specifically adapted for this model, making use of principal components analysis, independent components analysis, monotonic splines, and flow methods. Results show that the best transformation is case-dependent, and that they provide similar results to the ones used for geostatistical simulation. An open-source implementation of the methods is available to the community.

Keywords

Variational Gaussian process; normal score transform; mineral deposit modelling.

Introduction

Spatial prediction has become an indispensable component of modern earth and mineral systems analysis, underpinning decision-making across exploration targeting, orebody characterisation, resource estimation, and geometallurgical modelling. However, spatial datasets used in geological, geochemical, and geometallurgical investigations are characterised by marked deviations from Gaussian distribution. Heavy-tailed grade distributions, multi-modal responses associated with lithological and geochemical heterogeneity, compositional constraints, and complex spatial dependence structures introduce asymmetry and nonlinearity that challenge widely used modelling frameworks. Gaussian process (GP) models (Rasmussen, Williams, 2006), and particularly their variational formulations (Hensman et al., 2015), have become increasingly attractive for spatial inference because they can capture nonlinear relationships, propagate uncertainty rigorously, and incorporate anisotropic kernels that reflect geological architecture. However, a core assumption embedded in the latent structure of Gaussian processes is that the transformed data should approximate a multivariate Gaussian distribution. When this assumption is violated, kernel hyperparameters may be poorly estimated, predictive means may exhibit structural bias, and uncertainty intervals may display under- or over-confidence. These issues can propagate into downstream technical and economic evaluations of mineral deposits, including recoverable resource estimates, metallurgical process response predictions, and mine planning decisions that depend on stable representations of spatial uncertainty.

Normalisation is therefore not a peripheral preprocessing decision, but a central element of the modelling workflow. The way spatial data are transformed prior to variational Gaussian process (VGP) modelling (Gonçalves et al., 2022) determines the geometry of the latent space, the smoothness and anisotropy recovered by the kernel, and the calibration of predictive probabilistic outputs. Despite its importance, normalisation is often treated as a routine operation, implemented through simple Z-score scaling or logarithmic transforms without consideration of spatial effects, compositional structure, or multivariate dependencies. In practice, these choices determine whether the resulting VGP captures the true spatial behaviour of the underlying process or produces misleadingly precise predictions that obscure geological variability.

The challenge is amplified in natural resource applications where datasets are typically sparse, irregularly sampled, and affected by both geological and analytical noise. Geochemical variables may possess strong positive skewness, reflecting enrichment processes, while mineralogical and geometallurgical attributes can exhibit non-linear cross-correlations arising from mineral associations, alteration fronts, or process domains. Compositional datasets, such as oxide proportions, reside in the simplex and obey closure constraints that invalidate many conventional transformations. Spatial cross-validation further reveals that naïve normalisation can result in models that perform well in random splits but degrade when honouring spatial autocorrelation structures.

Recent methodological advances have begun to address aspects of these challenges. Spline-based univariate Gaussianisation has been demonstrated as an effective means of handling skewness while preserving rank structure (Gonçalves et al., 2020). More flexible nonlinear transformations, including projection pursuit multivariate transformation (PPMT) (Barnett et al., 2014) and normalising flows (Grathwohl et al., 2019), have gained traction in the geostatistical and machine learning literature, respectively, although their data demands and interpretability constraints limit routine application in natural resource settings. Yet, there has been no systematic analysis of how these various normalisation strategies compare when applied to spatial datasets under a unified VGP framework, nor how they influence predictive accuracy, uncertainty calibration, and spatial continuity.

This study addresses that gap by examining a suite of univariate, multivariate, and nonlinear normalisation methods applied to representative geometallurgical datasets. The focus is on understanding how each transformation reshapes the latent Gaussian geometry and how these differences manifest in spatial predictions and associated uncertainty measures. Variational Gaussian processes are used as the modelling platform because of their suitability for large datasets and their explicit reliance on latent Gaussian structure. Spatial cross-validation is employed to ensure that model comparisons reflect realistic generalisation behaviour rather than artefacts of random partitioning.

Theory

Variational Gaussian process modelling

The variational Gaussian process is a generalisation of the standard Gaussian process (Rasmussen, Williams, 2006), which in turn is a Bayesian derivation of kriging. It is based on the concept of latent variables, i.e., random Gaussian fields that are functions of the input coordinates. A vector-valued field $\boldsymbol{f(x)}$ is generated from a number M of inducing points t distributed over the same region as the data points $\boldsymbol{x}$, with corresponding latent variables $\boldsymbol{u(t)}$. Aggregating all the N data points in D dimensions in an $N\times D$ matrix $\boldsymbol{X}$, and the corresponding P latent variables in an $N\times P$ matrix $\boldsymbol{F}$, the corresponding probability distribution is given by (Hensman et al., 2015; Gonçalves et al., 2022, Gonçalves, Wellmann, 2025):

$$q(\boldsymbol{F}) = \prod_{p=1}^{P} q(\boldsymbol{f}_p) \quad [1]$$

$$q(\boldsymbol{f}_p) \sim \mathcal{N}(\boldsymbol{f}_p; \boldsymbol{\mu}_p, \boldsymbol{\Sigma}_p) \quad [2]$$

$$\boldsymbol{\mu}_p = \boldsymbol{K}_{fu}\boldsymbol{K}_{uu}^{-1}\boldsymbol{m}_p \quad [3]$$

$$\boldsymbol{\Sigma}_p = \boldsymbol{K}_{ff} + \boldsymbol{K}_{fu}\boldsymbol{K}_{uu}^{-1}(\boldsymbol{S}_p - \boldsymbol{K}_{uu})\boldsymbol{K}_{uu}^{-1}\boldsymbol{K}_{uf} \quad [4]$$

where $\boldsymbol{K}_{ij}$ is a covariance matrix computed for all pairs of inputs with any of the usual functions (Gaussian, spherical, exponential, etc.). For the sake of simplicity, they are assumed to be the same for all p latent variables, but this is not a requirement. Furthermore, $\boldsymbol{m}_p$ and $\boldsymbol{S}_p$ are the posterior mean and covariance matrix of the $\boldsymbol{u}_p$ latent variables, which are trainable parameters in the model. In other words, a field is generated by kriging the latent variables from the inducing points to the data points. If $M<N$, a computational gain is obtained at the potential cost of a loss in spatial detail.

The latent variables are linked to the measurements $\boldsymbol{y}$ through a likelihood function $p(\boldsymbol{y}|\boldsymbol{f})$. If the likelihood is Gaussian, the model will seek to minimise the squared error of predictions similarly to kriging, but other likelihood functions can be used (Hensman et al., 2015; Gonçalves et al., 2022), even allowing the model to deal with categorical data (Gonçalves et al., 2023). The model is trained by maximising the evidence lower bound (ELBO), an approximation of the true data likelihood $p(\boldsymbol{Y})$:

$$\log p(\boldsymbol{Y}) \geq \sum_{n=1}^{N} \mathbb{E}_{q(\boldsymbol{f}_n)}[\log p(\boldsymbol{y}_n|\boldsymbol{f}_n)] - \sum_{p=1}^{P} \mathrm{KL}[q(\boldsymbol{u}_p)||p(\boldsymbol{u}_p)] \quad [5]$$

where $\boldsymbol{Y}$ is the matrix aggregating all measurements, $p(\boldsymbol{up})\sim N(\boldsymbol{0},\boldsymbol{Kuu})$ is the prior distribution of the inducing points' latent variables, $q(\boldsymbol{up})\sim N(\boldsymbol{m_p},\boldsymbol{S_p})$ is the posterior, and $\mathrm{KL}[q(\boldsymbol{u}_p)||p(\boldsymbol{u}_p)]$ is the Kullback-Leibler divergence between them. The expectation $E_{q(\boldsymbol{f}_n)}$ is computed by Monte Carlo, with samples generated from $q(\boldsymbol{u}_p)$ that are interpolated to the data points. The same approach can be used to generate samples at new spatial locations after the model is trained.

Normalisation methods

The formulation in the aforementioned assumes that the measurements $\boldsymbol{y}$ are Gaussian-distributed, which is often not the case in the geosciences, as data distributions can be skewed and/or constrained. In the GP literature, data normalisation is denominated as warping (Snelson et al., 2004, Gonçalves et al., 2020). Warping consists of a bijective function $\boldsymbol{y}=\varphi(\boldsymbol{z})$ that maps the measurements $\boldsymbol{z}$ to normalised values. It is essentially similar to the Gaussian anamorphosis employed in geostatistics, but it can have trainable parameters. To incorporate warping in VGP training, Equation 5 is extended with a change of variables formula:

$$\log p(\boldsymbol{Z}) = \log p(\boldsymbol{Y}) - \sum_{n=1}^{N} \log \det \left| \frac{\partial \varphi(\boldsymbol{z}_n)}{\partial \boldsymbol{z}_n} \right| \quad [6]$$

where $\frac{\partial \varphi(\boldsymbol{z}_n)}{\partial \boldsymbol{z}_n}$ is the Jacobian of the transformation. Specific warping functions are described in the following, according to their uses.

Scaling

By default, the VGP implementation presented here assumes that the data has zero mean and unit variance. The Z score warping is simply given by:

$$\varphi(z_p) = \frac{z_p - \mu_p}{\sigma_p} \quad [7]$$

where μ_p and σ_p are the trainable mean and variance. They can be initialised with the empirical mean and variance of the data.

Constraints

Quantities such as chemical assays have non-negativity constraints. In such cases the log ($\varphi(z_p)=\log z_p$) or softplus ($\varphi^{-1}(y_p)=\log(1+e(y_p))$) functions can be applied. Chemical assays are often treated as compositional (Pawlowsky-Glahn et al., 2015), encompassing a positive and constant sum constraint. This can be dealt with by applying the centred log-ratio transformation (clr):

$$\varphi(z_p) = \log z_p - \frac{1}{P}\sum_{k=1}^{P} \log z_k \quad [8]$$

Note that, due to the subtraction of the mean, the resulting vector is linearly dependent. This can be avoided by using a vector basis in a (P-1)-dimensional space. The isometric log-ratio transformation (ilr) derives such a basis from sequential partitions of the original composition, but it is not explored in this work.

Decorrelation

The P latent variables are assumed to be independent for the sake of computational efficiency. If the measurements are not decorrelated, the final results may be sub-optimal. The simplest decorrelation method is principal components analysis (PCA) (Jolliffe, 2002). PCA finds the directions of maximum variance in the data space and returns an orthonormal matrix $\mathbf{Q}$. The transformation is given by $\varphi(\boldsymbol{z})=\mathbf{Q}\boldsymbol{z}$. If the variables have strong correlations, it is possible to retain only the first few principal components that explain the majority of the variability in the data.

Datasets often contain extreme values and/or outliers that can influence the PCA results. In this case a robust method can be used, such as the minimum covariance determinant (MCD) after Rousseeuw and Van Driessen (1999). It consists of computing the data covariance matrix using only a fraction of the data points that are closer to the mean in the Mahalanobis sense.

Gaussianisation

Data distributions may present asymmetry and/or heavy tails. A simple way to correct this is to employ monotonic splines (Gonçalves et al., 2020), defining a set of (y,z) knots, which are interpolated to obtain $\varphi(z_p)$ and $\varphi^{-1}(y_p)$. These splines are defined separately for each of the P directions, so if the data have a complex distribution this method may not be sufficient for a complete Gaussianisation. The data can be rotated prior to the application of splines to deal with non-Gaussianity in oblique directions. The method known as independent components analysis (ICA) (Hyvärinen, Oja, 2000) can be used to find the directions of maximum non-Gaussianity.

Flow methods generate a time-variant vector field that is used to move the data points closer to a Gaussian distribution (Grathwohl et al., 2019). This vector field is integrated for a number of time steps, and over time they may aggregate in a region of space, spread apart in others, make curves, etc. The flow can be reversed by integrating the vector field backwards in time. In this work, the vector field is defined by the mean of a GP generated from inducing points spread over $\boldsymbol{z}$ space.

Warping chains

A single warping function may not be sufficient to obtain good results. In this case, multiple warping functions can be chained together, each one working with the output of the previous one. Such chains can be built to mimic the behaviour of the known approaches in geostatistics. Figure 1 demonstrates the chains employed in this work. By choosing different paths and partially traversing the graph, different normalisation strategies can be tested, also considering the characteristics of the data (non-negative, compositional, etc.).

Methodology

The evaluation presented in this study is based on a comparative analysis of univariate and multivariate normalisation strategies applied to representative spatial datasets drawn from geometallurgical domains. The purpose of this section is to describe the datasets, justify their selection, outline the transformation procedures applied, and detail the variational Gaussian process modelling and cross-validation framework used to assess predictive accuracy, uncertainty calibration, and spatial structure recovery.

Data description

Four datasets were selected to capture a spectrum of distributional and dependence structures commonly encountered in mineral-systems analysis:

- **Jura:** the Jura dataset (Goovaerts, 1997) contains 7 variables that represent heavy metal contamination. Each variable has a different degree of asymmetry and they present non-linear relationships.
- **Synthetic:** this dataset contains 6 variables with highly non-linear relationships, generated explicitly to test a geostatistical simulation algorithm (De Figueiredo et al., 2021).
- **Copper:** a Cu deposit in Brazil, where many different elements were analysed (Hoffimann et al., 2022).
- **Assen:** a Fe deposit in South Africa, presenting multiple lithological domains and multi-modal distributions.

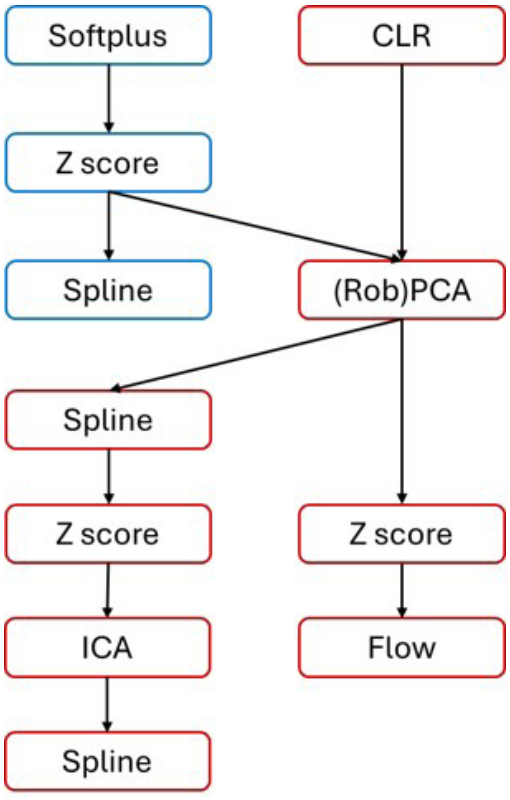


Figure 1—Warping chains employed in this work. Blue: univariate path. Red: multivariate path

The characteristics of each dataset and the methodologies employed are described in Table 1. To explore a range of practical situations, a different validation strategy was applied to each one.

The last two datasets are compositional, consisting of normalised oxide proportions from a set of drillhole or bench-scale samples. Because compositional data lie in the simplex and obey a closure constraint (Pawlowsky-Glahn et al., 2015), they provide an essential test of normalisation strategies that must respect this geometry. These datasets therefore provide a realistic basis for evaluating warping methods. Across all datasets, the spatial configuration reflects irregular sampling common in exploration and operational settings. Sampling density varies across structural domains, and drilling patterns reflect operational priorities, leading to non-isotropic spatial support. These characteristics ensure that the evaluation framework reflects conditions typical of real-world geometallurgical modelling, where data sparsity and anisotropy interact strongly with the normalisation and GP modelling choices.

Validation metrics

To assess the efficiency of the proposed approach over a wide range of situations, different validation methods were applied to each dataset (Table 1). Different validation metrics were also tested. The following metrics are available in the scikit-learn library (Pedregosa et al., 2011): Root mean square error (RMSE), mean absolute error (MAE), and median absolute error (MDAE). In the figures below, a subscript p indicates that the metric was computed from the prediction, averaging all predictive samples, while the subscript s indicates the aggregation of the errors of the individual samples. The samples were also used to compute the squared bias (average deviation from the true values) and variance, treating the samples as the predictions from individual models. The interval score (IS) (Gneiting, Raftery, 2007) was also computed with a significance of 0.05. IS assesses the accuracy of confidence intervals, which in this case were derived empirically from the samples. For compositional data the Aitchison distance is also computed (Pawlowsky-Glahn et al., 2015). For all metrics a lower value is better, and 25 samples were found to be sufficient to obtain stable results.

Results

Jura

Figure 2 displays the performance of different warping chains on the Jura dataset, using a set of 100 validation points as reference. The results vary for each element but overall, the ICA path performed best in terms of the interval score and, thus, has the best calibrated uncertainty.

Figure 3 displays the obtained normal scores, while Figure 4 allows a comparison of the distributions of the original data and samples computed at the data points. A close correspondence can be observed between the two.

Figure 4—Distribution of the Jura data (lower triangle, bars) and samples computed on the data locations (upper triangle, black lines)

Synthetic data

This dataset was designed specifically to challenge the standard normalisation methods. In this case the flow path performed better overall. Figures 5 to 7 display the results.

Copper

Due to the sheer number of variables, the distribution of the Copper dataset is best visualised through a PCA (Figure 8). It can be seen that the point cloud is not perfectly aligned with the principal components axes and there are some extreme points in the lower right corner of the plot. Due to this the robust PCA version was used in the warping chain for this dataset.

The validation strategy here was to divide the data into 5 folds, using the methodology by Linnenbrink et al. (2024). Four folds were used for training, and one was held out for validation. Figures 9 to 10 display the results on the validation set. In this case the ICA path displayed better results overall.

Assen

The Assen dataset displays a strong multimodality induced by differences in lithology (Figure 11). To take this feature into account, this deposit was modelled with the method described in Gonçalves and Wellmann (2025), where extra latent variables are assigned to the lithological classes and reused to influence the latent variables responsible for the assays. Due to the relatively small sample size concentrated in a small number of drillholes, a simple random train/test split was made, keeping 80% of the data for training.

Table 1

Description of the datasets

Dataset	Dimensionality	Data points	Variables	Characteristics	Validation method
Jura	2	259	7	Variable marginal distributions; non-linear relationships; non-negative.	Held-out data in a regular grid.
Synthetic	2	200	6	Non-linear and non-monotonic relationships.	Held-out data in a regular grid.
Copper	3	2000	19	Compositional.	Spatial k-fold (Linnenbrink et al., 2024).
Assen	3	807	9	Compositional; multimodal.	Random train/test split.

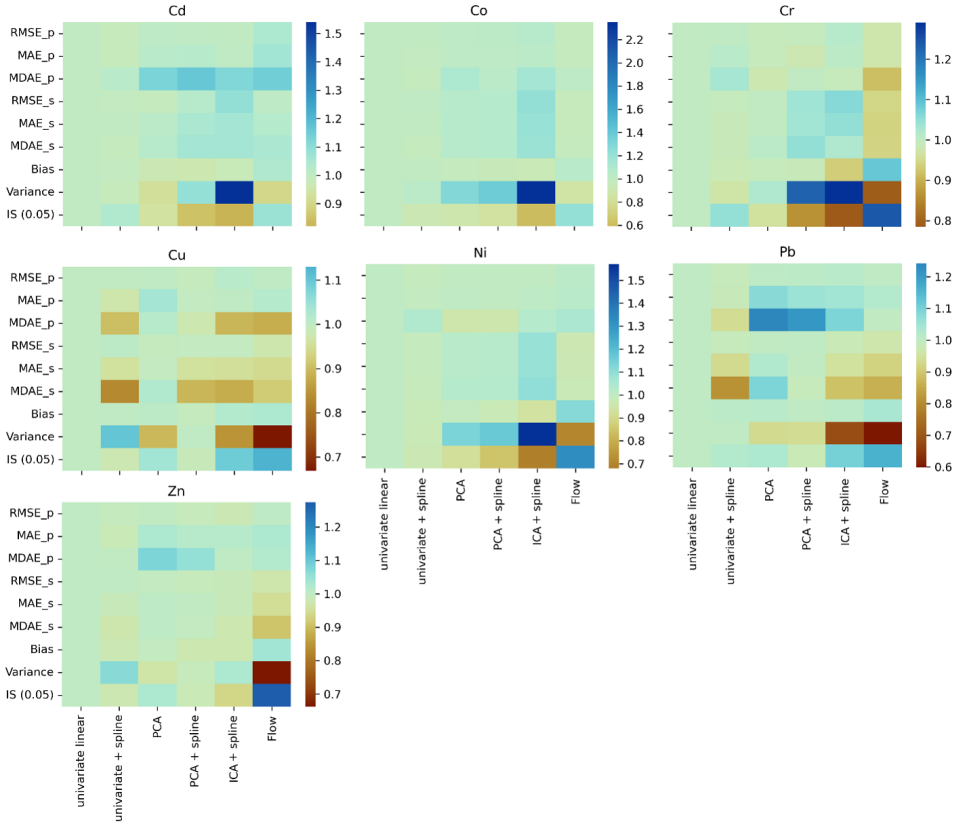


Figure 2—Validation results for the Jura dataset, relative to a simple linear rescaling

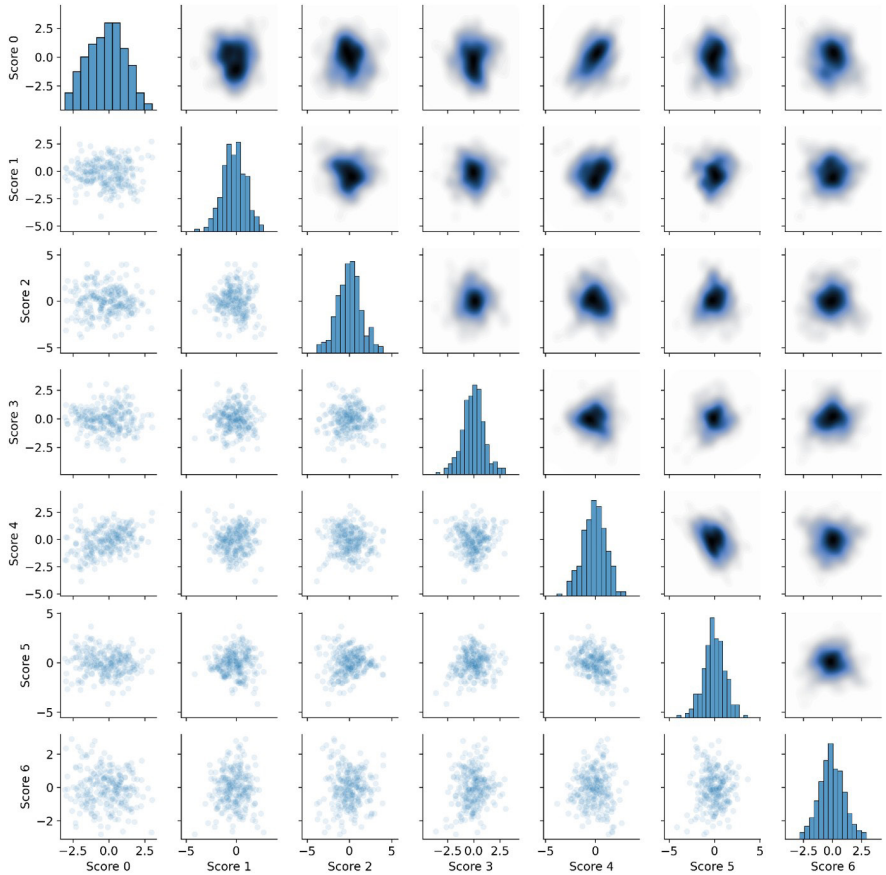


Figure 3—Normal scores obtained for the Jura dataset

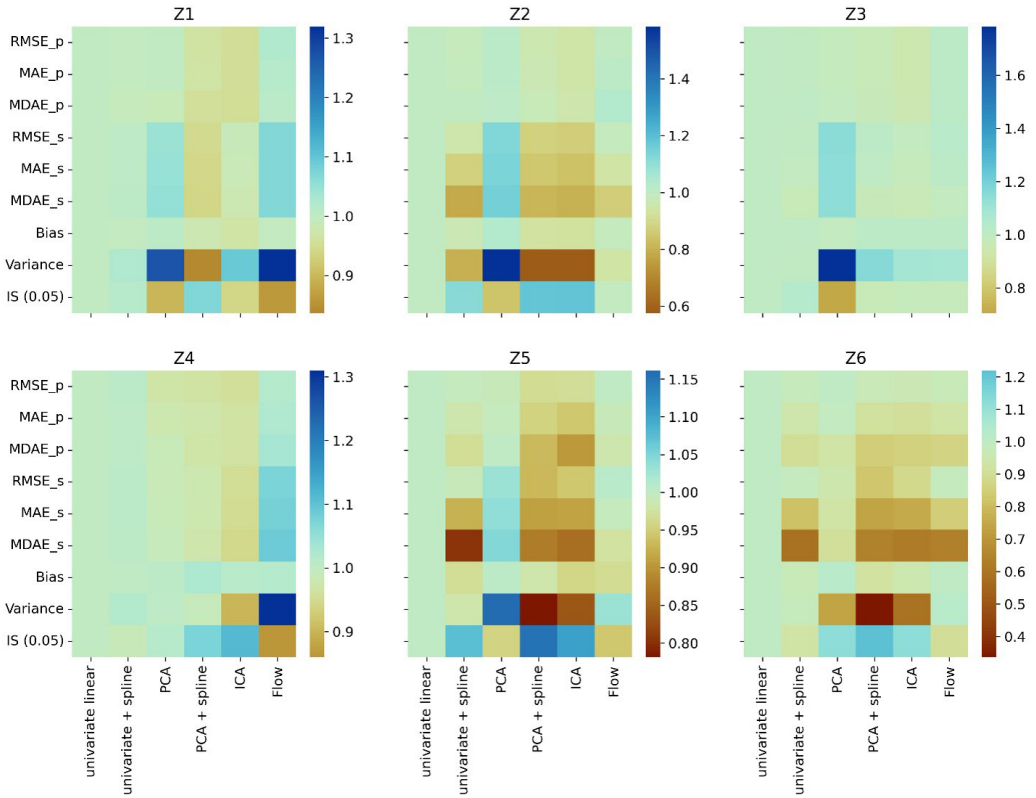


Figure 5—Validation results for the Synthetic dataset, relative to a simple linear rescaling

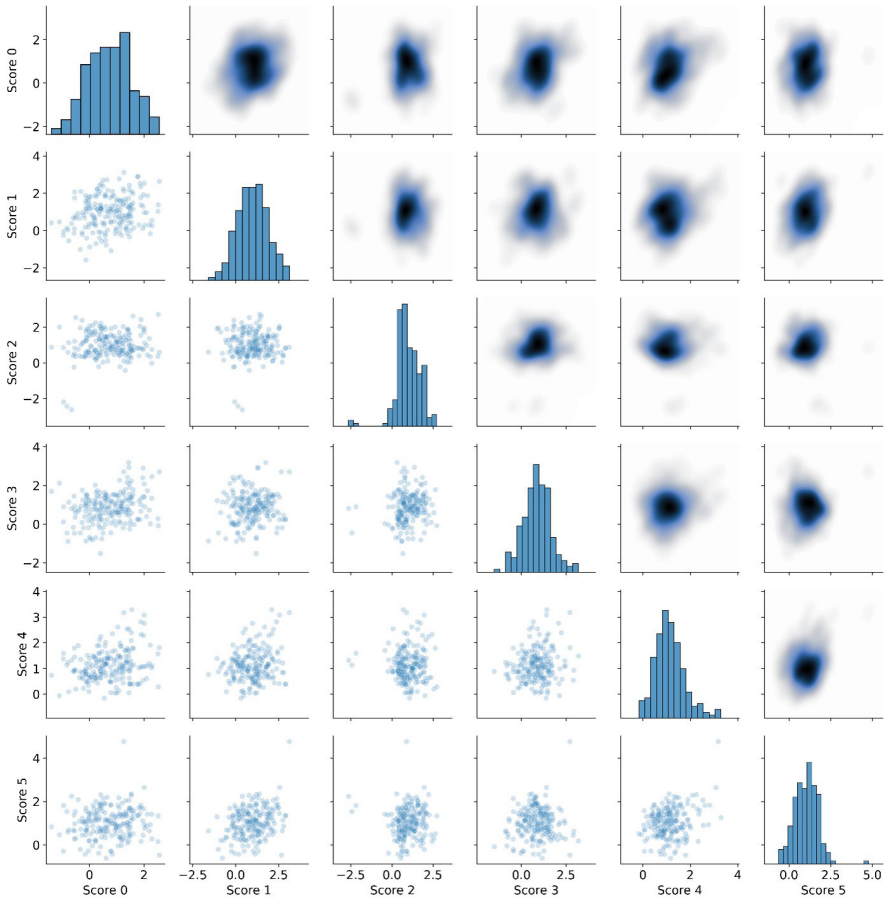


Figure 6—Normal scores obtained for the Synthetic dataset

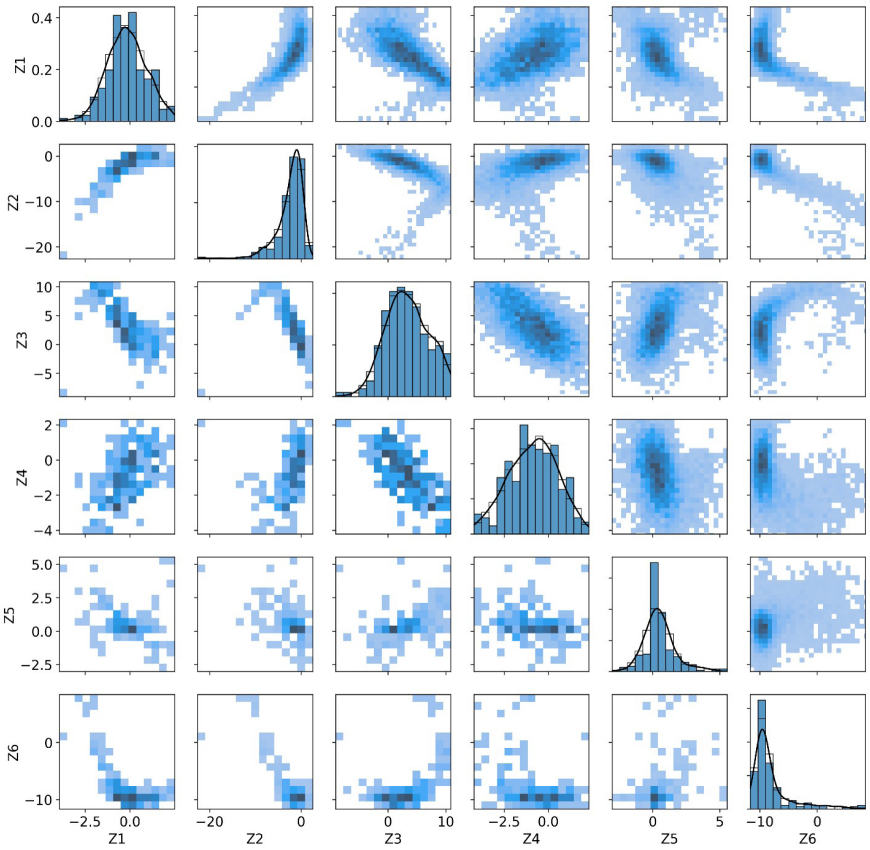


Figure 7—Distribution of the Synthetic data (lower triangle, bars) and samples computed on the data locations (upper triangle, black lines)

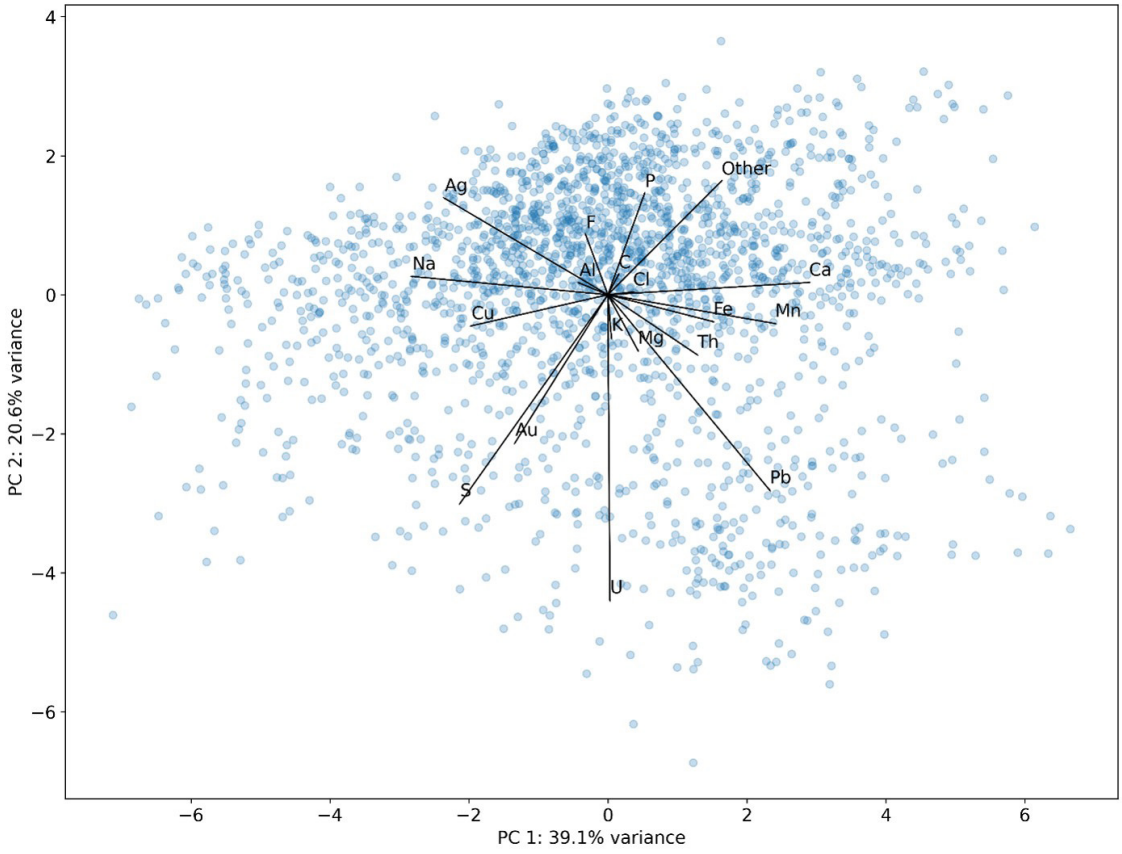


Figure 8—The first two principal components of the clr-transformed Copper dataset

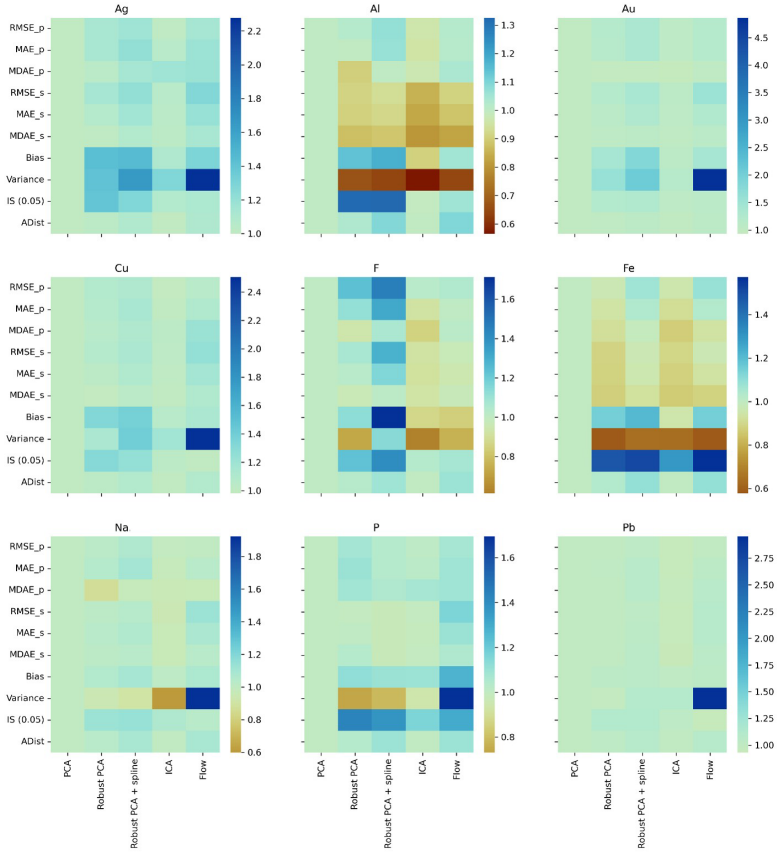


Figure 9—Validation results for the Copper dataset, relative to a simple linear rescaling of the clr-transformed values

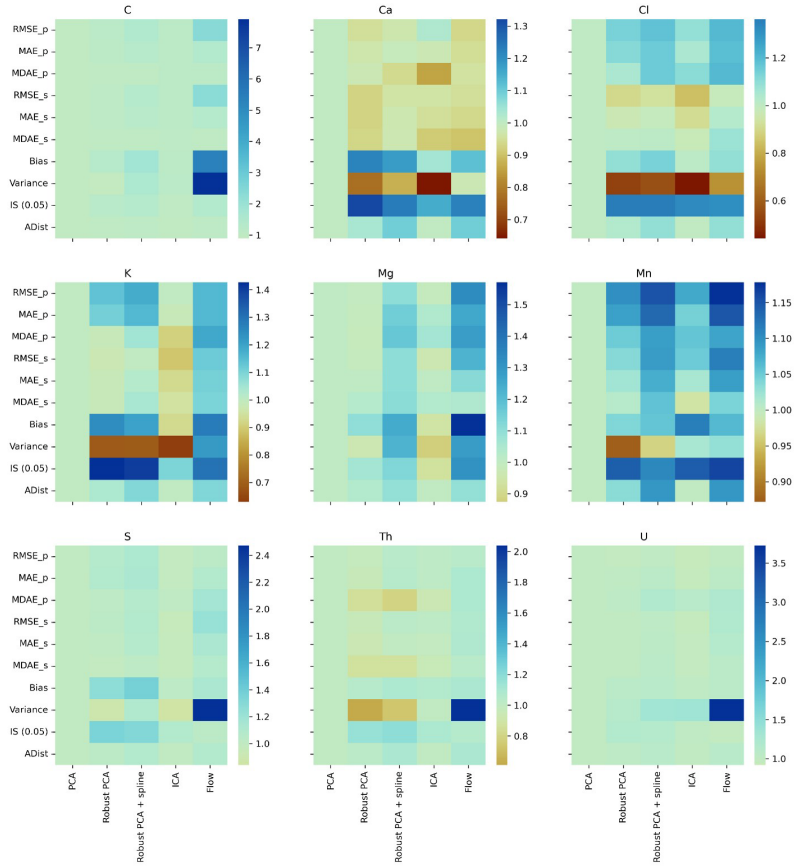


Figure 10—Validation results for the Copper dataset, relative to a simple linear rescaling of the clr-transformed values (continued)

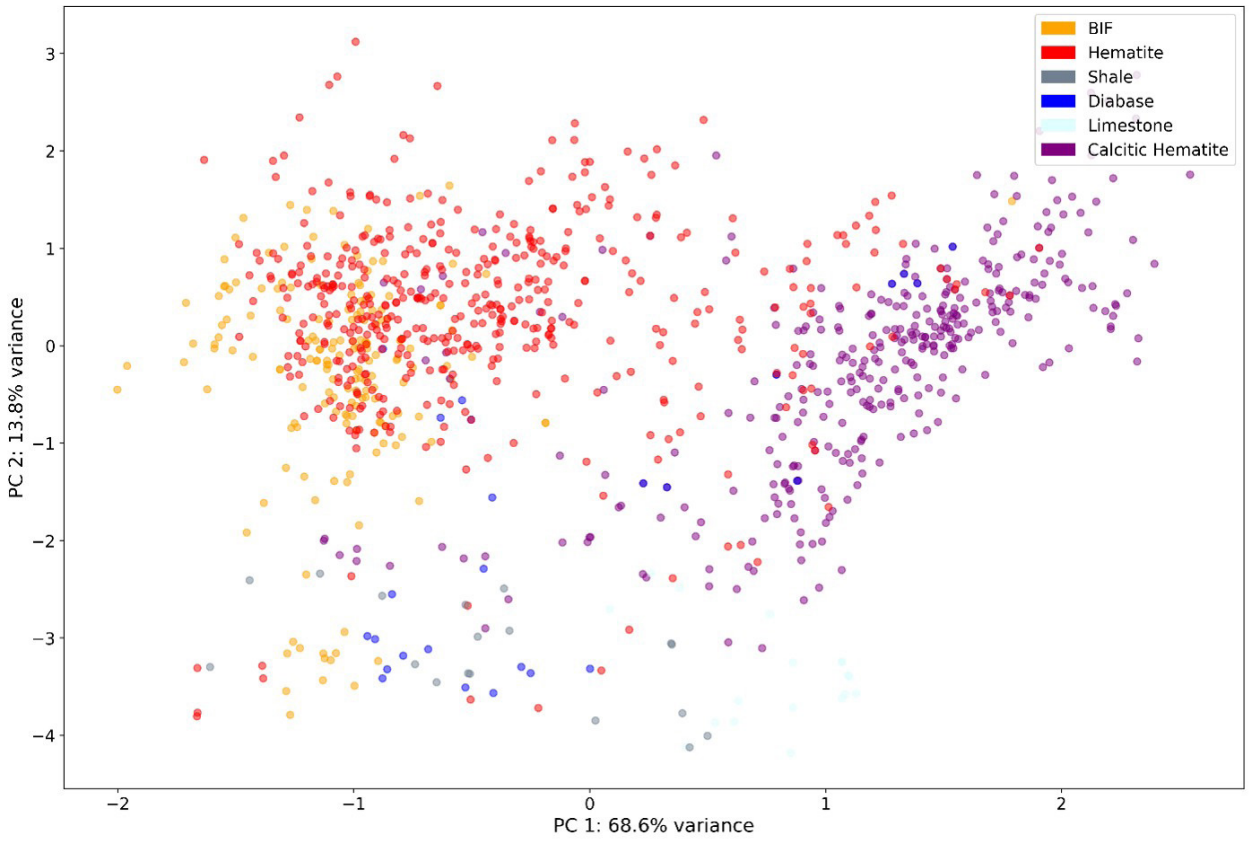


Figure 11—The first two robust principal components of the Assen dataset

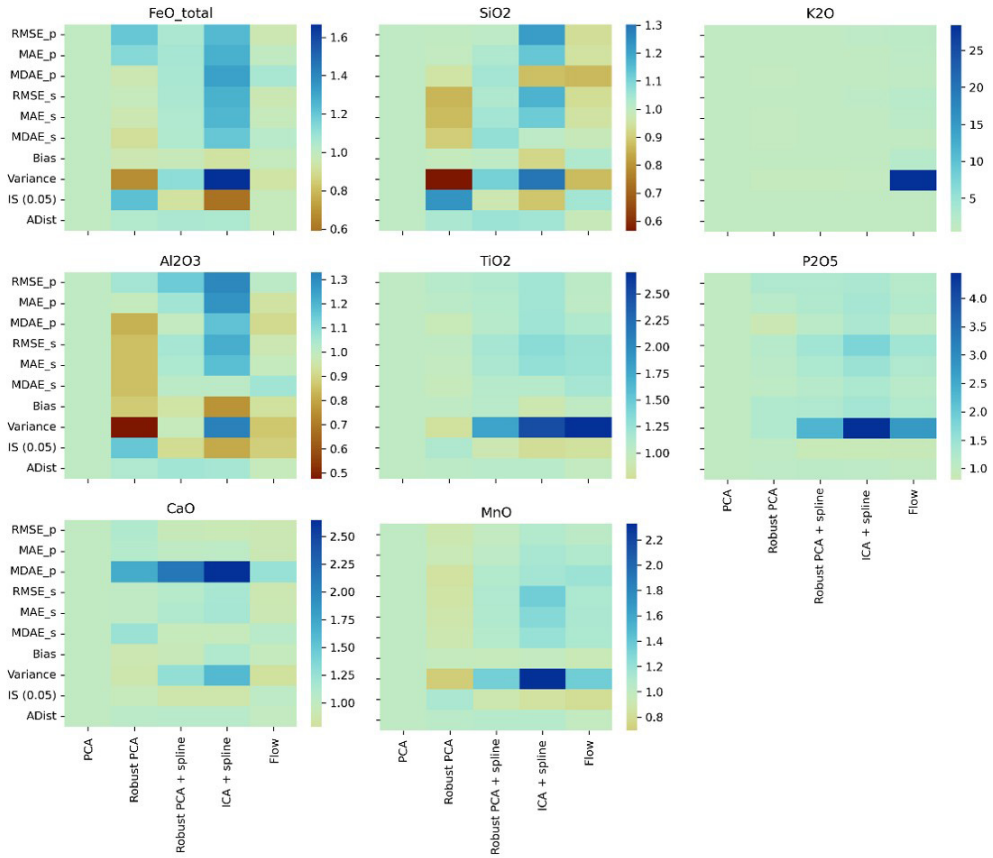


Figure 12—Validation results for the Assen dataset, relative to a simple linear rescaling of the clr-transformed values

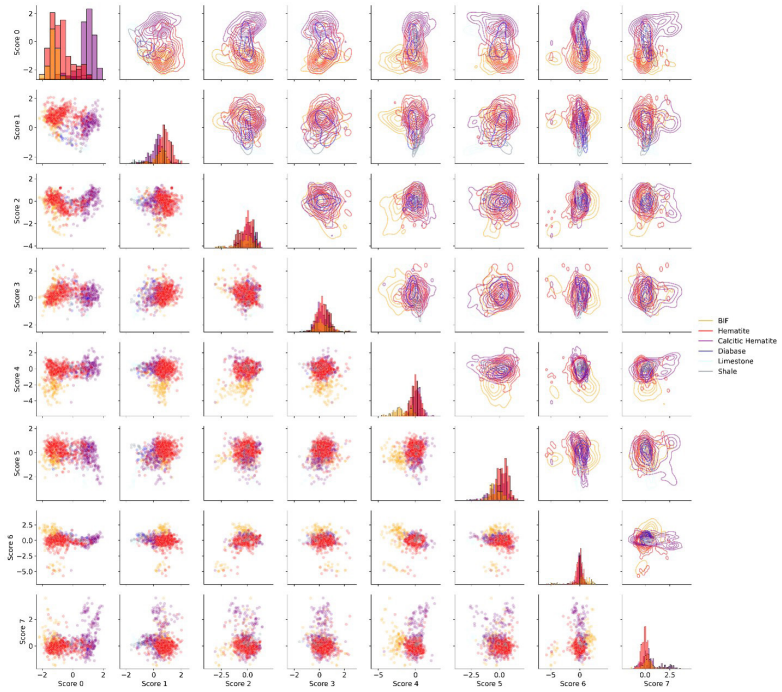


Figure 13—Normal scores obtained for the Assen dataset

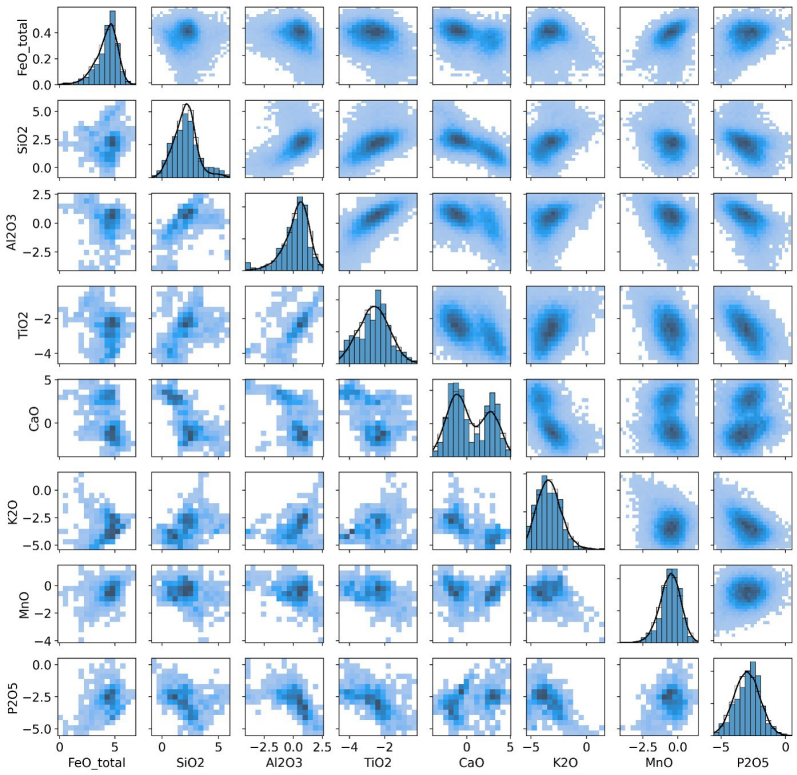


Figure 14—Distribution of the clr-transformed Assen data (lower triangle, bars) and samples computed on the data locations (upper triangle, black lines)

The validation results are displayed in Figure 12. Some elements were insensitive to the choice of method, while for others the best results were obtained with simpler warping paths. The flow path seems to be slightly better overall.

Due to the influence of lithology, the model normalised the data only within each category (Figure 13). The two clusters visible in Figure 11 were not merged by the flow warping as would be expected. Nevertheless, the relationships between the variables were recovered (Figure 14).

Discussion

The results did not reveal a clear winner among the methods studied, although the ICA and flow paths performed better than the simpler methods overall. All methods performed better than

simple rescaling, demonstrating that asymmetries and correlations do impact modelling results and should not be ignored. Apart from this, the choice of method is a matter of geological interpretation and an accuracy/complexity trade-off.

For the Assen data normalisation was not complete, but that does not necessarily mean that the model was sub-optimal. The corresponding procedure in traditional geostatistics would be to make separate models for each geological domain, in which it would be possible to fully normalise the data. VGP achieves a similar result without the need for separate models.

From the results is seems that approximate Gaussianity is sufficient to obtain good results, without the need for multiple passes as in PPMT. In machine learning it is always beneficial to avoid unnecessarily complex models due to the risk of overfitting and sub-optimal training convergence.

Conclusions

This study demonstrates that spatial data normalisation is a decisive factor in the reliability of variational Gaussian process inference for geological, geochemical, and geometallurgical datasets. Transformations that merely rescale or symmetrise the data are insufficient when confronted with the heavy-tailed distributions, non-linear dependencies and compositional constraints that characterise natural-resource systems. Effective normalisation must regularise both marginal and joint behaviour while preserving spatial continuity and geological structure.

The comparative evaluation across the four datasets demonstrates that the choice of normalisation strategy exerts some influence on predictive accuracy, uncertainty calibration, and spatial structure. All transformations improved accuracy relative to modelling on raw variables, but the choice of method still depends on the interpretation and experience of the professional.

Acknowledgements

The authors acknowledge the Brazilian National Council for Scientific and Technological Development (CNPq) for financial support (process 407752/2022-6).

CRediT author statement

IG: Conceptualisation, methodology, software, investigation, writing; GN: Conceptualisation, methodology, investigation, writing.

References

Barnett, R.M., Manchuk, J.G., Deutsch, C.v. 2014. Projection Pursuit Multivariate Transform. Mathematical Geosciences, vol. 46, no3, pp. 337–359. https://doi.org/10.1007/s11004-013-9497-7

De Figueiredo, L.P., Schmitz, T., Lunelli, R., Roisenberg, M., Santana de Freitas, D., Grana, D. 2021. Direct Multivariate Simulation - A stepwise conditional transformation for multivariate geostatistical simulation. *Computers and Geosciences*, 147(November 2020), 104659. https://doi.org/10.1016/j.cageo.2020.104659

Gneiting, T., Raftery, A.E. 2007. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, vol. 102, no. 477,pp. 359–378. https://doi.org/10.1198/016214506000001437

Gonçalves, Í.G., Echer, E., Frigo, E. 2020. Sunspot Cycle Prediction Using Warped Gaussian Process Regression. *Advances in Space Research*, vol. 65, pp. 677–683. https://doi.org/10.1016/j.asr.2019.11.011

Gonçalves, Í.G., Guadagnin, F., Cordova, D.P. 2022. Learning spatial patterns with variational Gaussian processes: Regression. *Computers & Geosciences*, 161, 105056. https://doi.org/10.1016/j.cageo.2022.105056

Gonçalves, Í.G., Guadagnin, F., Cordova, D.P. 2023. Variational Gaussian processes for implicit geological modeling. *Computers & Geosciences*, vol. 174. https://doi.org/10.1016/j.cageo.2023.105323

Gonçalves, Í. G., & Wellmann, F. (2025). Uncertainty propagation in deep Gaussian process networks. *Mathematical Geosciences*, vol. 57, no. 6, pp. 1115–1133. https://doi.org/10.1007/s11004-025-10187-4

Goovaerts, P. 1997. Geoestatistics for Natural Resources Evaluation. *Oxford University Press.*

Grathwohl, W., Chen, R.T.Q., Bettencourt, J., Sutskever, I., Duvenaud, D. 2019. Ffjord: Free-form continuous dynamics for scalable reversible generative models. *7th International Conference on Learning Representations*, ICLR 2019, pp. 1–13.

Hensman, J., Matthews, A.G., Ghahramani, Z. 2015. Scalable variational Gaussian process classification. *Journal of Machine Learning Research*, vol. 38, pp. 351–360.

Hoffimann, J., Augusto, J., Resende, L., Mathias, M., Mazzinghy, D., Bianchetti, M., Mendes, M., Souza, T., Andrade, V., Domingues, T., Silva, W., Silva, R., Couto, D., Fonseca, E., Gonçalves, K. 2022. Modeling Geospatial Uncertainty of Geometallurgical Variables with Bayesian Models and Hilbert–Kriging. *Mathematical Geosciences*, vol. 54, no. 7, pp. 1227–1253. https://doi.org/10.1007/s11004-022-10013-1

Hyvärinen, A., Oja, E. 2000. Independent component analysis: algorithms and applications. Neural Networks, vol. 13, pp. 411–430. www.elsevier.com/locate/neunet

Jolliffe, I.T. 2002. Principal Component Analysis (2nd ed.). Springer.

Linnenbrink, J., Milà, C., Ludwig, M., Meyer, H. 2024. KNNDM CV: K-fold nearest-neighbour distance matching cross-validation for map accuracy estimation. *Geoscientific Model Development*, vol. 17, no. 15, pp. 5897–5912. https://doi.org/10.5194/gmd-17-5897-2024

Pawlowsky-Glahn, V., Egozcue, J.J., Tolosana-Delgado, R. 2015. Modeling and Analysis of Compositional Data. John Wiley & Sons.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830. http://scikit-learn.sourceforge.net.

Rasmussen, C.E., Williams, C.K.I. 2006. Gaussian processes for machine learning. MIT Press. https://doi.org/10.1142/S0129065704001899

Rousseeuw, P., Van Driessen, K. 1999. A fast algorithm for the minimum covariance determinant estimator. Technometrics, vol 41, no. 3. http://amstat.tandfonline.com/doi/abs/10.1080/00401706.1999.10485670

Snelson, E., Rasmussen, C.E., Ghahramani, Z. 2004. Warped Gaussian processes. Adv. Neural Inf. Proc. Syst., vol. 16, pp. 337–344. ◆

MINE SAFE 2026
MINE HEALTH AND SAFETY CONFERENCE

ENGINEERING HEALTH & SAFETY FOR SUSTAINABLE FUTURE-READY MINES



HYBRID CONFERENCE

11-12 NOVEMBER 2026 — CONFERENCE
13 NOVEMBER 2026 — INDUSTRY AWARDS DAY
GALLAGHER CONVENTION CENTRE, JOHANNESBURG, SOUTH AFRICA

ZERO HARM

Abstract submission – 31 July 2026 I Extended Abstract/Papers submission- 31 August 2026

BACKGROUND

We are excited to announce our upcoming industry Conference and Industry Awards Day, dedicated to advancing mining safety and health, innovation, and sustainability and recognising the excellent work in creating ZERO HARM future ready mines. This year's theme, "Engineering Health, Safety for Sustainable Future-Ready Mines", underscores our collective commitment to achieving zero harm while embracing modernisation, technology, and collaboration as the foundation for safer, future-ready mining operations.

This conference will serve as a vital platform for knowledge-sharing and idea exchange among key stakeholders, including mining companies, the Department of Mineral Resources and Energy (DMRE), the Minerals Council South Africa, labour unions, and health and safety practitioners at all levels in the minerals industry.

CONFERENCE THEMES

1. Driving Zero Harm Forward
- Reinforcing the vision of eliminating fatalities.
- Sharing progress, lessons learned, and renewed commitments.

2. Technology in Action
- Case studies of mines implementing breakthrough technologies.
- Practical examples of problems solved through innovation.

3. Future-Ready Mine Design
- Strategies for long-term sustainability and resilience.
- Integrating safety and health into planning and operational frameworks.

4. Engineering Safety Excellence
- Advances in engineering solutions that enhance safety.
- Best practices in design, maintenance, and risk management.

5. Engineering in the Digital Age
- Modernisation and digital transformation in mining operations.
- Leveraging data, automation, and AI for safer outcomes.

6. Technology Advancements and Innovation
- Exploring robotics, wearables, sensors, and predictive analytics.
- Emerging technologies shaping the future of mining safety.

7. Health and Wellbeing in Mining
- Protecting workers' physical and mental health.
- Occupational health strategies, wellness programs, and medical innovations.

8. Sustainability and Responsible Mining
- Environmental stewardship and climate-conscious practices.
- Linking safety with sustainability for long-term industry resilience.

Why Attend?
- Gain insights from global case studies and pioneering projects.
- Network with industry leaders, engineers, and technology innovators.
- Explore practical solutions for building safer, sustainable mines.
- Be part of shaping the future of mining through collaboration and innovation.

Join us as we work together towards Safe Mines, Healthy Lives, and Sustainable Futures!

PARTNERSHIP OPPORTUNITIES

Sponsorship opportunities are available. Companies wishing To partner on this event should contact the Conference Coordinator.

WHO SHOULD ATTEND

The conference should be of value to:
- Safety practitioners
- Mine management
- Mine health and safety officials
- Engineering managers
- Underground production supervisors
- Surface production supervisors
- Environmental scientists
- Minimizing of waste
- Operations manager
- Processing manager
- Contractors (mining)
- Including mining consultants, suppliers and manufacturers
- Education and training • Energy solving projects
- Water solving projects
- Unions
- Academics and students
- DMRE
- Occupational health practitioners
- Leadership and ManagementTechnical and Engineers

SUBMIT AN ABSTRACT

CALL FOR PAPERS

Call for papers on the topics of safety, health and environment
Prospective authors are invited to submit titles and abstracts of their presentations in English and not longer than 500 words. Abstracts should be submitted to:
Include Abstract Submission Deadline: **31 July 2026**
Extended Abstract/Paper Submission: **31 August 2026**

For further information contact:
Gugu Charlie, Conference Co-ordinators
E-mail:gugu@saimm.co.za | Tel: +27 11 538-0238
| Web: www.saimm.co.za