A Program for Data Transformations and Kernel Density Estimation

John G. Manchuk and Clayton V. Deutsch

Modeling applications in geostatistics often involve multiple variables that are not multivariate Gaussian after a normal score transformation. In these cases, other transformation techniques are necessary to obtain Gaussian variables. There are practically an unlimited number of transformations to choose from, a few of the more common ones for geological variables discussed in this paper are the normal score transform, log ratio transforms for compositional variables, data sphering which is similar to principal component analysis, and the stepwise conditional transform. Transformations are useful to prepare data for density estimation. Most non-parametric density estimators are sensitive to features of the data including outliers and long tails that can be detected and mitigated through data transformation. A kernel density network for estimating multivariate densities is discussed. Unlike classical kernel estimation that positions a kernel at each datum, networks are trained to best-fit the data with a limited number of kernels. Using fewer kernels improved efficiency of post-processing, such as computing densities, performing the stepwise transform, and performing inverse transforms. Kernel density networks are applied to a four variable nickel laterite data set and densities from the model are used to perform a continuous stepwise conditional transform. Marginal densities fit the data well visually and stepwise transformed variables are nearly uncorrelated.

1. Introduction

Multivariate data typically pose a challenge for statistical analysis and modeling because of the increased dimensionality of the problems and of the typically non-Gaussian shape of the multivariate distribution. It is possible to achieve Gaussianity of the univariate marginal distributions using a normal score transformation; however this does not guarantee that the bivariate and higher dimensional distributions are Gaussian. In fact, it is rare that the latter is true. A variety of data transformation procedures exist and can be used in an attempt to obtain Gaussian multivariate data. This is ideal for Geostatistics, especially for popular algorithms like sequential Gaussian simulation (SGS). When this cannot be accomplished, other non-parametric methods to estimate the multivariate density function can be used.

Data transformations and density estimates have a variety of uses in statistical analysis and geostatistical modeling. In applications such as regression and correlation analysis, transformations can be used to linearize the relationship between two variables, such as porosity and permeability or bitumen grade and fines. Logarithms and the normal score transform are common. Other transformations are useful for outlier detection, which is an important task for various unsupervised density estimation techniques. Searching for the optimal locations to place kernels or selecting the optimal kernel parameters is often sensitive to, and adversely affected by outliers. Centering and sphering the data is a useful transform for detecting outliers in multivariate data. The normal score transformation has various uses with univariate data. It provides a mapping to Gaussian space for modeling purposes. The normal score transform also removes heteroskedastic features like the proportional effect that is common with lognormal and exponential distributed variables.

Density estimation has several applications when the data are multivariate, but is usually not needed for univariate problems. For a single variable, kernel smoothing is the simplest and most likely the best method. In cases where data are not multivariate Gaussian after a normal score transform, density estimation can be used to quantify that non-Gaussian behavior. For SGS, the model can be used to transform the data to MV Gaussian using a continuous stepwise transform, removing the need for a complicated linear model of coregionalization. Density estimates can also be used to obtain more robust estimates of correlation between variables.

Some data transformations are appropriate for specific types or classes of data. A few common data types encountered in Geostatistics include continuous and often non-negative variables, compositional data, and categorical data. Transformations discussed in this paper are concerned with the first two and include: the normal score transform with and without tie breaking; the additive and centered log ratio transforms for compositional data; a data sphering transform; and the stepwise transform. Descriptions and applications of each are provided. Several of the transformations are useful for preparing data for multivariate density estimation. An adaptive kernel density estimation procedure that follows from Kramer and Johnston (1994) is developed. The stepwise transformation is applied using the densities directly from the model and is hence continuous.

2. Transformations

The list of transformations that can be applied to random variables is endless. A small set of common transformations used in geostatistics applications is considered, with two others not as common, but that are useful for density estimation. Before describing each of the transformations, some preliminary definitions are covered. First, define a multivariate data set with N random variables and n samples as $\mathbf{Z}$: $Z_{ij}, i = 1, \dots, n, j = 1, \dots, N$. For any transformation, define the resulting set of transformed variables as $\mathbf{Y}$: $Y_{ij}, i = 1, \dots, m, j = 1, \dots, M$, where the number of transformed samples, m, and transformed variables M, are not necessarily equal to n and N respectively as some transformations lead to a change in these dimensions.

Probably the most well known transformation in geostatistics is the normal score transform. It is a univariate transformation that operates independently on each variable. There are two requirements: a distribution function for each $Z_{ij}, \forall i, j = 1, \dots, N$, that is, the univariate marginal distributions of $\mathbf{Z}$, and an approximation for the inverse Gaussian distribution. Distributions for $Z_{ij}, \forall i, j = 1, \dots, N$ are typically estimated because they are unknown and do not tend to follow any particular family of parametric distribution function like the Gaussian, lognormal, Chi, etc. A typical choice is the empirical cumulative distribution function that assumes all samples have a density equal to $1/N$. Approximations to the inverse Gaussian distribution vary and are usually hidden from the user in all statistical software. They are required since the Gaussian density function cannot be integrated exactly. Let $F_j(\mathbf{Z})$ be the density function for variable j of $\mathbf{Z}$ and $G^{-1}(p)$ represent the inverse Gaussian distribution, then the normal score transform is defined by Equation (1).

$$Y_{ij} = G^{-1}\left(F_j(Z_i)\right), i = 1,...,n,\ j = 1,...,N \tag{1}$$

The normal score transform is fairly rigid. Some flexibility is possible through how $F_j(\mathbf{Z})$ is estimated. One characteristic of the data that can be handled in different ways is the occurrence of repeated values, referred to as ties. A common approach to break ties is to add a small random deviate to each Z_{ij}. If there are repeated vectors in the data, that is, $Z_{ij} = Z_{kj} \forall j, i \neq k$, the random deviates must be different for each variable, otherwise the resulting normal score values will form a straight line in N-dimensional space, rather than a cloud, and can significantly alter statistics such as the correlation matrix of $\mathbf{Y}$. Another way to handle ties is to assign them all the same normal score values, which is very similar to the transformation used in mapping categories to Gaussian values for truncated Gaussian simulation. In this case, a histogram of the resulting univariate distributions, $f_j(\mathbf{Y})$, will not necessarily appear Gaussian. There will be spikes where a significant number of repeated values exist.

For purposes of density estimation, is it often desirable to transform all variables to have equal spread. A transformation called sphering (Fukunaga, 1972), Equation (2), achieved this where $\mathbf{C}$ and $\boldsymbol{\mu}$ are the covariance matrix and mean vector of $\boldsymbol{Z}$ respectively and $\boldsymbol{C}^{-1/2}$ is computed from the spectral decomposition of $\mathbf{C}$ in Equation (3). Transformed variables are centered and have unit variance and zero covariance.

$$\mathbf{Y} = \mathbf{C}^{-1/2}\left[\mathbf{Z} - \boldsymbol{\mu}\right] \tag{2}$$

$$\mathbf{C} = \mathbf{V}\mathbf{D}\mathbf{V}^T \qquad \mathbf{C}^{-1/2} = \mathbf{V}^T\mathbf{D}^{-1/2}\mathbf{V} \tag{3}$$

Data sphering can also be used for outlier detection and removal that is important for density estimation as outliers can have a significant impact on results. Outliers can be identified and removed by an iterative process where samples having a norm larger than a specified threshold, $\|\mathbf{y}\| > \alpha$, are removed and data a sphered again (Hwang et al, 1994). The process continues until no samples are considered outliers.

When random variables are compositional in nature, another class of transformations called log ratios is useful (Pawlowsky-Glahn, 2004). Compositional data consist of components of a whole and therefore sum to a constant, usually equal to one. Because of this, the data exist in an N-dimensional simplex. Log ratio transforms map the data from the simplex to coordinate systems where traditional statistical analyses can be applied. Further detail is available in paper ### in this report. Two log ratio transformation are discussed briefly: the additive (ALR) and centered log ratio (CLR) transforms. A limitation of these transformations is the random variables must be positive.

Forward and inverse ALR transformations are expressed by Equations (4) and (5). The denominator, z_k, can be any one of the components of $\mathbf{Z}$, but it must remain consistent when applying this transform to a complete set of compositions. The number of variables is $M = N - 1$. Note that the resulting space is not isometric.

Instead of being orthogonal, the coordinate axes are separated by 60 degrees (Pawlowsky-Glahn and Egozcue, 2006).

$$Y_{ij} = \log\left(Z_{ij} / Z_{ik}\right), \quad i = 1,...,n,\ j = 1,...,N,\ j \neq k \tag{4}$$

$$Z_{ij} = \frac{\exp\left(Y_{ij}\right)}{\sum_{j=1,\, j\neq k}^{N} \exp(Y_{ij}) + 1}, \quad i = 1,...,n,\ j = 1,...,N \tag{5}$$

The CLR transform results in orthogonal axes which simplifies further multivariate computations; however, the data are still constrainted to a subspace of $\mathbb{R}^N$. Transformed vectors have a zero sum, hence the subspace is a plane. Equation (6) is the forward CLR transform where $g(Z_{ij})$ is the geometric mean of $Z_{ij} \forall j$ and Equation (7) is the inverse transform.

$$Y_{ij} = \log\left(Z_{ij} / g(Z_{ij})\right), i = 1,...,n,\ j = 1,...,N \tag{6}$$

$$Z_{ij} = \frac{\exp(Y_{ij})}{\sum_{j=1}^{N} \exp(Y_{ij})}, i = 1,...,n,\ j = 1,...,N \tag{7}$$

A common transformation used in geostatistical modeling of multivatiate data is the stepwise conditional transform (Leuangthong and Deutsch, 2003). A continuous version of this transformation is applied after a density model is obtained and will be discussed in the following section.

3. Kernel Density Estimation

Kernel density estimation (KDE) is probably one of the most widely used approaches. In most applications, a probability density function is fit by placing a kernel at each datum. Gaussian kernels are typical. The width or variance of each kernel is chosen to achieve an acceptable fit and this tends to be a subjective process. In the case of fixed width kernel estimators, all kernels are assigned the same width parameters. A class of adaptive kernel density estimator allows all kernels to have different widths. Two downfalls of these approaches are the challenge in obtaining a good fit in both the tails and main body of a distribution and the curse of dimensionality (Hwang et al, 1994, Silverman, 1986).

The computational demand of KDE becomes substantial as the number of samples grows, making it impractical for applications in geostatistical modeling. Obtaining a density estimate at a specific vector, $\mathbf{z}$, from the model requires evaluating the distance to all samples and the density from the kernels at each sample. Computing a density is done using Equation (8), where h is the width parameter and $\varphi_i(\|\mathbf{z} - \mathbf{z}_i\|, h)$ is a kernel function centered at $\mathbf{z}_i$.

$$f(\mathbf{z}) = \frac{1}{nh^N} \sum_{i=1}^{n} \varphi_i\left(\left\|\mathbf{z} - \mathbf{z}_i\right\|, h\right) \tag{8}$$

To improve efficiency, a class of KDE called kernel density networks (KDN) fits the density using a limited number of kernels. These networks are also referred to as elliptic basis function networks (Johnston and Kramer, 1994). Radial basis function networks are also similar (Yee and Haykin, 2001). Kernels can be fixed width or have adaptive widths as with KDE. It is also possible to introduce anisotropy, that is, the covariance matrix for each kernel is not restricted to an identity matrix.

Three parameters must be determined to derive a KDN: 1 – the number of kernels, L; 2 – the locations or means for the kernels, $\boldsymbol{\mu}_l, l = 1, \ldots, L$; and 3 – the covariance matrix of each kernel, $\boldsymbol{\Sigma}_l, l = 1, \ldots, L$. Various methods exist for selecting L such as orthogonal least squares (Bishop, 1995) and cross validation (Kramer and Johnston, 1994). Determining parameters 2 and 3 is accomplished using an iterative optimization process involving K-means clustering in the following workflow:

1. Initialize kernel centers $\boldsymbol{\mu}_l, l = 1, \ldots, L$ using a random subset from $\mathbf{Z}$ and initialize $\boldsymbol{\Sigma}_l, l = 1, \ldots, L$ to identity matrices. This defines an initial set of isotropic Gaussian kernels.
2. Update $\boldsymbol{\mu}_l, l = 1, \ldots, L$ using K-means clustering. The Mahalanobis distance is used.

3. Update the kernel covariance matrices based on the updated centers. There are two approaches for this discussed later.
4. Repeat Steps 2 and 3 until mean and covariance parameters do not change.

Computing the updated covariance matrices in Step 3 is done either locally or globally. The local update uses samples from each K-means cluster to compute a weighted covariance matrix by Equation (9), where S_l is the set of samples in cluster l, w_i is a weight assigned to sample i computed by Equation (10), and W is the sum of the weights. In Equation (10), h is a width parameter that is equal to the average distance from kernel center l to the nearest O clusters, Equation (11).

$$\Sigma_l = \frac{1}{W}\sum_{i\in S_l} w_i \left(\mathbf{z}_i - \boldsymbol{\mu}_l\right)\left(\mathbf{z}_i - \boldsymbol{\mu}_l\right)^T$$
$$W = \sum_{i\in S_l} w_i \tag{9}$$

$$w_i = \exp\left(-\left\|\mathbf{z}_i - \boldsymbol{\mu}_l\right\|^2 / h_l^2\right) \tag{10}$$

$$h_l = \left[\frac{1}{O}\sum_{k=1}^{O_l}\left\|\boldsymbol{\mu}_k - \boldsymbol{\mu}_l\right\|^2\right]^{1/2} \tag{11}$$

The global update uses all the sample data in Equation (9) rather than those from each cluster. They are still weighted by Equation (10). In the local approach, resulting density estimates will fit the local characteristics of the sample set, but are less smooth than the global approach. For data sets that appear multi-modal or that have clear non-linear features, the local approach may be better suited. The global approach may be better suited to data sets that appear smooth or uni-modal.

In both cases, it is possible for the covariance matrices to become highly anisotropic, for example, if a cluster becomes small during iteration and points fall approximately in a line. To handle such behavior, the anisotropy is constrained to a maximum α using the following procedure:

1. Compute the decomposition $\boldsymbol{\Sigma}_l = VDV^T$ where D is a diagonal matrix of eigenvalues and V is an orthonormal matrix of eigenvectors.
2. Check the anisotropy of $\boldsymbol{\Sigma}_l$: $\alpha_l = \max(D)/\min(D)$
3. If $\alpha_l > \alpha$
    a. Let $\min(D) = \max(D)/\alpha$
    b. Update the covariance using $\boldsymbol{\Sigma}_l = VDV^T$
    c. Iterate from Step 1 using the updated $\boldsymbol{\Sigma}_l$

This process ensures the anisotropy is constrained for all dimensions involved.

Once a density model is obtained, the probabilities can be used directly for the stepwise conditional transformation. Since Gaussian kernels are used, all marginal and conditional probabilities can be computed. For a sample, $\mathbf{z} = [z_1\ z_2\ \ldots\ z_N]$, the transformation progresses as follows:

1. Compute the cumulative marginal probability of z_1 from the set of kernels using
$$F(z_1) = \sum_{l=1}^{L} \lambda_l G(z_1, \mu_1, \Sigma_{11})$$
where λ_l are local densities that ensure the area under the kernels integrates to 1 and $G(z, m, s)$ is the cumulative Normal distribution with mean m and variance s evaluated at point z.
2. Compute the cumulative probability of z_2 conditional to z_1 based on the marginal bivariate distribution between z_1 and z_2. The conditional mean and variance are defined by

$$\mu_2(Z_1 = z_1) = \mu_2 + \Sigma_{12}\Sigma_{11}^{-1}(z_1 - \mu_1)$$
$$\Sigma_{22}(Z_1 = z_1) = \Sigma_{22} - \Sigma_{12}\Sigma_{11}^{-1}\Sigma_{21}$$

Then the cumulative probability is defined as

$$F(z_2 \mid Z_1 = z_1) = \sum_{l=1}^{L} f_l(z_1)\lambda_l G(z_2, \mu_2(Z_1 = z_1), \Sigma_{22}(Z_1 = z_1))$$

Where $f_l(z_1)$ is the marginal density of z_1 from kernel l.

3. Continue on for all variables involved, with z_3 conditional to z_2 and z_1, and so on.
4. Compute the inverse Gaussian distribution for all cumulative probabilities accumulated from Steps 1, 2 and 3, similar to Equation (1).

4. Program and Parameters

A program called MVDENSITYFIT was compiled to perform all data transformation and construct a KDN. Parameters are as follows:

Table 1: Parameters for MVDENSITYFIT

Line	Parameters	
1	data.dat	-file with data
2	2 4 5	- number of variables and columns
3	-998. 9999.	- trimming limits
4	2 3 2	-number of transforms and types (see notes)
5	0.03 1	-flags for transform 1
6	10	-number of kernels for density estimation
7	0	-global (0) or local(1) covariance
8	0 8	-kernel type and overlap (see notes)
9	5	-maximum kernel anisotropy
10	5	-maximum iterations
11	trans.out	-file base-name for transformation output
12	1	-output densities to a grid: 1-yes (mainly for checking)
13	100 100	- number of cells for each variable
14	densities.out	- file for density output
15	1	-generate and output a random sample, 1-yes
16	10000	- number of samples to generate
17	randomsamples.out	- file for sample output
18	1	-perform stepwise conditional transform, 1-yes
19	stepwise.out	- file for stepwise output
20	1	-perform all transforms forward (1) or inverse (-1)
21	mvmodel.out	-file for multivariate density model input/output
22	sgsim.out	-file with data to inverse transform
23	2 1 2	- number of variables and columns
24	100 100 1	- nx, ny, nz per realization, set to zero if not this format

Most parameters are explained adequately in Table 1; however, some require further explanation. Data files are all typical GSLIB format. The transformations and types on Line 4 are: 1 – normal scores; 2 – normal scores without tie breaking; 3 – ALR; 4 – CLR; 6 – unique values; 7 – sphere. The unique values utility was not explained above. It removes all duplicate vectors from a set of data, which can be useful to remove large spikes of zeros for example. Some transforms require additional parameters and Line 5 is an example for the ALR, which includes a small value to replace zeros (0.03 in this case) and the variable to use as the denominator (k in Equation (4)). The kernel type parameter on Line 8 has been added as a future version will have other kernels available, for example the Bartlett kernel. The overlap parameter also on Line 8 defines the number of nearest kernels to use in computing the width parameter h_l in Equation (11). A maximum iteration parameter is available on Line 10 as convergence can be slow in some cases and users may want to stop the fitting process early. The file base name on Line 11 is used to generate files with results from each of the transformations requested on Line 4. In this case, output files would be called trans_1.out for ALR and trans_2.out for the normal scores without tie breaking.

It is possible to perform transformations forward and inverted with the parameter on Line 20. For forward transformation, the density model is output; for inverse, it is read in by the program. This file contains pertinent information related to the transforms and the resulting locations and covariances of the kernel density network. Data for inverse transformations can be from geostatistical simulation for example.

5. Examples

Four examples are provided to demonstrate the KDN approach and program. Examples 1 – 3 are synthetic cases involving sample data derived from Gaussian mixture models for validation purposes. The fourth example involves compositional data from an actual Nickel Laterite deposit. Parameters for the synthetic examples are as follows:

1. A set of 1000 samples drawn from a single trivariate Gaussian kernel with mean zero, unit variance, and covariance between variables 1 and 2 of 0.7, between 1 and 3 of 0.3, and between 2 and 3 of 0.5.
2. A set of 1200 samples drawn from a mixture model with three trivariate Gaussian kernels with means $[1\ 1\ 0]$, $[2\ 1.75\ 0]$, $[1.5\ 3\ 0]$ (400 samples per kernel). Correlation between variables 1 and 2 for kernel 1 is 0.5; between variables 1 and 3 for kernel 2 is 0.5; and between 2 and 3 for kernel 3 is 0.5. All other correlations are zero. The variance of all variables is 0.1 to provide some separation between the clusters.
3. Same samples and covariance matrices as 2, but with means of $[1\ 1\ 0]$, $[2\ 1\ 0]$, $[1.5\ 2\ 0]$ to provide more overlap of the sample data.

Quality of the estimated density functions is assessed by comparing the known kernel parameters to the estimated parameters as well as the true to estimated densities. Comparison of the kernel means is provided in Table 2, variances in Table 3, and covariances in Table 4. Estimates of the density at each of the sample points were compared to the known densities (Figure 1). Models match inputs well in all cases and correlation coefficients between true and estimated densities are high, greater than 0.98. In the third case, variances are slightly underestimated, which leads to overestimation of the larger densities (estimates drift above the 45 degree line for high density values in Figure 1). Of these three cases, the most interesting for applying the stepwise transformation was case 2 (Figure 2). Any interesting features such as non-linearity and multi-modality have been removed. Mean values are all close to zero and variances close to one. Correlation coefficients of the stepwise transformed variables are no larger than 0.015 in absolute value.

Table 2: Comparison of kernel means for synthetic cases.

Case	Kernel	Input Mean			Output Mean			SSE
1	1	0	0	0	0.002	0.015	-0.003	5.39E-06
2	1	1	1	0	0.985	0.972	-0.026	2.20E-04
	2	2	1.75	0	1.977	1.759	0.028	5.40E-04
	3	1.5	3	0	1.487	3.024	-0.004	1.59E-04
3	1	1	1	0	1.019	0.988	-0.015	3.71E-04
	2	2	1	0	2.061	0.979	-0.023	3.75E-03
	3	1.5	2	0	1.520	2.016	0.004	4.05E-04

Table 3: Comparison of kernel variance for synthetic cases

Case	Kernel	Input Variance			Output Variance			SSE
1	1	1	1	1	1.035	0.966	0.960	1.22E-03
2	1	0.1	0.1	0.1	0.096	0.093	0.107	1.33E-05
	2	0.1	0.1	0.1	0.103	0.100	0.097	7.30E-06
	3	0.1	0.1	0.1	0.106	0.091	0.095	3.20E-05
3	1	0.1	0.1	0.1	0.091	0.083	0.109	7.84E-05
	2	0.1	0.1	0.1	0.080	0.092	0.086	4.04E-04
	3	0.1	0.1	0.1	0.097	0.093	0.095	1.07E-05

Table 4: Comparison of kernel covariance for synthetic cases

Case	Kernel	Input Covariance			Output Covariance			SSE
1	1	0.7	0.3	0.5	0.715	0.277	0.465	2.18E-04

2	1	0.05	0	0	0.039	0.004	0.004	1.20E-04
	2	0	0.05	0	-0.002	0.049	-0.004	3.95E-06
	3	0	0	0.05	-0.001	-0.002	0.050	6.62E-07
3	1	0.05	0	0	0.039	0.001	0.003	1.16E-04
	2	0	0.05	0	-0.001	0.041	-0.001	6.32E-07
	3	0	0	0.05	0.005	-0.004	0.043	2.23E-05

A four-variate density function is estimated for the Nickel example. The variables are percent grade of Ni, Fe, SiO_2 and MgO and they are compositional. Not all samples sum to 100 % so a fifth variable is added that is equal to the remainder. Two transformations are applied to the data: 1 – the additive log ratio transform is used with the remainder variable as the denominator, and 2 – the data are sphered, so variables have unit variance and zero covariance. The density function is fit to the sphered variables and then the stepwise transform is applied to obtain a set of Gaussian variables for other processes such as SGS.

Data consist of a total of 18270 samples. A random subset of 4363 samples was used for density estimation. The number of kernels chosen was based on checking the quality of fit for the univariate and bivariate marginal densities. Since the densities of data is not known ahead of time, this was done visually; however, one could consider cross validation techniques to be more rigorous in other studies. After some iteration, 50 kernels with an overlap of 5 nearest neighbors were chosen. Covariance matrices of each kernel were updated globally. Resulting univariate and bivariate densities are shown in Figure 3 and stepwise transform results in Figure 4. The stepwise transform results in variables that are nearly uncorrelated, with the absolute maximum correlation not exceeding 0.085. Non-zero correlation indicates that some regions of the multivariate density are not covered by the sample data. One may consider applying a normal score transform on each stepwise transformed variable independently to reduce the correlation and obtain variables with zero mean and unit variance.

6. Conclusions and Future Work

This work applied kernel density estimation networks to model multivariate densities. Having a density model permits the stepwise conditional transform to be applied in a continuous fashion, using the actual density estimates rather than a multivariate histogram. Effects of binning are not observed in the final cross plots between various stepwise transformed variables. The method was shown to accurately reproduce known densities for basic trivariate Gaussian mixture models. With actual data, the true density is not known and checking results is more challenging. This motivates an area of future work to develop tools to check multivariate density estimates. For example, when enough data are available a multivariate histogram could be used to compare density in the histogram bins to density predicted by the model. Cross validation tools to select parameters for the KDN are also needed.

References

Bishop, C.M., 1995, Neural networks for pattern recognition. Oxford University Press, 504

Fukunuga, K., 1990, Introduction to statistical pattern recognition. Academic Press, 592

Hwang, J.-N., Lay, S.-R., Lippman, A., 1994, Nonparametric multivariate density estimation: a comparative study. IEEE Transactions on Signal Processing, 42(10), pp. 2795-2810

Johnston, L.P.M and Kramer, M.A., 1994, Probability density estimation using elliptic basis functions. American Institute of Chemical Engineers (AIChE) Journal, 40(10), pp. 1639-1649

Leuangthong, O., and Deutsch, C.V. (2003) Stepwise conditional transformation for simulation of multiple variables, Mathematical Geology, Vol. 35, No. 2, pp. 155-173

Pawlowsky-Glahn, V., 2004, Geostatistical analysis of compositional data. Oxford University Press, 304

Pawlowsky-Glahn, V., and Egozcue, J.J., 2006, Compositional data and their analysis: an introduction: in Buccianti, A., Mateu-Figueras, G., and Pawlowsky-Glahn, V. (eds), Compositional data analysis in the geosciences: from theory to practice: Geological Society, London, Special Publications, 264, pp. 1-10

Silverman, B.W., 1986, Density estimation for statistics and data analysis. Chapman and Hall, 176

Yee, P.V. and Haykin, S., 2001, Regularized radial basis function networks: theory and applications. John Wiley & Sons, 208

Figures

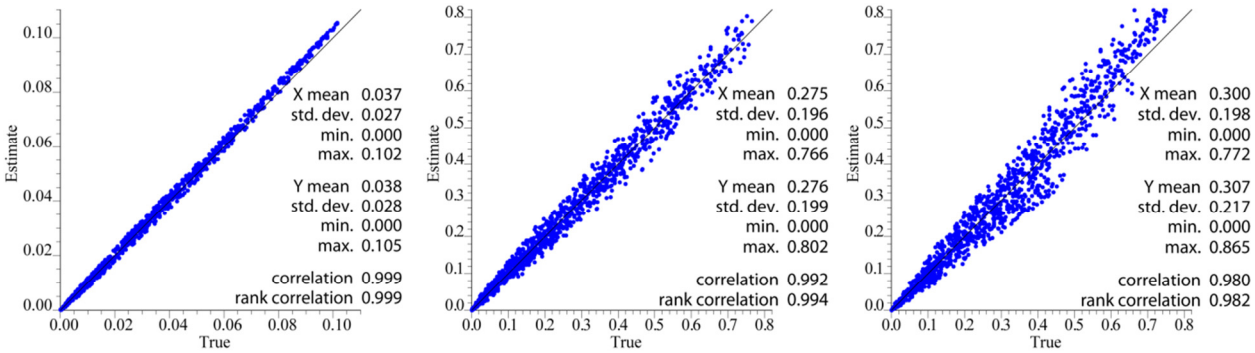

Figure 1: Cross plots of true vs estimated densities for examples 1, 2 and 3.

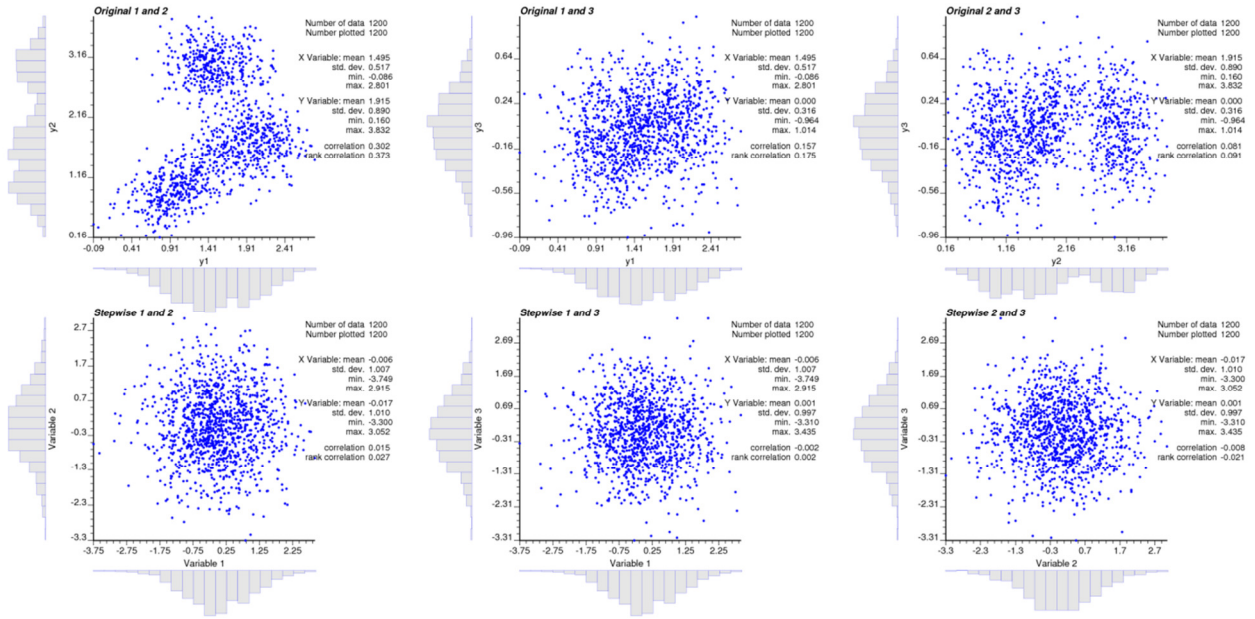

Figure 2: Cross plots of original (top) and stepwise transformed (bottom) variables from case 2.

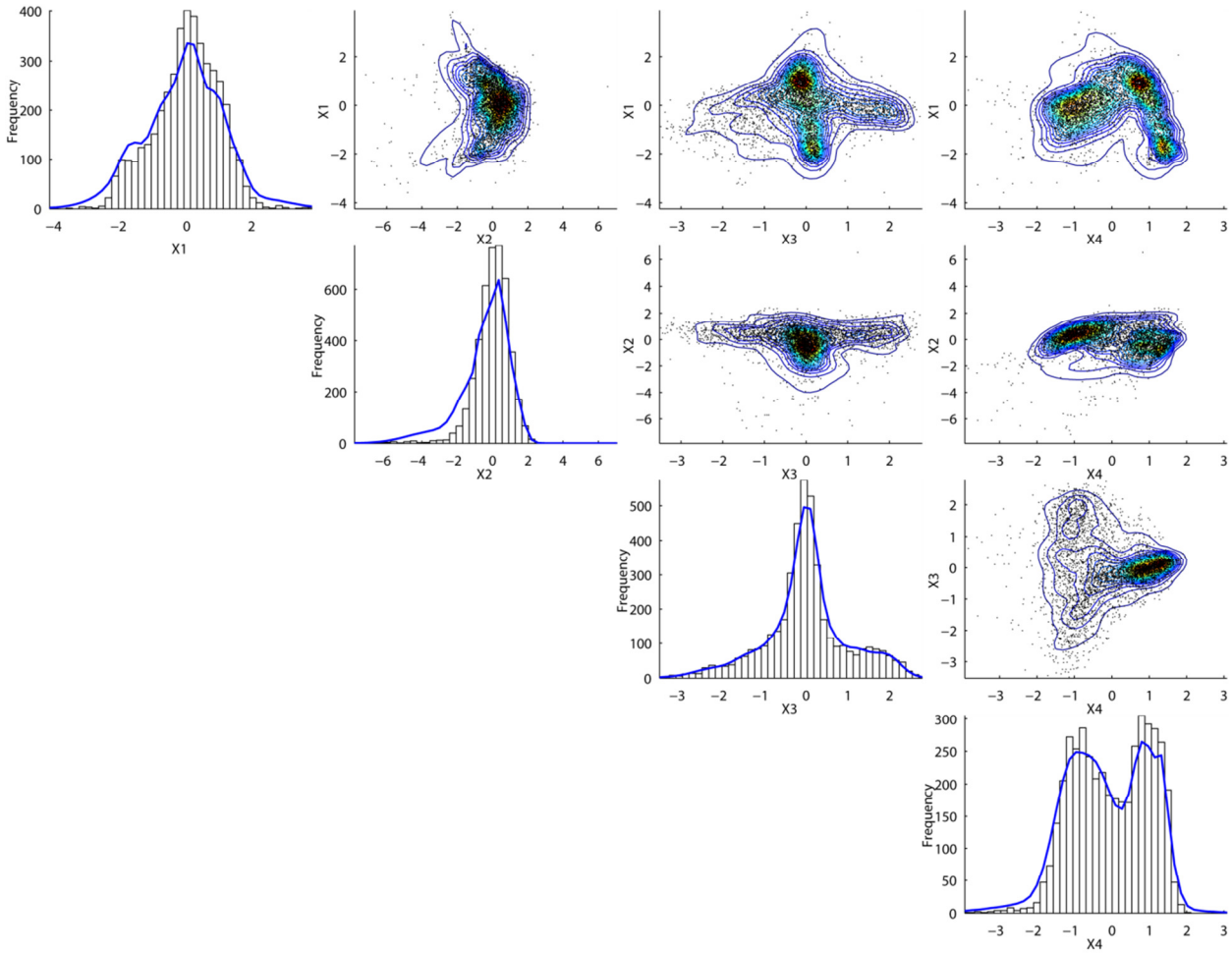

Figure 3: Univariate and bivariate marginal density estimates for the nickel laterite data. X1 to X4 are the sphered, ALR transformed variables corresponding to Ni, Fe, SiO_2 and MgO.

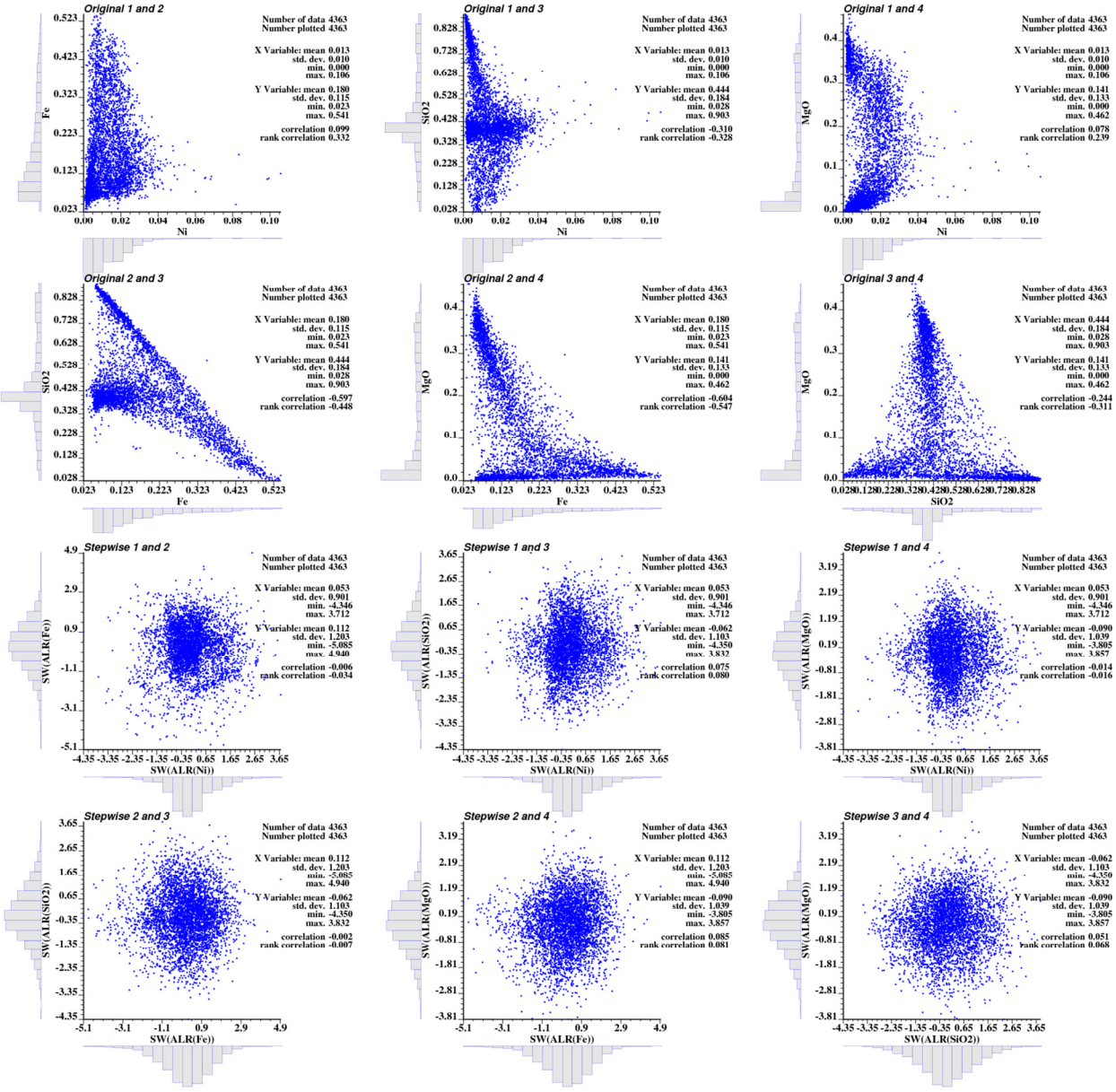

Figure 4: Original nickel laterite variables (top two rows) and stepwise transformed variables (bottom two rows).