Why Logratios are a Bad Idea for Multiscale Facies Modeling

Zhou Lan, Oy Leuangthong and Clayton V. Deutsch

Centre for Computational Geostatistics
Department of Civil and Environmental Engineering
University of Alberta

Indicator kriging is an important unbiased estimator and is widely used in multiscale facies modeling, but applying indicator kriging will sometimes lead to estimated facies proportions which are negative or do not sum to one. Logratio transformation can help solve this problem and is attracting the interest of more and more geostatisticians. However, two critical problems are unavoidable in logratio transformation and thus make it not suitable for the multiscale facies modeling. This paper will discuss the main procedures of multiscale facies modeling; the basic formulation and common application of logratio transformation; and problems of logratio transformation in multiscale facies modeling. Finally, a brief discussion on the alternative approaches in solving the negative proportion and sum constraint problem in indicator kriging is presented.

Introduction

The distribution of lithofacies is one of the most important factors in reservoir modeling. Many critical petrophysical properties in reservoir analysis and estimation, such as porosity and permeability, are highly correlated with facies type. A realistic model capturing the spatial distribution of various facies categories over the area of interest will be significantly helpful for modeling of reservoir characteristics and performance. Various approaches are used in multiscale facies modeling, among which are indicator kriging and sequential indicator simulation which are widely applied and are proven to be practical and efficient in modeling the spatial uncertainty of facies distribution.

An issue arises with indicator kriging because it can lead to a negative proportion of the predicted facies and/or the predicted proportions of all the facies over the same area do not sum to 1. Logratio transformation is an alternative to solve this constant-sum constraint and its special properties make it frequently applied in compositional data analysis. However, two longstanding issues remain a challenge in the logratio approach: the occurrence of zero proportion of certain facies categories in any sampled or un-sampled location will make the transformed logratio non existent; furthermore, the nonlinearity of logratio transformation can lead to bias when we perform linear estimation and back transform the result to the original data. These two issues make logratios unsuitable for multiscale facies modeling.

In Part 1 of this paper, we give a brief description on the background and available approaches to multiscale facies modeling. In Part 2, we introduce the essential formalism of the logratio transform and its common applications. In Part 3, we discuss the application and problems of applying logratios for multiscale facies modeling. We conclude with a brief discussion on some alternatives that may help solve the negative proportion and sum constraint problems in the indicator kriging and simulation approach.

Background on Multiscale Facies Modeling

Multiscale facies modeling is often applied in reservoir analysis. In the production process and various geological and reservoir tests, people obtain facies information at the sampled locations. As shown in Figure 1, suppose we have K (say, 4) facies, F_1, …, F_K in a 3-dimensional space V, and use $F(\mathbf{u})$ to denote the facies type at point $\mathbf{u}$. For a single well in this space, we have a vertical series of these 4 facies. Equivalently, for each small sampled location point $\mathbf{u}_\alpha$, we have an indicator variable defined as:

$$I(\mathbf{u}_\alpha;k)=\begin{cases}1 & \text{if } F(\mathbf{u}_\alpha)=F_k \\ 0 & \text{otherwise}\end{cases}$$

where $k=1, \ldots, K$. Scaling the indicator over a block volume $\mathrm{v}_\beta \subseteq V$, we obtain the proportion of the k^{th} facies $p_{k\beta}$ as

$$p_{k\beta}=\frac{1}{v_\beta}\int_{\mathbf{v}_\beta} I(\mathbf{u};k)d\mathbf{u}, \quad \mathbf{u}\in v_\beta$$

Depending on the scale of support v_β, different values and distributions will be obtained for facies proportion $p_{k\beta}$.

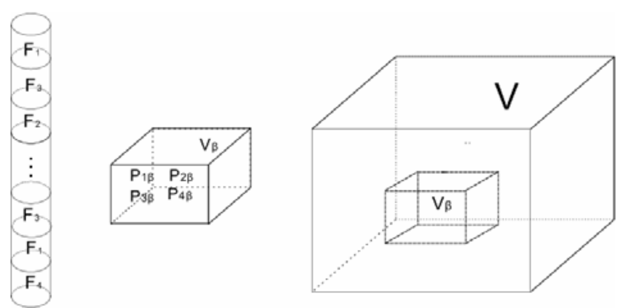

Figure 1: Multiscale facies modeling.

This scale-dependent transition from a categorical to continuous property poses two interesting issues. Firstly, inferring facies proportion, p_{k0}, for each facies category at a specified volume centred at an unsampled location $\mathbf{u}_0$ based on available point sampled facies data is a non-trivial problem. Secondly, once the facies proportions for a specific volume can be inferred, extending this distribution of proportions to reflect alternate scales of support remains an outstanding challenge.

Suppose we consider n sampled points data $I(\mathbf{u}_\alpha;k)$, ($\alpha=1,2,\ldots,n$) within a certain neighborhood of an unsampled location $\mathbf{u}_0$. We assume that a prior mean facies proportion $\tilde{p}_{k0}$ is stationary over this neighborhood, then for the k^{th} facies category, the probability p_{k0} at location $\mathbf{u}_0$ can be estimated by indicator kriging (IK) as follows:

$$\hat{p}_{k0}^{IK} = \sum_{\alpha=1}^{n} \lambda_{k\alpha} I(\mathbf{u}_\alpha; k) + \left[1 - \sum_{\alpha=1}^{n} \lambda_{k\alpha}\right] \cdot \tilde{p}_{k0}$$

where $\lambda_{k\alpha}$ is the weight corresponding to the indicator data at location $\mathbf{u}_\alpha$ for category k, $k=1, \ldots, K$. The weights $\lambda_{k\alpha}$, $\alpha=1, \ldots, n$ are determined by minimizing the kriging variance:

$$\sigma_{IK}^2 = E[\hat{p}_{k0}^{IK} - p_{k0}]^2$$

and thus reaching the simple kriging system:

$$\sum_{\beta=1}^{n} \lambda_{k\beta} C_I(\mathbf{u}_\alpha, \mathbf{u}_\beta) = C_I(\mathbf{u}_0, \mathbf{u}_\alpha) \quad \forall \alpha = 1, 2, \ldots, n$$

where $C_I(\mathbf{u}_\alpha,\mathbf{u}_\beta)$, is the indicator covariance between locations $\mathbf{u}_\alpha$ and $\mathbf{u}_\beta$.

The extension of this framework to simulation, known as sequential indicator simulation (SIS), can provide a distribution of uncertainty of facies proportions at an unsampled location. The general methodology for SIS is as follows:

1. The entire space is gridded and all the grid nodes are visited via a random path
2. At each grid node $\mathbf{u}_\nu$:
    (a) Search for nearby data and previous simulated values;
    (b) Perform IK to estimate the facies probability for each category at this grid node and build the cumulative probability function;
    (c) Draw a simulated category for the grid node based on the distribution function built in (b) above.

Unlike IK which yields estimated proportions, SIS returns a simulated category. Thus, while SIS constructs a local distribution of facies proportion, Monte Carlo simulation returns only a discrete value.

The problem of order relations is not new to indicator methods. The requirements to satisfy order relations are twofold: (1) $0 \le p_{k\beta} \le 1 \ \ \forall k$ and β ; and (2) $\sum_{k=1}^{K} p_{k\beta} = 1 \ \ \forall \beta$. Unfortunately, these two constraints are not automatically guaranteed under indicator kriging. Various corrections have been suggested to solve this problem. The logratio transformation of the facies proportions is attracting the interest of more and more geostatisticians due to its special properties.

Logratio Formalism

A logratio value, denoted by r_k, for the facies proportion p_k is defined as:

$$r_k = \log[\frac{p_k}{p_q}], \quad k = 1, \ldots, q-1, q+1, \ldots, K$$

where the denominator p_q can be any one fixed proportion among p_k, $k=1, \ldots, K$. This transform is reversible via:

$$p_k = \frac{\exp(r_k)}{1 + \sum_{l=1, l \neq q}^{K} \exp(r_l)} \quad \text{for } k = 1, ..., q-1, q+1, ..., K$$

and

$$p_q = \frac{1}{1 + \sum_{l=1, l \neq q}^{K} \exp(r_l)}$$

Directly from the above formulas, we have $0 \le p_k \le 1$ and $\sum_{k=1}^{K} p_k = 1$ for any given set of $\{r_1, ..., r_{q-1}, r_{q+1}, ..., r_K\}$ and thus satisfies the order relation requirements as mentioned above.

Aitchison (1986) gave a detailed introduction on logratio analysis for compositional data. Some notable properties of the transform include:

- There is a one-to-one correspondence between the original data (p_1, …, p_K) and the logratio vector (r_1, …, r_{q-1}, r_{q+1}, …, r_K), therefore any statement in terms of logratios can be expressed as an equivalent statement in terms of it original components. It turns out that after a complete circle of forward and back transformation, the result is independent of the selection of the denominator p_q in $\log\left[p_k / p_q \right]$ (Aitchison, 1999).
- Logratio inference obtained from any subcomposition ($p_{(1)}$, …, $p_{(d)}$), where $d \le K$, from the parent composition (p_1, …, p_K) will be exactly the same as the inference from the parent composition provided that we apply the same component as the denominator (Aitchison, 1986). For instance, suppose x_1, …, x_K denote the number of samples belonging to facies F_1, …, F_K, respectively; and $x_{(1)}$, …, $x_{(d)}$ denote, respectively, number of samples belong to facies $F_{(1)}$, …, $F_{(d)}$ and the set of $F_{(1)}$, …, $F_{(d)}$ is a subset of F_1, …, F_K, that is { $F_{(1)}$, …, $F_{(d)}$} $\subseteq${F_1, …, F_K}. Then the logratio r_k for {F_1, …, F_K} is

$$r_k = \log\left(\frac{x_k}{\sum_{l=1}^{K} x_l} \Big/ \frac{x_q}{\sum_{l=1}^{K} x_l} \right) = \log \frac{x_k}{x_q}$$

  and for {$F_{(1)}$, …, $F_{(d)}$},

$$r_{(k)} = \log\left(\frac{x_{(k)}}{\sum_{l=1}^{d} x_{(l)}} \Big/ \frac{x_{(q)}}{\sum_{l=1}^{d} x_{(l)}} \right) = \log \frac{x_{(k)}}{x_{(q)}}$$

  Therefore, r_k=$r_{(k)}$ whenever $F_{(k)}$=F_k and $F_{(q)}$=F_q. This conclusion permits the application of logratios on any known collection of facies collection, even in regions where the exact number of facies are unknown.

- The logarithmic operation on the ratio p_k/p_q often leads to an approximately normally distributed data which satisfies the distribution assumption for linear or non-linear regression models and many other statistical tests.
- The covariance structure of the original compositional data can be expressed in terms of a logratio covariance structure, determined by matrix

$$\mathbf{C} = [\mathrm{cov}\{\log(p_k / p_q), \log(p_l / p_q)\}] \quad \text{where } k, l = 1, \ldots, K$$

Further details about the logratio covariance structure are discussed by Aitchison (1986).

Due to the above characteristics, logratio transformations and modeling is frequently applied in compositional data analysis in various fields such as ecology, geology and environmental science, where original compositional data is transformed to logratio values. Then a series of statistical analysis are applied on the logratio data:

A linear or non-linear model can be built regarding the logratio values and against a series of regressor variables acting as factors that will determine the compositional variable. In this way, the conditional logratio values can be estimated based on the given regressor variables; further, the significance of each regressor can be tested. The fitted logratio values are then back transformed to get the estimated compositional value given the values of the regressors.

In geostatistics, simple kriging can be applied on the known logratio values in the sampled area to estimate the conditional logratio value for an unsampled location and then back transformed to get estimated proportions.

Application of Logratios to Facies Modeling

Given the explicit control of order relations in logratio transformation, the possibility of applying the transform to facies proportion modeling was explored. The following procedure was considered:

1. Transform facies proportions to logratios.
2. Kriging is performed to estimate the conditional logratio value at an unsampled location.
3. Back transform the estimated logratio value to obtain the estimated facies proportion at the unsampled location.

In Sequential Indicator Simulation with logratios transformed data, the procedures are the same as discussed above except for the following adjustments:

1. For each of the randomly visited grid node, perform kriging on the logratios, for each of the *K-1* proportions of interest. These are then back transformed to obtain the estimated facies proportions, which are subsequently used to construct the conditional cumulative distribution function (ccdf).
2. Perform Monte Carlo simulation (MCS) to draw a facies category and assign it to the grid node. Add this to the database and proceed to the next location.

Despite the above seemingly straightforward application, two critical issues remain in the proposed implementation: (1) the zero proportions, and (2) non-linearity.

Problems of Zero-proportion

As previously discussed, at a point scale, facies data can be resolved into discrete categories and the application of an indicator transform is straightforward. As we consider scaling this information to represent a block support, this categorical information transitions into a continuous property. We can further imagine that if we consider a sufficiently small block volume, the possibility exists that we will have only one or several (but not all) facies represented. This results in at least one facies whose proportion at that scale is zero.

The presence of a zero proportion facies is only a problem in the logratio transform if that particular facies is chosen as the one represented in the denominator. Recall that the transform is based on taking the ratio of p_k relative to p_q, where q is one of the K facies. A zero proportion for p_q leads to a non-existent logratio. Of course, one could argue that the choice of the q^{th} facies should be the one that does not yield zero proportions; however, in practice this may be unrealistic as this is dependent on the support volume of interest and it may be difficult to ensure that all locations will satisfy this constraint.

One simple solution is to apply some arbitrary small value, such as 10^{-5} or 10^{-20}, to substitute the zero facies proportions. This arbitrary choice can lead to quite different logratio values. However, a consistent output is required in our analysis and modeling. According to repeated tests via simulated examples, we find that the logratios are preserved after a Gaussian or uniform score transform and this may partially solve the problem in case we apply a second transformation for subsequent modeling. However, in many common facies modeling methodologies, such as SIS, neither a Gaussian nor uniform score transformation is used; therefore such an inconsistency problem is hard to avoid if logratios are adopted. The problem arises when averaging logratios with and without zeros; the back transformed values would be different.

Non-linearity Problem

A General View

Recall in multiscale facies modeling, facies categories are scaled up to facies proportions at various supports, and kriging is applied to estimate facies proportion at an unsampled location. Each of these processes (up-scaling and kriging) involves a linear process applied on our sample data; however, the logratio transformation is not a linear mapping of the data to an alternate set of variables. This non-linearity means that, in general,

$$a \log\left(\frac{p_k}{p_q}\right) + b \log\left(\frac{p_l}{p_q}\right) \neq \log\left(a \cdot \frac{p_k}{p_q} + b \cdot \frac{p_l}{p_q}\right)$$

The consequence of this non-linear mapping means that after we obtain the mean of logratio values, we can not simply do the reverse transform to obtain the mean of the ratio p_k/p_q. Similarly, when we obtain an estimated logratio value using kriging, we can not obtain the correct estimated value of the ratio p_k/p_q or the proportion p_k simply by a straightforward back transformation.

The nonlinearity problem can be illustrated through a small example. Here assume we have only two facies and take two samples from location $\mathbf{u}_1$ and $\mathbf{u}_2$ with proportion for facies 1 as p_{11} and p_{12} respectively. We plot the points of all the possible combinations of proportions p_{11}

and p_{12} which result at the same estimated proportion $\hat{\bar{p}}_1$ and get the percentile contour as in Figure 2 (solid curves). Comparing this figure with the dot lines in Figure 2, which represents the percentile contour of linear averaging, we can see clearly the nonlinearity of the logratio back transformed proportion $\hat{\bar{p}}_1$.

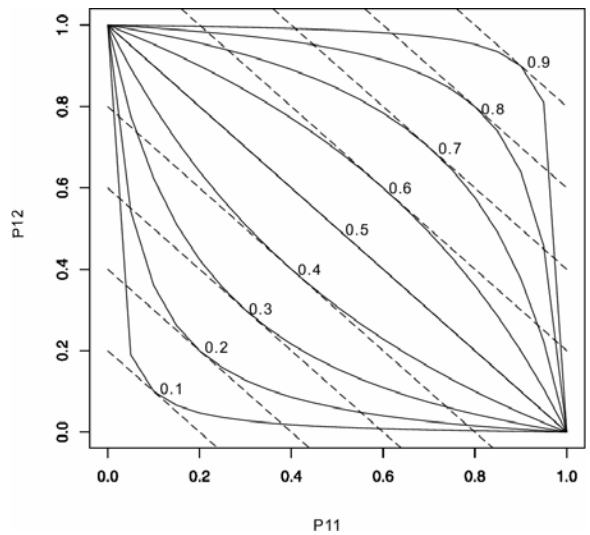

Figure 2: Percentile contour for estimated proportion of facies 1, $\hat{\bar{p}}_1$. Each of the solid curves is the combination of p_{11} and p_{12} that gives an identical logratio back transformed estimated facies proportion $\hat{\bar{p}}_1$. Each of the dotted lines is the combination of p_{11} and p_{12} that yields an identical arithmetic average of $\bar{p}_1$.

The differences (errors) between the estimated $\hat{\bar{p}}_1$ and the true $\bar{p}_1$ (that is: $\bar{p}_1 - \hat{\bar{p}}_1$) are obvious. Figure 3 gives the map of the errors vs p_{11} and p_{12} while Figure 4 gives the curves showing a series of the errors vs p_{11} for each $p_{12} \in \{0.05, 0.1, 0.15, ..., 0.95\}$. From these two figures, we can see that the difference is spreading symmetrically as p_{11} and p_{12} diverges away from 0.5 and the range is within $(-0.5, 0.5)$. By the way, in Figure 4, when p_{11} takes a value between approximately 0.1 to 0.9, the error is relatively small, but still goes up to ±0.20. Errors are significantly greater (±0.50) when p_{11} takes value less than 0.1 or greater than 0.9.

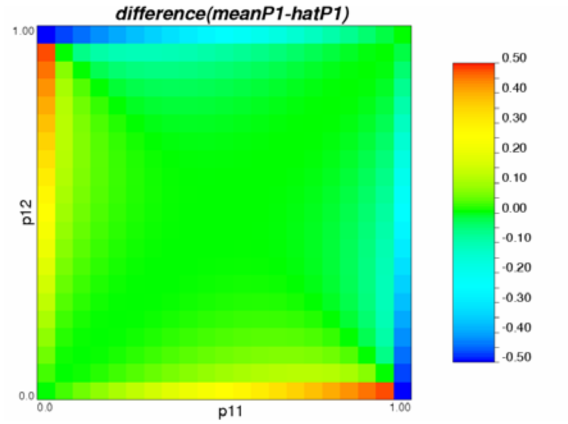

Figure 3: Map of difference between $\hat{\bar{p}}_1$ and $\bar{p}_1$.

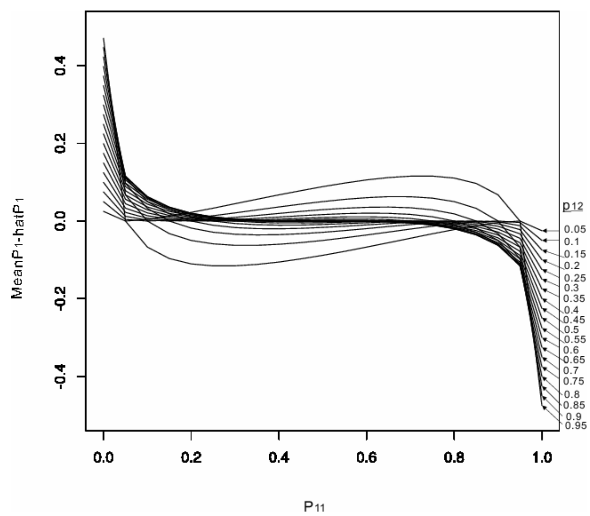

Figure 4: Difference between $\hat{\bar{p}}_1$ (hatP1) and $\bar{p}_1$ (meanP1). Each curve represents the difference $\bar{p}_1 - \hat{\bar{p}}_1$ vs p_{11} for a specific value of p_{12}.

3.2.2 Arithmetic Average versus Geometric Average

Recall the back transform of the logratio as shown in Part 2 above:

$$\hat{\bar{p}}_k = \frac{\exp(\bar{r}_k)}{1 + \sum_{l=1, l \neq q}^{K} \exp(\bar{r}_l)} \quad \text{for } k = 1, \ldots, q-1, q+1, \ldots, K$$

and

$$\hat{\bar{p}}_q = \frac{1}{1 + \sum_{l=1, l \neq q}^{K} \exp(\bar{r}_l)}$$

From this, we have

$$\overline{r}_k = \frac{1}{n}\sum_{\alpha=1}^{n} r_{k\alpha} = \frac{1}{n}\sum_{\alpha=1}^{n}\log\frac{p_{k\alpha}}{p_{q\alpha}} = \log\sqrt[n]{\prod_{\alpha=1}^{n}\frac{p_{k\alpha}}{p_{q\alpha}}} = \log\frac{\dot{p}_k}{\dot{p}_q} \quad \text{for } k = 1,\ldots,q-1,q+1,\ldots,K$$

And back transformed to finally yield

$$\hat{\overline{p}}_k = \frac{\dot{p}_k}{\sum_{l=1}^{K}\dot{p}_l} \quad \text{for } l = 1,\ldots,K$$

where $p_{k\alpha}$ is the proportion of the i^{th} facies at location $\mathbf{u}_\alpha$, $\alpha=1, \ldots, n$, and n is the number of data, $\dot{p}_k$'s the geometric average of $p_{k\alpha}$'s in the sample.

This process shows that, when we back-transform the arithmetic average of logratio of facies proportions, we obtain a standardized geometric average of the proportion, rather than the arithmetic average of the facies proportion. Applying logratios in either up-scaling or down-scaling process is therefore inappropriate.

In the example with two facies and two samples above, we have

$$\hat{\overline{p}}_k = \frac{\dot{p}_k}{\dot{p}_1 + \dot{p}_2}$$

where $k= 1, 2$. Given a fixed estimated value $\hat{\overline{p}}_1 = a$, we find

$$p_{12} = \frac{1-p_{11}}{1+\frac{1-2a}{a^2}\cdot p_{11}} = \frac{1-p_{11}}{1+\xi\cdot p_{11}} \text{ with } \xi = \tfrac{1-2a}{a^2}$$

Notice now that $\xi = 0$ if and only if $a=0.50$. That is to say, given a fixed estimated value a for the back transformed estimated logratio proportion, the relation of the two sample proportion values p_{11} and p_{12} for facies 1 are nonlinear unless $a=0.50$. This is what we see in Figure 2.

Such a problem also brings a strong effect on the kriging process. For example, if we estimate p_{i0} (proportion for i^{th} facies at position $\mathbf{u}_0$) using ordinary kriging, we get the estimate written as:

$$\hat{p}_{k0} = \lambda_1 p_{k1} + \ldots + \lambda_n p_{kn} \quad \text{with} \quad \sum_{\alpha=1}^{n}\lambda_\alpha = 1$$

But applying logratios, we have:

$$\hat{r}_{k0} = \lambda_1^* r_{k1} + \ldots + \lambda_n^* r_{kn} \quad \text{with} \quad \sum_{\alpha=1}^{n}\lambda_\alpha^* = 1$$

that is:

$$\hat{r}_{k0} = \lambda_1^* r_{k1} + ... + \lambda_n^* r_{kn} = \log \frac{\prod_{\alpha=1}^{n} (p_{k\alpha})^{\lambda_\alpha^*}}{\prod_{\alpha=1}^{n} (p_{q\alpha})^{\lambda_\alpha^*}}$$

The reverse transform is given as

$$\hat{p}_{k0}^* = \frac{\prod_{\alpha=1}^{n} (p_{k\alpha})^{\lambda_\alpha^*}}{\sum_{l=1}^{K} \prod_{\alpha=1}^{n} (p_{l\alpha})^{\lambda_\alpha^*}}$$

Clearly, $\hat{p}_{i0} \neq \hat{p}_{i0}^*$ and expected value

$$E[\hat{p}_{k0}^*] = E\left[\frac{\prod_{\alpha=1}^{n} (p_{k\alpha})^{\lambda_\alpha^*}}{\sum_{l=1}^{K} \prod_{\alpha=1}^{n} (p_{l\alpha})^{\lambda_\alpha^*}}\right] \neq E[\hat{p}_{k\alpha}] = E[\hat{p}_{k0}]$$

in general. By the way, the λ_α's and λ_α^*'s ($\alpha = 1, 2, ..., n$) in the above are estimated via two different linear regression models independently. It seems difficult to identify the relationship between λ_α's and λ_α^*'s which is necessary if we want to model and fix the difference (errors) between $\hat{p}_{i0}$ and $\hat{p}_{i0}^*$.

Conclusion

The benefit of applying the logratio transform to this problem of facies proportions that naturally describe the facies composition at various volume supports are twofold: (1) it explicitly honours the non-negative requirement for estimated facies proportions, and (2) it reproduces the constant sum constraint that is inherent to honouring the order relations. Despite these benefits, the use of a linear approach, such as kriging, to directly estimate and model the facies proportion yields a bias. Furthermore, back transforming the arithmetic average of logratio values leads to a standardized geometric average of facies proportion, which is completely different from the arithmetic average that is desired in the up-scaling process. Further, the zero proportion problem is hard to avoid when no normal score or uniform score transformation is adopted. These problems make it inappropriate to apply logratios in the multiscale facies modeling.

Some Future Works

It is concluded in this paper that logratio transformation is not an appropriate approach in multiscale facies modeling and we need to consider other approaches to cope with the problems regarding the negative proportion and sum constraints.

- Compositional kriging (CK) suggested by de Gruijter et al. (2001) is one choice. The essential idea is similar to ordinary kriging, but all the constraints are simultaneously accounted for via a corresponding Lagrange multiplier. The kriging weights are estimated by minimizing the kriging variance subject to all the constraints. A further study on the application of CK in multiscale facies modeling will be valuable in solving our problems.
- Another important alternative is posteriori processing. Here we set the kriging estimated facies probability $\hat{p}_{k\nu}^{IK}$ at the unsampled location $\mathbf{u}_\nu$ to zero if it is negative and then

reset the estimated facies probability $\hat{p}^*_{k\nu}$ according to the formula below:

$$\hat{p}^*_{k\nu} = \frac{\hat{p}^{IK}_{k\nu}}{\sum_{l=1}^{K} \hat{p}^{IK}_{l\nu}} \quad \text{where } k = 1, \ldots, K$$

After posteriori processing, we obtain the estimated facies probabilities which fully satisfy the order relation requirements. A relative adjustment is made on the SIS algorithm by adding the posteriori correction after indicator kriging and before Monte Carlo simulation step at each of the randomly visited grid nodes. Further work will be needed to solve various technical problems in applications of this posteriori correction.

References

Aitchison, J. (1986), *The Statistical Analysis of Compositional Data*, Chapman and Hall, New York.

Aitchison, J. (1981). *"A New approach to null correlations of proportions"*. Mathematical Geology, Vol. 13, p.175-189.

Aitchison, J. (1999). *"Logratio and Natural Laws in Compositional Data Analysis"*. Mathematical Geology, Vol. 31, No.5, p.563-579.

Deutsch, C.V. and Journel, A.G. (1998), *GSLIB Geostatistical Software Library and User's Guide*, Oxford University Press.

Deutsch, C.V. (2002), *Geostatistical Reservoir Modeling*, Oxford University Press.

Walvoort, D.J.J., and de Gruijter, J.J. (2001) *"Compositional Kriging: A Spatial Interpolation Method for Compositional"*, Mathematical Geology, Vol. 33, No. 8, p.951-966.