# Incorporating Secondary Data in the Prediction of Reservoir Properties Using Bayesian Updating

Chad Neufeld and Clayton V. Deutsch

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

## Abstract

*Bayesian updating is a robust method that can be used to incorporate secondary data in the estimation of reservoir properties. It's ability to capture non-linear relationships between primary and secondary data give it an advantage over other estimation methods. The required theory and equations are derived. Some implementation details are presented. A short, comparative case study shows the benefits of sequential Gaussian simulation with Bayesian updating. A new version of* `sgsim`*,* `sgsim_bu`*, has been developed to perform Bayesian updating.*

## Introduction

Most geological phenomenon exhibit complex behaviour. These complex behaviours are also present in the relationship between the reservoir properties that we are trying to predict and the secondary data that we have available. It is difficult to reproduce the complex non-linear relationship between data types in the estimation of reservoir properties with existing methods. Bayesian updating is developed as a way to incorporate secondary data in the estimation of reservoir properties with the ability reproduce these complex relationships between different data sources.

Collocated cokriging is used extensively to incorporate two- or three-dimensional secondary information in the prediction of a primary variable. The primary variable may be a petrophysical property of the reservoir, i.e. porosity, or it may be a geological control, i.e. the top surface of the reservoir. The secondary information is typically a seismic response or interpreted from seismic such as fraction of shale or impedance. A limitation of conventional collocated cokriging is the assumption that there is a linear relationship between the primary and secondary variable. Consider a case where the correlation is high in areas where the secondary variable is low and the correlation is low when the secondary variable is high. When the secondary data is low this limitation reduces the influence of the secondary variable and when the secondary variable is high the influence of the secondary data is increased. Bayesian updating overcomes this limitation.

A cross plot of the primary variable versus the secondary is used to calibrate the secondary data to the primary. Unlike collocated cokriging, which uses the secondary variable and a constant correlation, Bayesian updating uses the mean and variance of the primary variable given the secondary variable. This allows Bayesian updating to adjust the influence of the secondary data according to the amount of information that it contains.

The standard version of `sgsim` has been updated to allow Bayesian updating in addition to the standard estimation options. The new `sgsim_bu` allows simple kriging or collocated kriging to be used to calculate the prior distribution that will be updated using the secondary data. Currently only one secondary variable can be used to perform the Bayesian updating in `sgsim_bu`. With many possible sources of data available there may be several data types that could be included in the estimation of reservoir properties. Incorporating multiple secondary variables into the estimation process is relatively simple and straightforward [4]. The methodology developed in this paper deals with a single secondary variable, but could easily be expanded to $n$ variables.

### Notation

Let $Z_\alpha(\mathbf{u})$ be the primary data available in the field of interest where $\mathbf{u}$ is the location vector. $Z_\alpha(\mathbf{u})$ is the reservoir property being predicted available at the well locations. Let $X(\mathbf{u})$ be the secondary data available at all locations over the field of interest. $X(\mathbf{u})$ is the seismic variable available at all $\mathbf{u}$ used to aid in the prediction of $Z(\mathbf{u})$.

Let $y(\mathbf{u})$ be the estimate and $\sigma^2(\mathbf{u})$ be the variance of the estimate at the location $\mathbf{u}$. At each location three estimates of $Z(\mathbf{u})$ are considered: the likelihood $y_L(\mathbf{u})$ and $\sigma^2_L(\mathbf{u})$, the prior $y_P(\mathbf{u})$ and $\sigma^2_P(\mathbf{u})$, and the updated $y_U(\mathbf{u})$ and $\sigma^2_U(\mathbf{u})$. The likelihood distribution is calculated at each location from the bi-variate relationship between the secondary and primary data. Kriging is used to determine the prior distribution at each location. The Bayesian updating process combines the likelihood and prior distributions to get the updated distribution.

## Bayesian Updating Equations

The Bayesian updating methodology presented in this paper is expanded from Doyen's work [2]. Taken directly from his paper:

$$x_i^{SCC} = \frac{\rho z_i \sigma^2_{SK}(i) + x_i^{SK}[1-\rho^2]}{\rho^2[\sigma^2_{SK}(i) - 1] + 1} \quad (1)$$

$$\sigma^2_{SCC}(i) = \frac{\sigma^2_{SK}(i)[1-\rho^2]}{\rho^2[\sigma^2_{SK}(i) - 1] + 1} \quad (2)$$

where $x_i^{SCC}$ is the updated estimate, $\sigma^2_{SCC}$ is the updated variance, $x_i^{SK}$ and $\sigma^2_{SK}$ are the simple kriging estimate and variance using the primary data, $\rho$ is the correlation between the primary and secondary variables, and $z_i$ is the secondary variable. We want to rewrite Doyen's equations into our notation.

Table 1: Notation differences.

| | **Doyens Notation** | **New Notation** |
|---|---|---|
| Updated Mean | $x_i^{SCC}$ | $Y_U(\mathbf{u})$ |
| Updated Variance | $\sigma^2_{SCC}(i)$ | $\sigma^2_U(\mathbf{u})$ |
| Prior Mean | $x_i^{SK}$ | $Y_P^2(\mathbf{u})$ |
| Prior Variance | $\sigma^2_{SK}(i)$ | $\sigma^2_P(\mathbf{u})$ |
| Secondary Variable | $z_i$ | $X(\mathbf{u})$ |

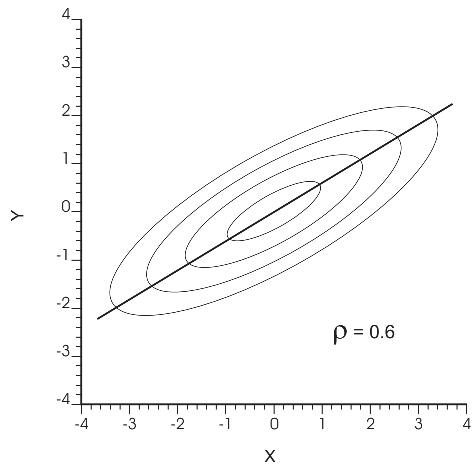


Figure 1: Bi-variate Gaussian relationship.

We will replace the $SK$ subscript with $P$ for the prior distribution, the $SCC$ with $U$ for the updated distribution, and $z_i$ with $X(\mathbf{u})$ for the secondary variable. Table 1 contains the different notations. Now rewrite equations (1) and (2) in the new notation:

$$Y_U(\mathbf{u}) = \frac{\rho X(\mathbf{u})\sigma^2_P(\mathbf{u}) + Y_P(\mathbf{u})[1-\rho^2]}{\rho^2[\sigma^2_P(\mathbf{u})-1]+1} \tag{3}$$

$$\sigma^2_U(\mathbf{u}) = \frac{\sigma^2_P(\mathbf{u})[1-\rho^2]}{\rho^2[\sigma^2_P(\mathbf{u})-1]+1} \tag{4}$$

Working in Gaussian space allows equations 3 and 4 to be further simplified. Figure 1 shows a bi-variate Gaussian relationship. The two equations from this relationship that we use are:

$$\mathrm{E}\{Y|X\} = Y_L$$

$$Y_L(\mathbf{u}) = \rho \cdot X(\mathbf{u}) \tag{5}$$

$$\mathrm{Var}\{Y|X\} = \sigma^2_L$$

$$\sigma^2_L(\mathbf{u}) = 1-\rho^2 \tag{6}$$

Now sub equations 5 and 6 into equations 3 and 4 to get the final equations 7 and 8:

$$Y_U(\mathbf{u}) = \frac{Y_L(\mathbf{u})\sigma^2_P(\mathbf{u}) + Y_P(\mathbf{u})\sigma^2_L(\mathbf{u})}{[1-\sigma^2_L(\mathbf{u})][\sigma^2_P(\mathbf{u})-1]+1} \tag{7}$$

$$\sigma_U^2(\mathbf{u}) = \frac{\sigma_P^2(\mathbf{u})\sigma_L^2(\mathbf{u})}{[1-\sigma_L^2(\mathbf{u})][\sigma_P^2(\mathbf{u})-1]+1} \tag{8}$$

where $P$ refers to the parameters of the prior distribution obtained from kriging, $L$ refers to the likelihood distribution obtained from the secondary data, and $U$ is the updated distribution.

# Methodology

Estimating reservoir properties using Bayesian updating requires four distinct steps: (1) calibration of the primary and secondary data to define the likelihood distribution, (2) calculation of the prior distribution using kriging, (3) updating the prior distribution with the likelihood distribution to get the updated distribution, and (4) performing the simulation using the posterior distribution. Steps 1 and 3 will be discussed here. For details on steps 2 and 4 see [1].

## Calculation of the Likelihood Distribution

The likelihood distribution defines the bi-variate relationship between the primary and secondary variables. Many of these relationships are too complex to be defined by a single parameter; i.e. the correlation coefficient. Bayesian updating uses the mean and variance of the primary variable given a specific value of the secondary variable as the likelihood distribution. Complex heteroscedastic and non-linear relationships can be captured using this approach.

The mean and variance of the primary variable given the secondary variable are required to define the likelihood distribution. Splitting the primary variable into classes, or bins, according to the secondary variable is the first step. Calculating the mean and variance for each class is the second step. This set of means and variances defines the likelihood distribution. In some cases the bin-by-bin mean and variance are quite noisy. A moving window, inverse distance, or kriging type smoothing could be applied to the two calibration curves,

The mean and variance of the primary variable have to be transferred to the estimation grid for input to `sgsim_bu`. A simple program, `build_lh`, is included for this task. It takes the two calibration curves and populates the estimate grid given the value of the secondary variable at each node in the grid.

**Likelihood Consistency Check** It is a good idea to check the likelihood distribution for consistency. The likelihood distribution has a mean and variance for all values of the secondary variable. We can calculate the probability of the mean value given its associated variance. This probability can be used as a measure consistency.

The probability is calculated by determining the cumulative probability up to the likelihood mean using a normal curve with a mean of zero and a variance equal to one minus the likelihood variance. Extremely small or large probabilities may cause problems. These mean-variance pairs could be corrected. No correction scheme has been proposed yet. A simple plot to assess the consistency of the likelihood is shown in Figure 12 in the case study.

### Calculation of the Prior Distribution

Any form of kriging may be used to estimate the prior distribution at the location **u**. The updated version of `sgsim`, `sgsim_bu`, allows simple or collocated kriging to be used to calculate the prior distribution. In most cases simple kriging will be sufficient, but using collocated cokriging allows the integration of an additional data type.

### Bayesian Updating of the Likelihood and Prior Distributions

The Bayesian updating is performed after the prior distribution has been calculated in `sgsim_bu`. Since both the likelihood and prior distributions are Gaussian the resulting updated distribution is also Gaussian [4]. Recall that the mean for the updated distribution is:

$$Y_U(\mathbf{u}) = \frac{Y_L(\mathbf{u})\sigma_P^2(\mathbf{u}) + Y_P(\mathbf{u})\sigma_L^2(\mathbf{u})}{[1-\sigma_L^2(\mathbf{u})][\sigma_P^2(\mathbf{u})-1]+1}$$

and the variance for the updated distribution is:

$$\sigma_U^2(\mathbf{u}) = \frac{\sigma_P^2(\mathbf{u})\sigma_L^2(\mathbf{u})}{[1-\sigma_L^2(\mathbf{u})][\sigma_P^2(\mathbf{u})-1]+1}$$

where $y$ and $\sigma^2$ are the mean and standard deviation for a particular distribution, and $P$, $L$, and $U$ are the prior, likelihood, and updated posterior distributions respectively, The simulation is then performed using the updated distribution at each location.

## Case Study

This case study compares 3 simulation methods for modelling porosity in a 2-D reservoir. The methods used were sequential Gaussian simulation (sGs), sGs with collocated cokriging, and sGs with Bayesian updating. The well data is complemented by seismic data. The data set contains (1) the porosity at the well locations, and (2) the fraction of clay interpreted from seismic.

Figure 2 shows the well porosity data and Figure 3 shows the histogram of the porosity data. The fraction of clay from seismic is shown in Figure 4 and Figure 5 shows the fraction of clay histogram.

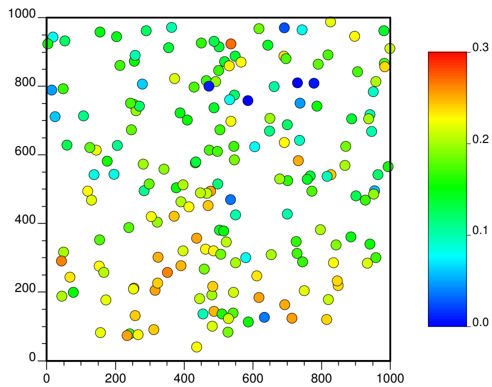


Figure 2: Porosity location map.

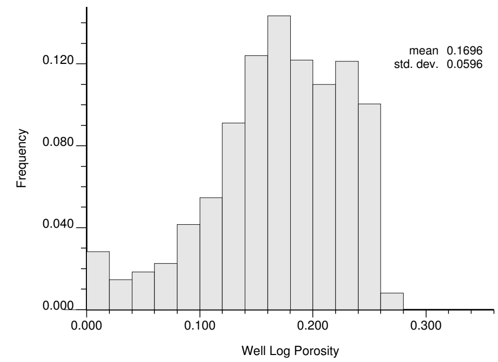


Figure 3: Porosity histogram.

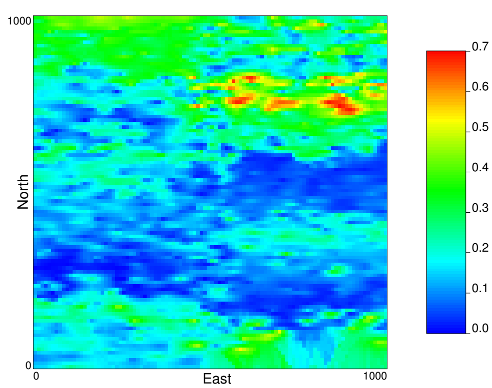


Figure 4: Fraction of clay from seismic.

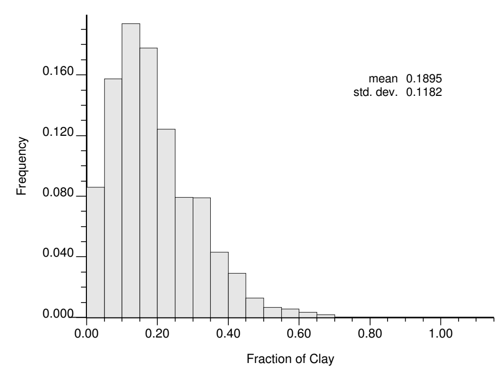


Figure 5: Fraction of clay histogram.

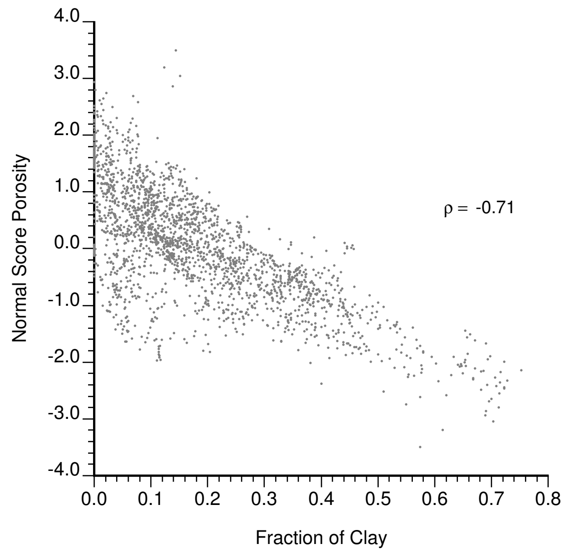


Figure 6: Scatterplot of normal score porosity versus fraction of clay.

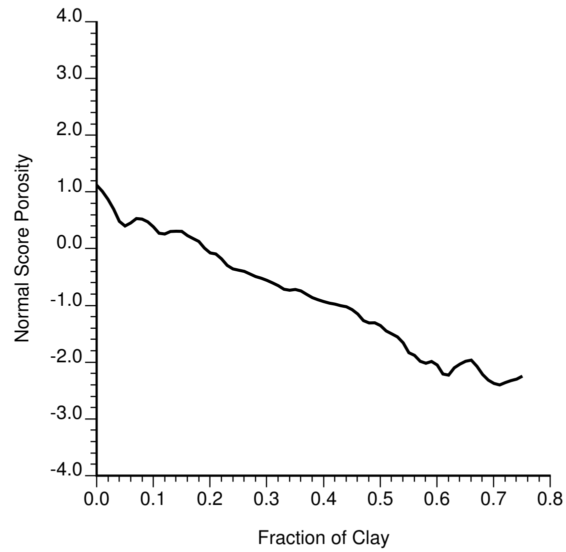


Figure 7: Smoothed porosity mean curve.

## Likelihood Distribution Calculation

First the porosity data were declustered and normal score transformed. The scatterplot of normal score porosity versus fraction of clay is shown in Figure 6. The correlation is good at -0.71 and there is an obvious decrease in the porosity variability as the fraction of clay increases.

The next step was to bin the data and calculate the mean and variance for each bin. The resulting mean and variance curves were very noisy, so an inverse distance smoothing of the curves was done. The smoothed mean and variance curves are shown in Figures 7 and 8. An interesting plot can be made by placing the smoothed mean curve and the $\pm 1$ standard deviation curves on top of the scatterplot. This is shown in Figure 9.

The mean and variance were then transferred to the simulation grid using `build_lh`. This program takes the mean and variance curves and populates the simulation grid based on the fraction of clay value at each grid node. These two grids are required as input to `sgsim_bu`. The mean and variance maps are shown in Figures 10 and 11.

The likelihood distribution was checked for consistency before running the simulations. This check is done for each point along then mean and variance curves. Figure 12 shows the consistency plot for the likelihood distribution. The points that lie at the extreme end of the probability distribution may cause problems during the simulation. The problems could be inflated variance or a shifted mean. No correction will be done.

It is interesting to note that bi-variate relationship is not simple. The decrease in porosity in relation to the fraction of clay is almost linear, but the variance is not constant or linear. Our goal is to reproduce the porosity versus fraction of clay relationship in the simulation.

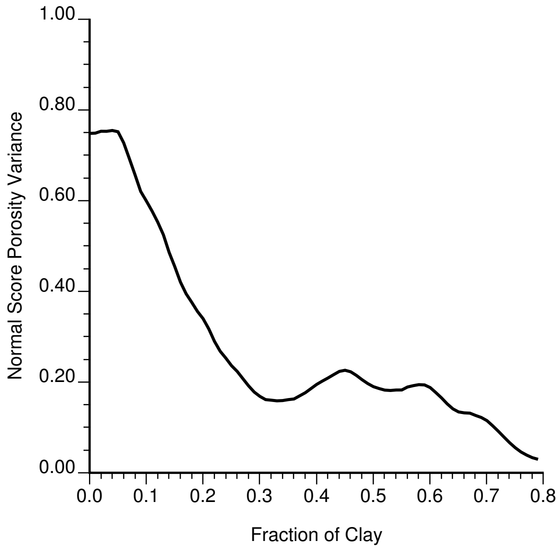


Figure 8: Smoothed porosity variance curve.

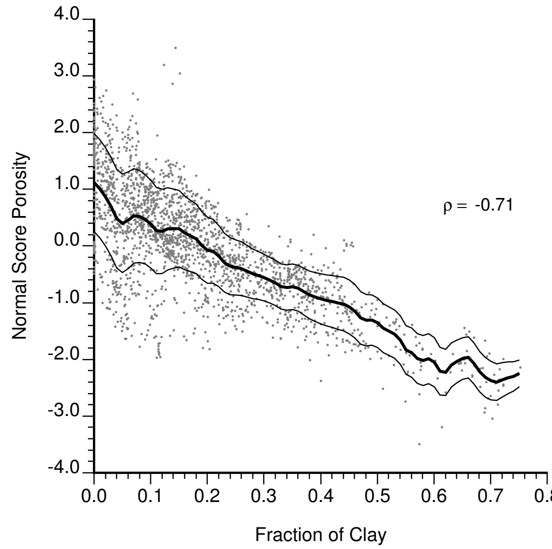


Figure 9: Scatterplot of normal score porosity versus fraction of clay with the mean porosity curve.

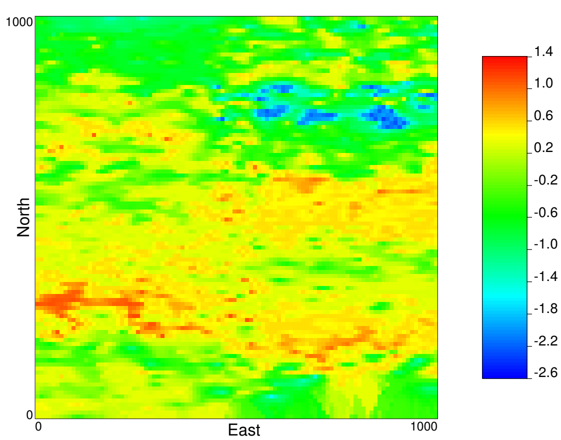


Figure 10: Gridded likelihood mean for Bayesian updating.

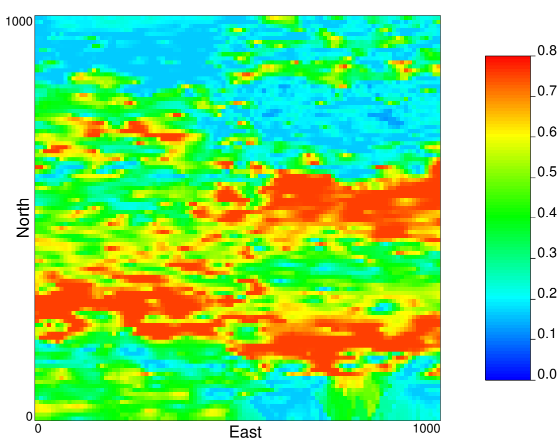


Figure 11: Gridded likelihood variance for Bayesian updating.

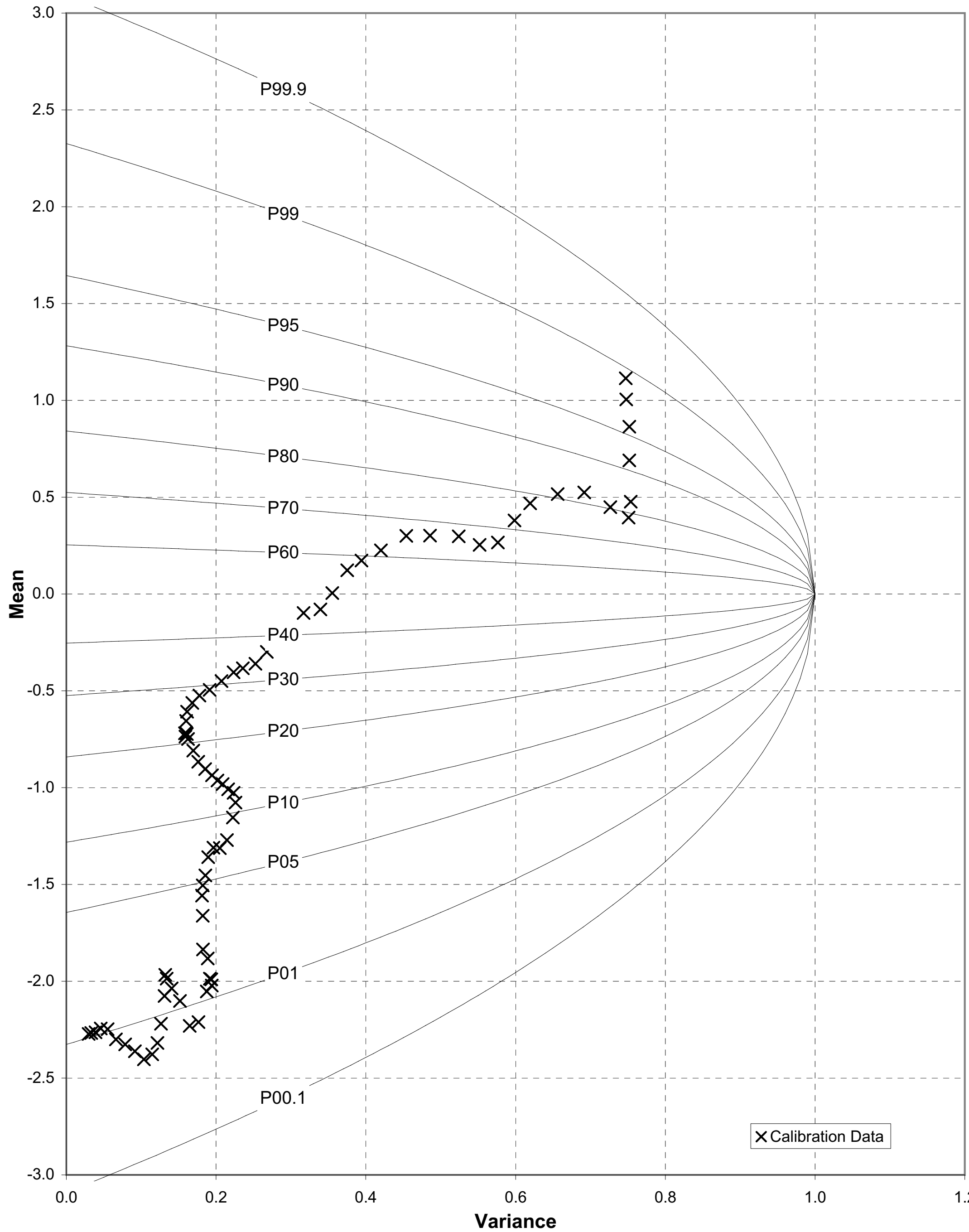


Figure 12: Consistency plot for the likelihood distribution.

### Simulation Comparison

An identical setup was used for the three simulations. The sGs used the porosity at the well locations. The sGs with collocated cokriging used the fraction of clay as the secondary variable with a correlation of -0.7. The sGs with Bayesian updating used the likelihood mean and variance as the secondary data. Several realizations were run for each case to compare.

Figure 13 shows the results of the 2nd realization for the three simulation methods. Sequential Gaussian simulation with Bayesian updating reproduces the bi-variate relationship between porosity and the fraction of clay the best.

## Implementation Details

The Bayesian updating method has not been without its growing pains. To date two issues have been identified. The first is variance inflation and the second is the mean shifting from zero. We do not propose a solution for the mean implementation issue here.

The variance inflation is very similar to the variance inflation that occurs with collocated cokriging. The approach used for collocated cokriging is to use a variance reduction factor. We take a similar approach here. A variance reduction factor can be specified in the parameter file. This reduction factor is applied to the variance after the Bayesian updating has been done and before the random number for the simulation is drawn. This has had good results.

## Future Work

The initial applications of sGs using Bayesian updating have given good results. However, there are many aspects of the Bayesian updating technique that need to be further explored and documented. These included, but are not limited to: (1) variance inflation, (2) reproduction of the input histogram, (3) variogram reproduction, and (4) honouring the non-linear bi-variate relationship between the primary and secondary data.

## Conclusions

Bayesian updating is a robust method that can be used to incorporate secondary data in the estimation of reservoir properties. The ability to capture non-linear relationships gives it an advantage over collocated cokriging. There have been several application of this idea to date. All of the applications have given good results, but there have been some implementation issues. Additional work is required to document and fix the implementation issues.

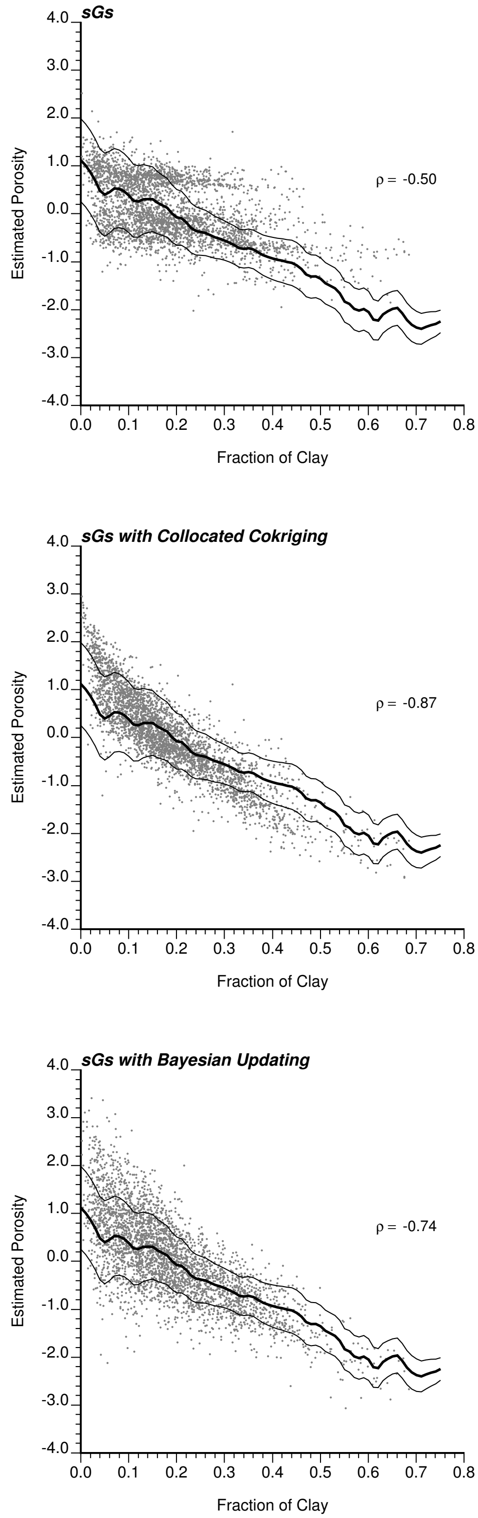


Figure 13: Results from the three simulation methods. sGs with Bayesian updating reproduces the bi-variate relationship the best.

## References

[1] C. V. Deutsch and A. G. Journel. *GSLIB Geostatistical Software Library and User's Guide*. Oxford University Press, New York, second edition, 1998.

[2] P. M. Doyen, L. D. den Boer, and W. R. Pillet. Seismic porosity mapping in the Ekofisk field using a new form of collocated cokriging. *Society of Petroleum Engineers*, 1996. SPE 36498.

[3] W. Xu, T. T. Tran, R. M. Srivastava, and A. G. Journel. Integrating seismic data in reservoir modeling: The collocated cokriging alternative. *Society of Petroleum Engineers*, 1992. SPE 24742.

[4] S. D. Zanon and C. V. Deutsch. Direct prediction of reservoir performance with Bayesian updating under a multivariate Gaussian model. In *5th Canadian International Petroleum Conference*, Calgary, Alberta, Canada, June 2004. Petroleum Society of the Canadian Institute of Mining, Metallurgy & Petroleum.