Large Scale Modeling by Bayesian Updating Techniques

Weishan Ren

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

Large scale models are useful for resource estimation. Building large scale models is an important step in the scale consistent modelling approach. Bayesian updating technique is recommended because of reliability and simplicity of integrating multiple types of data. The theory of Bayesian updating technique is reviewed in this paper. The CCG format of Bayesian updating is shown to be the same as other publications. Simulation with Bayesian updating can be performed sequentially or using a P-field approach. If the local uncertainties are already built in large scale modeling, the P-field simulation with Bayesian updating is recommended because of computational efficiency and the consistency between the local uncertainties and the simulation realizations.

Introduction

Large scale models are required for modeling large areas such as an entire lease or reservoir. Building fine scale models for such large areas is neither practical nor necessary. For example, 100 realizations of one variable for the Surmont lease at a scale of 20 m by 1 m by 1m would require over 1000 GB storage. This is a significant amount of storage and it takes time to process and understand the resulting models. A 2-D model for the Surmont lease at a scale of 100m by 100m only takes a few MB.

Large scale models are useful for resource estimation at the early phases of field development. Resource estimation focuses on volume averages rather than detailed heterogeneity. Fine scale 3-D models of heterogeneity are useful for flow simulation but not necessary for resource estimation. Reliable large scale models are appropriate for resource estimation and selecting areas of interested. Fine scale 3-D models can be constructed in these areas when they are needed.

The advantage of modeling at a large scale is that smaller scale data can be upscaled to the model scale so that the multiscale modeling is converted to single scale modeling. Gaussian-based techniques can be used without concern for non-linear averaging. Converting data to 2-D summaries further simplifies multiscale modeling. 2-D mapping is the most common approach to large scale modeling.

For the scale consistent modeling (Ren *et al.*, 2007), an important step is to construct a large scale model as accurate as possible by integrating all available data. There are several geostatistical techniques that can be used to integrate different data into a geological model including Gaussian-based Bayesian updating, indicator cokriging, and full cokriging. The Bayesian updating approach will be presented in detail due to its reliability and simplicity in data integration. The contribution of the primary and secondary information on the updated results can be easily understood. In the approach, well data including core and well log data are considered as primary data that we have the most confidence. The seismic data, geologic trends, structural information, geological interpretations and other indirectly measured data are used as secondary data that are normally extensively available.

Bayesian updating technique was originally introduced by Doyen *et al.* (1996). The technique decomposes the collocated cokriging estimate into two models: prior and likelihood. The prior model is built from well data, and the likelihood model is built from secondary information (the definition could be reversed with a different interpretation). The prior model is then updated with the likelihood model to build the final posterior or updated model. Deutsch and Zanon (2004) applied the Bayesian updating technique to predict reservoir performance. They showed the approach in a new format that has the advantage of easy implementation in mapping multiple primary variables using multiple secondary variables. It has been successfully applied in the McMurray formation (Ren *et al.*, 2006a and Ren *et al.*, 2006c).

Bayesian updating uses all related information to generate local distributions of uncertainty. Under a multivariate Gaussian model, local uncertainty in the prior estimates is given by the kriging variance that accounts for the closeness and redundancy of the well data. Sparse well data lead to significant uncertainty in the interwell regions. Integrating secondary information such as seismic and geological data can reduce the uncertainty and improve the 2-D modeling. For large areas or areas have complex geology, the modeling could be improved by accounting for non-stationarity in correlations and data precision (Ren *et al.*, 2006b).

This paper focuses on the theory of Bayesian updating, presents some interesting aspects of the technique and the methodology of simulation with Bayesian updating.

Theory of Bayesian Updating

Consider a random function Y that is stationary over the area of interest, *A*. It is the primary variable of interest. There are *m* random functions X_j, *j = 1…m* over the same model area. They are the secondary variables. Assume Y and X_j (*j = 1…m*) are jointly multi-Gaussian after univariate transformation.

Suppose there are *n* data of the primary variable available in the area of interest: $\{y(\mathbf{u}_i), i = 1 \ldots n\}$, where $\mathbf{u}_i$ is the location vector in *A*. In the context of this thesis, the results of simple kriging using only the primary data are considered as a ***prior*** distribution of uncertainty parameterized by the kriging mean and variance. The kriging mean is calculated by:

$$\overline{y}_P(\mathbf{u}) = \sum_{i=1}^{n} \lambda_i y(\mathbf{u}_i) \tag{1}$$

where $\mathbf{u}$ is the location being estimated and the weights λ_i, (*i =1…,n*) are calculated from the well known normal equations:

$$\sum_{i=1}^{n} \lambda_i C(\mathbf{u}_i - \mathbf{u}_k) = C(\mathbf{u} - \mathbf{u}_k), \quad k = 1, \ldots n \tag{2}$$

where $C(\mathbf{u}_i - \mathbf{u}_k)$ is the covariance between the two primary data $y(\mathbf{u}_i)$ and $y(\mathbf{u}_k)$, and $C(\mathbf{u} - \mathbf{u}_k)$ is the covariance between estimated location $y(\mathbf{u})$ and primary data $y(\mathbf{u}_k)$. The kriging variance is given by

$$\sigma_P^2(\mathbf{u}) = 1 - \sum_{i=1}^{n} \lambda_i C(\mathbf{u} - \mathbf{u}_i) \tag{3}$$

The simple kriging leads to the parameters of a Gaussian conditional distribution conditioning to the primary data. A conditional distribution is predicted at each unsampled location.

In general, secondary data are available at every location in the modelled area: $\{x_j(\mathbf{u}), \quad j = 1, \ldots m, \quad \forall\, \mathbf{u} \in A\}$. The results of prediction with all collocated secondary data provide another conditional distribution. This distribution is related to the likelihood distribution in a Bayesian context. It has been called the ***likelihood*** distribution; this thesis retains that name. The non-standard Gaussian likelihood distribution is fully defined by a mean and variance.

The likelihood mean is calculated by:

$$\overline{y}_L(\mathbf{u}) = \sum_{j=1}^{m} \lambda_j x_j(\mathbf{u}) \tag{4}$$

Here, the weights λ_j (*j=1, … n*) are also given by the well-known normal equations:

$$\sum_{j=1}^{m} \lambda_j \rho_{j,k} = \rho_{j,0}, \quad k = 1,...m \tag{5}$$

where $\rho_{j,k}$ is the correlation between different types of secondary data, and $\rho_{j,0}$ is the correlation between the secondary data and primary data. The likelihood variance is then given by:

$$\sigma_L^2(\mathbf{u}) = 1 - \sum_{j=1}^{m} \lambda_j \rho_{j,0} \tag{6}$$

The two conditional distributions (prior and likelihood) can be merged together to get the updated distribution. The mathematic combination is derived from the Bayesian statistical analysis of the posterior distribution. The posterior distribution of uncertainty at estimated location **u** is a conditional distribution conditioning to both primary and secondary data:

$$P\{y(\mathbf{u}) \mid X(\mathbf{u}), y(\mathbf{u}_1), ...y(\mathbf{u}_n)\} \tag{7}$$

where $X(\mathbf{u}) = \begin{pmatrix} x_1(\mathbf{u}) & x_2(\mathbf{u}) & \cdots & x_m(\mathbf{u}) \end{pmatrix}^T$ is a vector of the collocated secondary data. Under the assumption that collocated secondary data screen the influence of other secondary data that are further away (Journel, 1999), only collocated secondary data are considered. This distribution is equivalent to the results of collocated cokriging.

In the context of Bayesian statistical analysis, the posterior distribution can be decomposed into a product of two distributions (Doyen *et al.*, 1996; Besag, 1986):

$$P\{y(\mathbf{u}) \mid X(\mathbf{u}), y(\mathbf{u}_1), ...y(\mathbf{u}_n)\} \propto f\{X(\mathbf{u}) \mid y(\mathbf{u})\} P\{y(\mathbf{u}) \mid y(\mathbf{u}_1), ...y(\mathbf{u}_n)\} \tag{8}$$

where $P\{y(\mathbf{u}) \mid y(\mathbf{u}_1), ...y(\mathbf{u}_n)\}$ is the prior distribution only conditioning to primary data. As shown before, the prior distribution is a non-standard Gaussian distribution with simple kriging mean $\bar{y}_P(\mathbf{u})$ and variance $\sigma_P^2(\mathbf{u})$. We have

$$P\{y(\mathbf{u}) \mid y(\mathbf{u}_1), ...y(\mathbf{u}_n)\} \propto \exp\left(-\frac{\left[y(\mathbf{u}) - \bar{y}_P(\mathbf{u}) \right]^2}{2\sigma_P^2(\mathbf{u})} \right)$$

$$\text{Where } \exp\left(-\frac{\left[y(\mathbf{u}) - \bar{y}_P(\mathbf{u}) \right]^2}{2\sigma_P^2(\mathbf{u})} \right) = \exp\left(-\frac{y^2(\mathbf{u})}{2\sigma_P^2(\mathbf{u})} + \frac{y(\mathbf{u})\bar{y}_P(\mathbf{u})}{\sigma_P^2(\mathbf{u})} - \frac{\bar{y}_P{}^2(\mathbf{u})}{2\sigma_P^2(\mathbf{u})} \right) \tag{9}$$

$$\text{Thus, } P\{y(\mathbf{u}) \mid y(\mathbf{u}_1), ...y(\mathbf{u}_n)\} \propto \exp\left(-\frac{y^2(\mathbf{u})}{2\sigma_P^2(\mathbf{u})} + \frac{y(\mathbf{u})\bar{y}_P(\mathbf{u})}{\sigma_P^2(\mathbf{u})} \right)$$

The proportionality constants that are independent of $y(\mathbf{u})$ are eliminated.

The $f\{X(\mathbf{u}) \mid y(\mathbf{u})\}$ in Equation (8) is the likelihood function that is simplified by only conditioning to the collocated primary data under the assumption of a Markov model that the collocated primary data screen the influence of other primary data that are further away (Journel, 1999). Under the assumption of a multivariate Gaussian model, it is the density function of a multivariate Gaussian distribution parameterized by the conditional mean vector $E\{X(\mathbf{u}) \mid y(\mathbf{u})\} = \boldsymbol{\rho}\, y(\mathbf{u})$ and m x m conditional covariance matrix Σ, where $\boldsymbol{\rho} = \begin{pmatrix} \rho_{10} & \rho_{20} & \cdots & \rho_{n0} \end{pmatrix}^T$ is the vector of correlation coefficients between the primary variable

and secondary data. The location **u** is dropped from the notation for simplicity. The covariance matrix can be converted to the matrix of correlation coefficients:

$$\Sigma = \begin{pmatrix} \rho_{11} & \cdots & \rho_{1n} \\ \vdots & \ddots & \vdots \\ \rho_{n1} & \cdots & \rho_{nn} \end{pmatrix} - \begin{pmatrix} \rho_{10}\rho_{10} & \cdots & \rho_{10}\rho_{n0} \\ \vdots & \ddots & \vdots \\ \rho_{n0}\rho_{10} & \cdots & \rho_{n0}\rho_{n0} \end{pmatrix} = \boldsymbol{\rho}_{ij} - \boldsymbol{\rho}\,\boldsymbol{\rho}^t$$

Since the secondary data are known, the likelihood function is actually a function of $y(\mathbf{u})$. Eliminating the proportionality constants that are independent of $y(\mathbf{u})$, we have:

$$f\{X(\mathbf{u}) \mid y(\mathbf{u})\} \propto \exp\left(-\frac{1}{2}[X(\mathbf{u}) - \boldsymbol{\rho}\, y(\mathbf{u})]^t \Sigma^{-1}[X(\mathbf{u}) - \boldsymbol{\rho}\, y(\mathbf{u})]\right)$$

$$\exp\left(-\frac{1}{2}[X(\mathbf{u}) - \boldsymbol{\rho}\, y(\mathbf{u})]^t \Sigma^{-1}[X(\mathbf{u}) - \boldsymbol{\rho}\, y(\mathbf{u})]\right)$$

where $= \exp\left(-\frac{1}{2}[X^t(\mathbf{u}) - \boldsymbol{\rho}^t y(\mathbf{u})]\Sigma^{-1}[X(\mathbf{u}) - \boldsymbol{\rho}\, y(\mathbf{u})]\right)$ (10)

$$= \exp\left(-\frac{1}{2}[X^t(\mathbf{u})\Sigma^{-1}X(\mathbf{u}) - 2\boldsymbol{\rho}^t\Sigma^{-1}X(\mathbf{u})y(\mathbf{u}) + \boldsymbol{\rho}^t\Sigma^{-1}\boldsymbol{\rho}\, y^2(\mathbf{u})]\right)$$

Thus, $f\{X(\mathbf{u}) \mid y(\mathbf{u})\} \propto \exp\left(\boldsymbol{\rho}^t\Sigma^{-1}X(\mathbf{u})y(\mathbf{u}) - \frac{1}{2}\boldsymbol{\rho}^t\Sigma^{-1}\boldsymbol{\rho}\, y^2(\mathbf{u})]\right)$

Multiplying the Equations (9) and (10) gives the posterior distribution:

$$P\{y(\mathbf{u}) \mid X(\mathbf{u}), y(\mathbf{u}_1), \ldots y(\mathbf{u}_n)\} \propto \exp\left(-\frac{1}{2}[\boldsymbol{\rho}^t\Sigma^{-1}\boldsymbol{\rho} + \frac{1}{\sigma_P^2(\mathbf{u})}]y^2(\mathbf{u}) + [X^t(\mathbf{u})\Sigma^{-1}\boldsymbol{\rho} + \frac{\bar{y}_P(\mathbf{u})}{\sigma_P^2(\mathbf{u})}]y(\mathbf{u})\right)$$

This equation is in the form of $\exp\left(-Ax^2 + Bx\right)$, where the constants A and B define the mean and variance of a Gaussian kernel:

$$\bar{y}_U(\mathbf{u}) = \frac{B}{2A} = \frac{\boldsymbol{\rho}^t\Sigma^{-1}X(\mathbf{u}) + \frac{\bar{y}_P(\mathbf{u})}{\sigma_P^2(\mathbf{u})}}{\boldsymbol{\rho}^t\Sigma^{-1}\boldsymbol{\rho} + \frac{1}{\sigma_P^2(\mathbf{u})}} = \frac{\frac{\boldsymbol{\rho}^t\boldsymbol{\rho}_{ij}{}^{-1}X(\mathbf{u})}{(\boldsymbol{\rho}_{ij} - \boldsymbol{\rho}\,\boldsymbol{\rho}^t)\boldsymbol{\rho}_{ij}{}^{-1}} + \frac{\bar{y}_P(\mathbf{u})}{\sigma_P^2(\mathbf{u})}}{\frac{\boldsymbol{\rho}^t\boldsymbol{\rho}_{ij}{}^{-1}\boldsymbol{\rho}}{(\boldsymbol{\rho}_{ij} - \boldsymbol{\rho}\,\boldsymbol{\rho}^t)\boldsymbol{\rho}_{ij}{}^{-1}} + \frac{1}{\sigma_P^2(\mathbf{u})}} \quad (11)$$

$$= \frac{\boldsymbol{\rho}^t\boldsymbol{\rho}_{ij}{}^{-1}X(\mathbf{u})\sigma_P^2(\mathbf{u}) + \bar{y}_P(\mathbf{u})(1 - \boldsymbol{\rho}\,\boldsymbol{\rho}_{ij}{}^{-1}\boldsymbol{\rho}^t)}{\boldsymbol{\rho}^t\boldsymbol{\rho}_{ij}{}^{-1}\boldsymbol{\rho}\;\sigma_P^2(\mathbf{u}) + (1 - \boldsymbol{\rho}\,\boldsymbol{\rho}_{ij}{}^{-1}\boldsymbol{\rho}^t)}$$

From Equation (5), the vector of weights can be expressed as: $\boldsymbol{\lambda} = \boldsymbol{\rho}^t\boldsymbol{\rho}_{ij}{}^{-1}$. Thus, the likelihood mean and variance are: $\bar{y}_L(\mathbf{u}) = \boldsymbol{\rho}^t\boldsymbol{\rho}_{ij}{}^{-1}X(\mathbf{u})$ and $\sigma_L^2(\mathbf{u}) = 1 - \boldsymbol{\rho}^t\boldsymbol{\rho}_{ij}{}^{-1}\boldsymbol{\rho}$. Then, the updated mean is

$$\bar{y}_U(\mathbf{u}) = \frac{\boldsymbol{\rho}^t\boldsymbol{\rho}_{ij}{}^{-1}X(\mathbf{u})\sigma_P^2(\mathbf{u}) + \bar{y}_P(\mathbf{u})(1 - \boldsymbol{\rho}\,\boldsymbol{\rho}_{ij}{}^{-1}\boldsymbol{\rho}^t)}{\boldsymbol{\rho}^t\boldsymbol{\rho}_{ij}{}^{-1}\boldsymbol{\rho}\;\sigma_P^2(\mathbf{u}) + (1 - \boldsymbol{\rho}\,\boldsymbol{\rho}_{ij}{}^{-1}\boldsymbol{\rho}^t)} = \frac{\bar{y}_L(\mathbf{u})\sigma_P^2(\mathbf{u}) + \bar{y}_P(\mathbf{u})\sigma_L^2(\mathbf{u})}{\sigma_P^2(\mathbf{u}) - \sigma_P^2(\mathbf{u})\sigma_L^2(\mathbf{u}) + \sigma_L^2(\mathbf{u})} \quad (12)$$

And the updated variance is:

$$\sigma_U^2(\mathbf{u}) = \frac{1}{2A} = \frac{1}{\boldsymbol{\rho}^t \Sigma^{-1} \boldsymbol{\rho} + \frac{1}{\sigma_P^2(\mathbf{u})}} = \frac{(1 - \boldsymbol{\rho}\, \boldsymbol{\rho}_{ij}{}^{-1} \boldsymbol{\rho}^t)\sigma_P^2(\mathbf{u})}{\boldsymbol{\rho}^t \boldsymbol{\rho}_{ij}{}^{-1} \boldsymbol{\rho}\, \sigma_P^2(\mathbf{u}) + (1 - \boldsymbol{\rho}\, \boldsymbol{\rho}_{ij}{}^{-1} \boldsymbol{\rho}^t)}$$
$$= \frac{\sigma_L^2(\mathbf{u})\sigma_P^2(\mathbf{u})}{\sigma_P^2(\mathbf{u}) - \sigma_P^2(\mathbf{u})\sigma_L^2(\mathbf{u}) + \sigma_L^2(\mathbf{u})} \tag{13}$$

These results give the parameters of a posterior non-standard Gaussian distribution called the ***updated*** distribution of uncertainty.

A schematic illustration of the Bayesian updating technique is given in Figure 1.

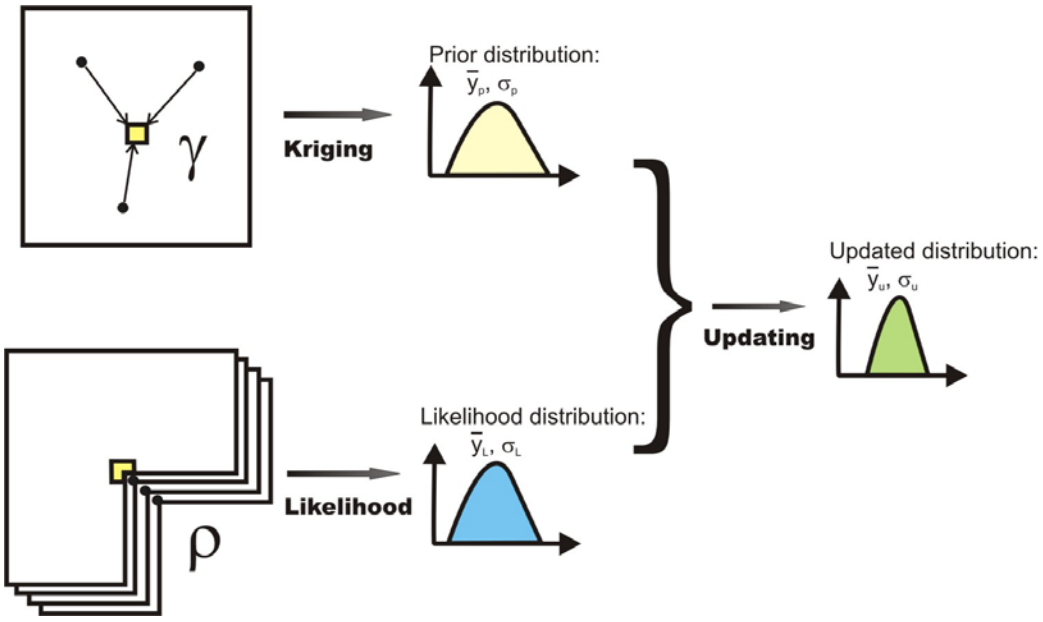

Figure 1: Schematic illustration of the Bayesian updating technique. The block dots are data, and the yellow square is a location being estimated.

Some Interesting Aspects of Bayesian Updating

The updated distribution obtained by merging the prior and likelihood distributions is non-convex because it does not always fall between the two distributions. **Error! Reference source not found.** shows the updated mean changes with increasing of prior means. The global distribution, a standard Gaussian distribution has an effect on whether the updated distribution falls between the two distributions. If the prior and likelihood distributions are on each side of the global distribution, the updated distribution will always fall in between. If the two distributions are both above the global distribution or both below the global distribution, the updated mean fall in between only when the means of the prior and likelihood distributions are far away enough (see cases 6 and 9 in **Error! Reference source not found.**). When the two distributions are very close, the updated mean will not fall in between (see cases 7 and 8 in **Error! Reference source not found.**). The updated distribution tends to be close to the distribution further away from the global mean (

Figure **2**).

Table 1: The Bayesian updating calculation showing the changes of updated mean (4th row) with increasing prior mean (second row). The variances of prior and likelihood are 0.4 for all cases.

Case	1	2	3	4	5	6	7	8	9
Prior Mean $\bar{y}_P$	-2.00	-1.50	-1.00	-0.50	0.00	0.50	1.00	1.50	2.00
Likelihood Mean $\bar{y}_L$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Updated Mean $\bar{y}_U$	-0.63	-0.31	0.00	0.31	0.63	0.94	1.25	1.56	1.88

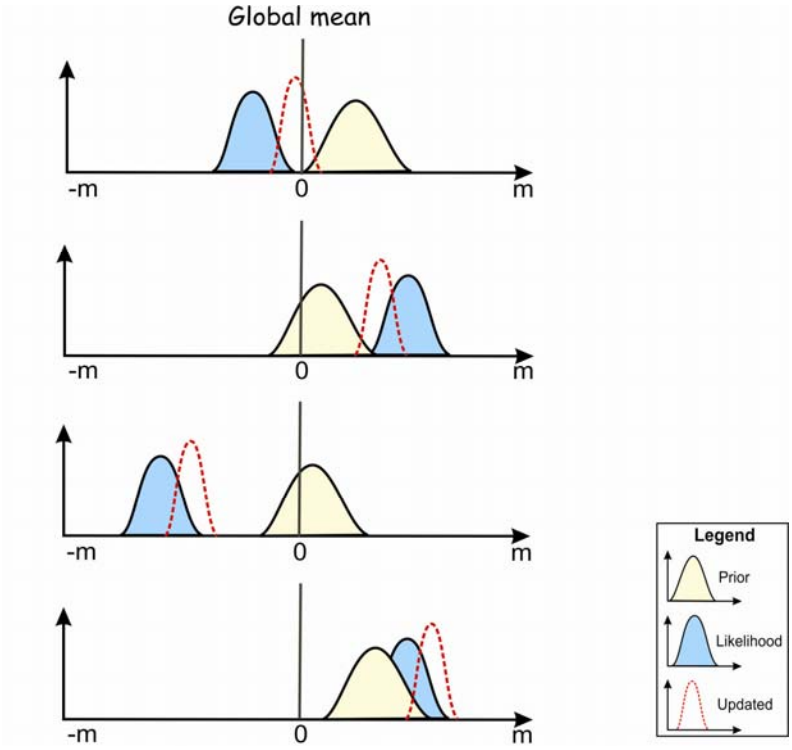

Figure 2: Schematic illustration of the location of updated distributions in Bayesian updating technique.

The updated variance is only affected by the prior and likelihood variances. It has the characteristic Gaussian property of homoscedasticity. **Error! Reference source not found.** shows the updated variance increases with increasing of prior variances. The updated variance is always the smallest variance among the three distributions. As shown in

Figure **2:** Schematic illustration of the location of updated distributions in Bayesian updating technique., the width of the updated distribution is always the narrowest, which indicates the uncertainty is reduced after Bayesian updating.

Table 2: The Bayesian updating calculation showing the increasing of updated variance (last row) with increasing of prior variance (second row).

Case	1	2	3	4	5	6	7
Prior Variance σ_P^2	0.10	0.30	0.40	0.50	0.60	0.70	0.90
Likelihood Variance σ_L^2	0.40	0.40	0.40	0.40	0.40	0.40	0.40
Updated Variance σ_U^2	0.09	0.21	0.25	0.29	0.32	0.34	0.38

The Bayesian updating technique relies on the multivariate Gaussian assumption. Under this assumption, all marginal and conditional distributions are Gaussian, and can be fully defined by an appropriate mean and variance. Therefore, the updated mean and variance provide a non-standard Gaussian distribution of uncertainty at the estimated location. Natural data may not follow Gaussian distribution. Normal score

transformation is required to transform all the variables into normal scores at the beginning. After Bayesian updating, the local conditional distribution of uncertainty must be back transformed to original units.

The multivariate Gaussian assumption also provides that all multivariate relationships are linear and can be characterized by correlation coefficients. The correlation coefficients obtained from the wells are considered to measure the true relationship of each pair of variables over the model area and applicable to the interwell regions under the assumption of stationary. However, in practice, the multivariate relationships may not be linear or stationary over a large area. Local correlation coefficient may be more representative of the true relationship. The reliability of secondary data should also be considered in the correlation with the primary data. Modifications were introduced to account for possible non-stationarity, complex multivariate relationships and quality of the secondary data in local uncertainty assessment (Ren *et al.*, 2006).

Secondary information, locally varying correlation and quality are used in Bayesian updating to improve the local estimation. However, regardless of how fine-tuned the estimates are, there are errors and uncertainties. Bayesian updating technique provides uncertainty distributions at all of locations. The mean values of updated distributions cannot be used as a realization because of smoothness. Simulated realizations are required for constructing fine scale models in the scale consistent modeling approach.

Simulation with Bayesian Updating

Under the multivariate Gaussian model, Bayesian updating can be performed within sequential Gaussian simulation (SGS). The basic procedure is to perform simple kriging sequentially using input data and previously simulated values to build local prior distributions (Equations 1 to 3), and establish the updated distribution using the prior distribution and the likelihood distribution (Equations 4 to 6) from collocated secondary data. Then, draw randomly from the updated distribution (Equations 12 and 13) to get the simulated value:

$$y_s(\mathbf{u}) = w(\mathbf{u})\sigma_u(\mathbf{u}) + \overline{y}_u(\mathbf{u}) \tag{14}$$

where $w(\mathbf{u})$ is a random number drawn from a standard normal distribution, $\sigma_u(\mathbf{u})$ and $\overline{y}_u(\mathbf{u})$ are the Bayesian updated standard deviation and mean, respectively.

For computing efficiency, rather than performing Bayesian updating sequentially, we can perform Bayesian updating to build local uncertainty distributions and use a simplified simulation approach, P-field simulation (Srivastava, 1992), to generate simulated realizations. P-field simulation starts with a probability field, that is, a set of spatially correlated probability values uniformly distributed between 0 and 1. Then, simulated values are drawn by sampling the local conditional distributions using the corresponding probability values.

The implementation of P-field simulation with Bayesian updating is to draw a set of standard normal deviates that are spatially correlated within the field *A*, and then condition these standard normal values with the Bayesian updated mean and standard deviation to get simulated values as shown in Equation 14. The only difference is that the $w(\mathbf{u})$ is the spatially correlated value from standard normal distribution.

An example of simulation with Bayesian updating is presented. A reference image of primary variable is given in the top left of Figure 3 for comparing purpose. The values in the black circles are used as primary data. A secondary variable (top right in Figure 3) is also provided for simulation. Both SGS with Bayesian updating and P-field simulation with Bayesian updating are performed using the primary and secondary data. Multiple realizations are generated. The P-field simulation realizations #1, #5 and #10 are shown together with the reference image, secondary data, and updated estimates in Figure 3. The SGS with Bayesian updating simulation realizations are shown in Figure 5. Figure 4 and Figure 6 shows the cross plots of reference versus the simulation realizations #1, #5 and #10. The collocated co-simulation (CC-SGS) is also performed to compare the two Bayesian updating simulation realizations. The collocated co-simulation is performed using the SGSim program from GSLIB (Deutsch and Journel, 1998). A variance

reduction factor of 0.6 is used to correct the variance inflation. The reference and the collocated co-simulation realizations #1, #5 and #10 are shown in Figure 7. The cross plots of the reference versus CC-SGS realizations #1, #5 and #10 are shown in Figure 8.

By visually comparing different simulation realizations with the reference image, the P-field simulation realizations are best in matching the overall trend of the reference. The correlations between the reference and P-field simulation realizations are the highest among the three simulation methods, and the points are closer to the 45^o line in the cross plots. The main reason for that is the updated results capture the trend of the reference very well (Figure 3). The variances of all simulation realizations are close to the variance of reference data. P-field simulation shows no variance inflation. SGS with Bayesian updating shows slightly higher variances. CC-SGS realizations would show much higher variances if no variance correction applied. The other two simulation methods show similar correlations between reference and simulation realizations. Overall, the P-field simulation approach has the best performance.

Conclusions

Bayesian updating technique is recommended for large scale modeling. The CCG format of Bayesian updating is same as the original format of Bayesian updating approach proposed by Doyen et al., 1996. The advantage of the CCG format is the easier implementation of integrating multiple secondary variables. Many sources of data can be used as secondary data. Secondary variables that contain trend or non-linear geological features can bring those features into Bayesian updating results. A large scale model can be very reliable after integrating different sources of data.

If the local uncertainties are already built in large scale modeling, the P-field simulation with Bayesian updating is recommended because of computational efficiency and the consistency between the local uncertainties and the simulation realizations. The comparison with collocated cosimulation, sequential Gaussian simulation with Bayesian updating indicates P-field simulation performs better in term of correlation between simulation realization and a reference image. P-field simulation may cause a slight error around wells. But they are considered negligible if the primary variable is quite spatially continuous.

References

Besag, J.: "On the Statistical Analysis of Dirty Pictures," J. R. Statist. Soc. B., 48, p. 259-302, 1986.

Deutsch, C.V. and Journel, A.G.: "GSLIB: Geostatistical software library and user's guide," 2nd edition, Oxford University Press, New York, 1998.

Deutsch, C. V. and Zanon, S.: "Direct Prediction of Reservoir Performance with Bayesian Updating Under a Multivariate Gaussian Model," Paper presented at the Petroleum Society's 5th Canadian International Petroleum Conference (55th Annual Technical Meeting), Calgary, Alberta, 2004.

Doyen, P.M., Den Boer, L.D., and Pilley, W.R.: "Seismic Porosity Mapping in the Ekofisk Field Using a New Form of Collocated Cokriging," Society of Petroleum Engineers, SPE 36498 presented at the SPE Annual Technical Conference and Exhibition, 1996.

Ren, W., McLennan, J., Leuangthong, O., and Deutsch, C.V.: "Reservoir Characterization of McMurray Formation by 2-D Geostatistical Modeling," Natural Resources Research, 2006a.

Ren, W., and Deutsch, C.V.: "Bayesian Updating with Local Varying Correlation," Centre for Computational Geostatistics Report Eight, Edmonton, Alberta, 2006b.

Ren, W., Deutsch, C.V., Garner, D., Wheeler, T.J., Richy, J.F., and Mus, E.: "Quantifying Resources for the Surmont Lease with 2D Mapping and Multivariate Statistics," SPE 102094 presented at the SPE Annual Technical Conference and Exhibition, October 2006c.

Ren, W.: PhD thesis: "Scale Consistent Geostatistical Modeling for Reservoir Characterization," University of Alberta, Edmonton, Alberta, 2007.

Srivastava, R. M.: "Reservoir Characterization with Probability Field Simulation," SPE Formation Evaluation, 7(4), pp. 927–937, 1992.

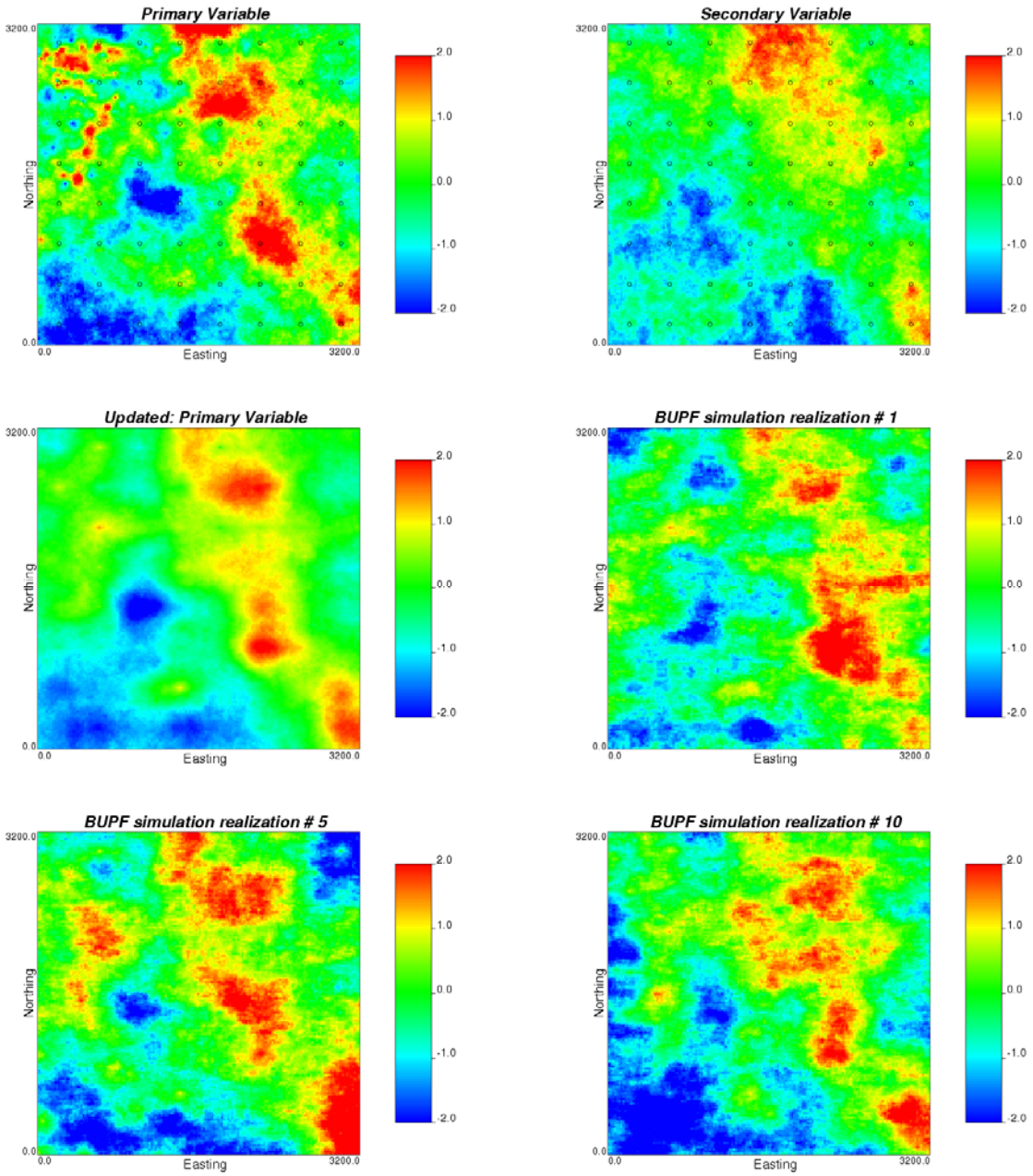

Figure 3: The primary variable (reference) and secondary variable used for Bayesian updating are in the top row. The Bayesian updated estimates and P-field simulation realizations are shown in the middle and bottom rows.

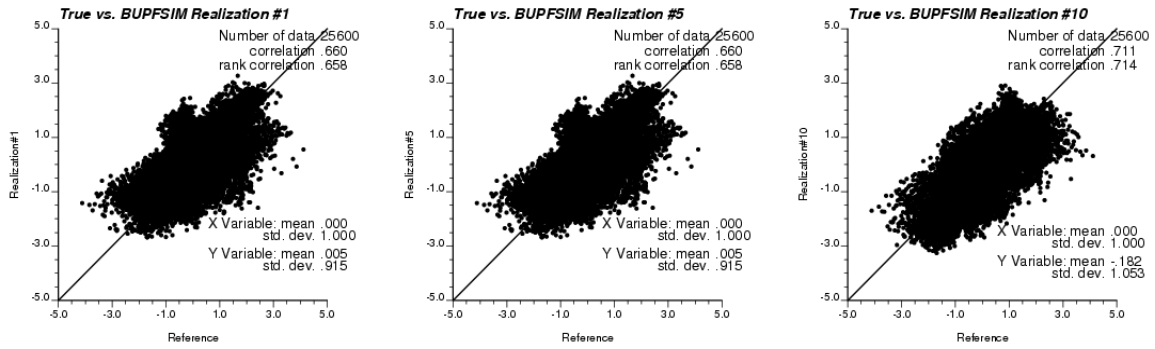

Figure 4: The cross plots of the reference versus the P-field simulation realizations #1, #5 and #10.

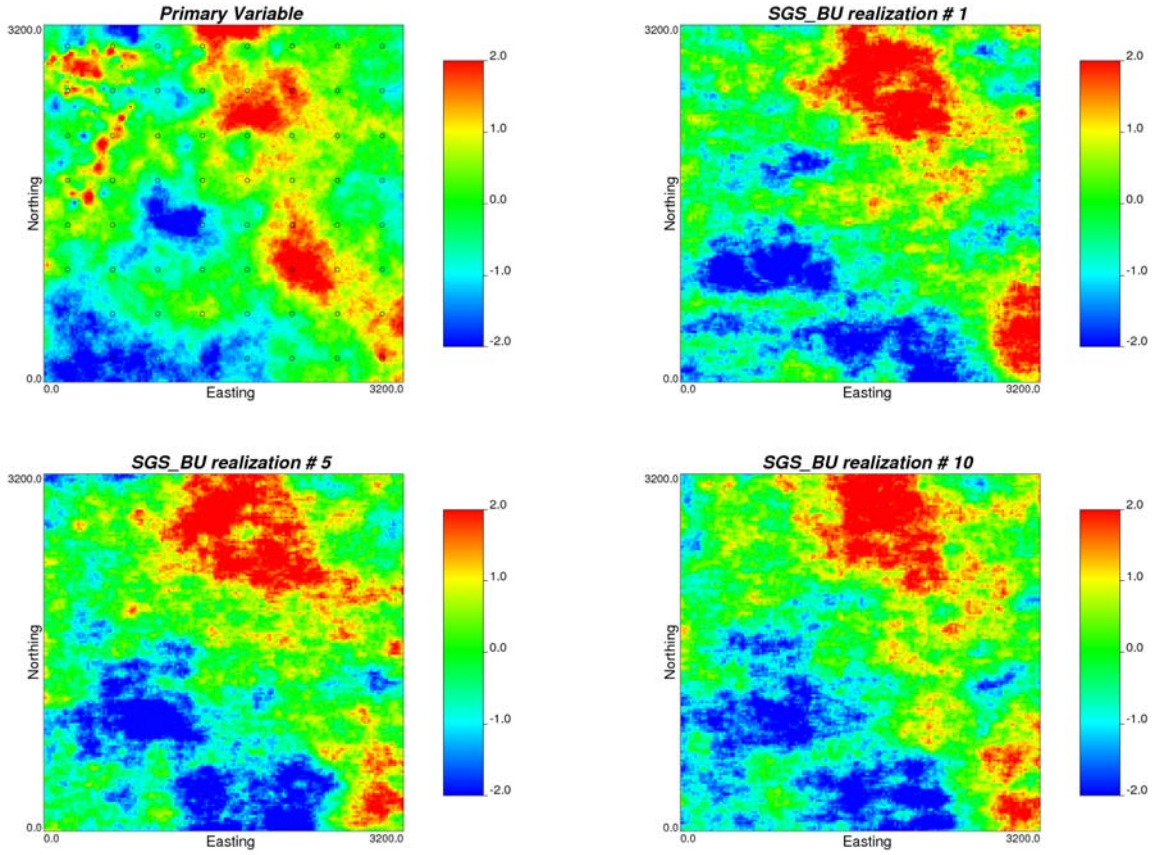

Figure 5: The primary variable (reference, up-left) and the SGS with Bayesian updating simulation realizations.

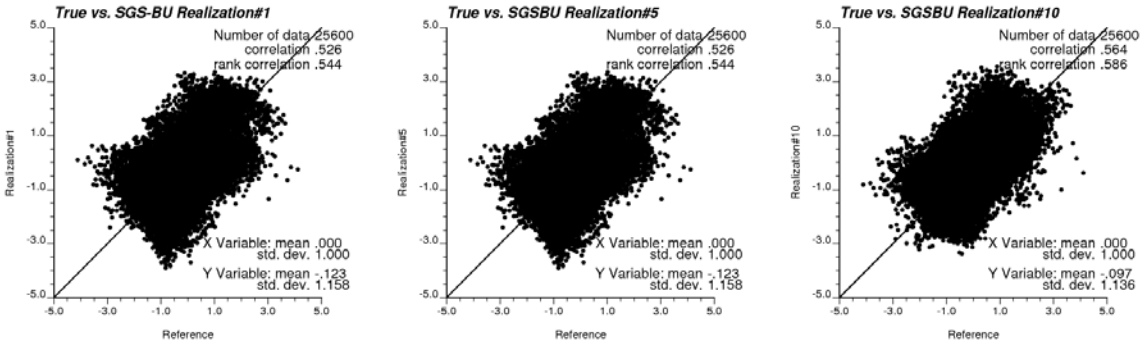

Figure 6: The cross plots of the reference versus the SGS with Bayesian updating simulation realizations #1, #5 and #10.

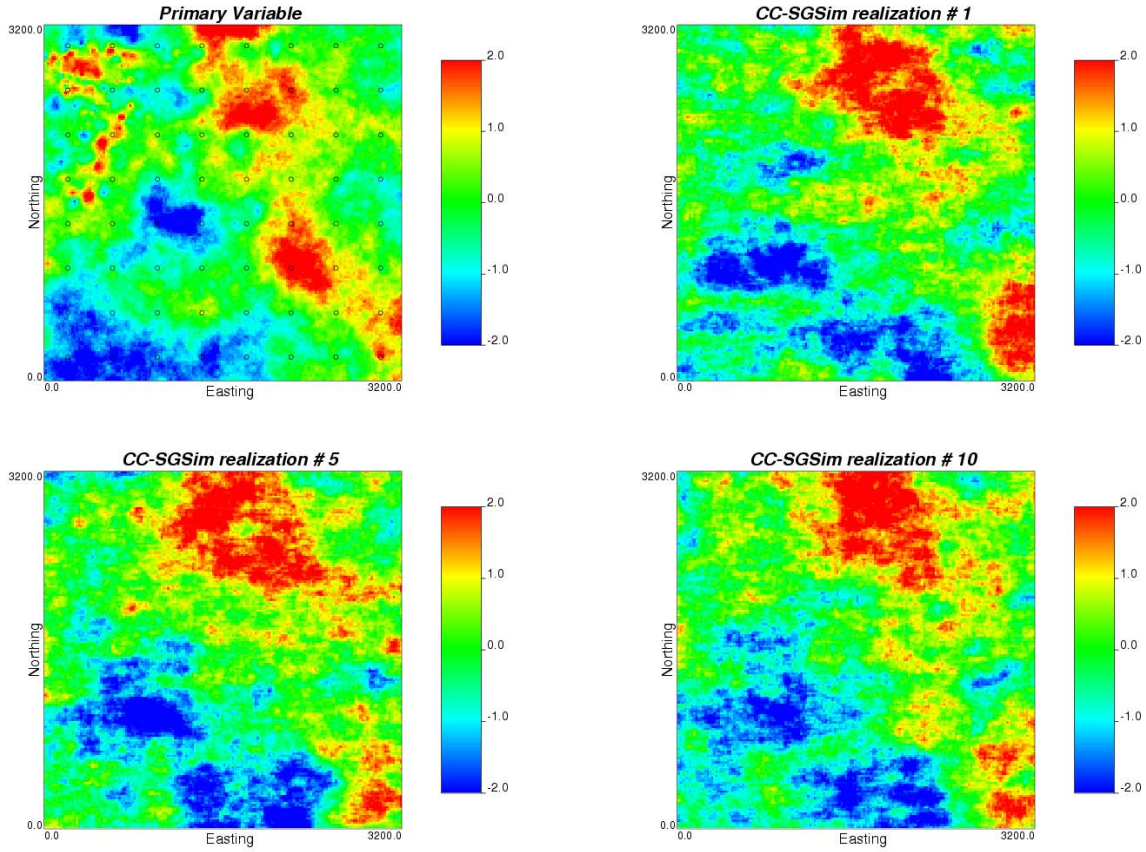

Figure 7: The primary variable (reference, up-left) and the collocated co-simulation realizations.

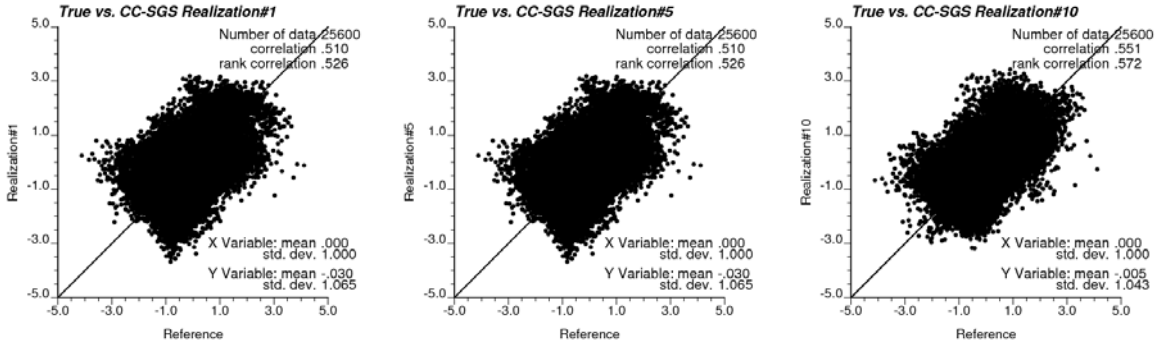

Figure 8: The cross plots of the reference versus the collocated co-simulation realizations #1, #5 and #10.