"Research Note"

BAYESIAN PREDICTION IN GEOSTATISTICAL MODELS WITH MATERN CORRELATION FUNCTION[*]

M. JAFARI KHALEDI

Department of Statistics, Tarbiat Modares University, P.O. Box 14115-134, Tehran, I. R. of Iran
Email: jafari-m@modares.ac.ir

Abstract – This paper deals with Bayesian geostatistical prediction under the Matern correlation function which involves a smoothness parameter in addition to the range parameter. In fact, we consider the reference prior for the range parameter and an inverse gamma prior for the smoothness parameter. We then propose an attractive and straightforward Monte Carlo method to sample the posterior distribution of the model parameters and achieve Bayesian prediction. In a sensitivity analysis, the importance of the prior choice is assessed. Since the posterior results greatly depend on the prior hyperparameters, the Monte Carlo EM algorithm is applied to determine their maximum likelihood estimates. Finally, we utilize this procedure in the geostatistical prediction of carbon monoxide concentrations in Tehran.

Keywords – Geostatistics, Bayesian, prior sensitivity, reference prior

1. INTRODUCTION

Geostatistical models are commonly used for analyzing spatial data and have been extensively used in different areas of spatial statistics. In these models, usually the general and flexible Matern family which involves a smoothness parameter in addition to the range parameter, is used to model the correlation structure [1]. In geostatistical data analysis, the main interest is to predict the random field in some unmeasured sites. For this purpose, in recent years, the Bayesian method has been very widely used [2, 3]. However, selection of the prior distribution for the parameters of the Matern family requires some extra care as they can be difficult to interpret and hence, difficult to elicit. Moreover, the smoothness parameter and the range parameter are usually highly correlated, so assuming them to be independent a priori would give nonsensical results. Assuming that the smoothness parameter is fixed, Berger *et al.* [4] introduced a reference prior for the range parameter which allows the posterior to be proper. One of the main advantages of this prior is that it depends on the smoothness parameter. But due to its complexity, it is quite difficult to use the usual the Monte Carlo methods for sampling from the corresponding posterior distribution. To the best of our knowledge, no published work using this prior for spatial prediction is available. Consequently, in the hierarchical Bayes approach, a vague proper prior is used [5, 6]. But, in this case, a sensible choice of the hyperparameters is crucial since they may have an unpleasant influence on prediction and inference. Also, to make Bayesian inference feasible, the range parameter is considered independent of the smoothness parameter. In the current paper, adopting a full Bayesian approach, we consider the reference prior and an inverse gamma prior for the range and smoothness parameters, respectively. We then propose an attractive and straightforward Monte Carlo method to sample the posterior distribution of the model parameters. Applying this method, we can generate independent

---

samples from the posterior distribution, which is much more efficient for spatial prediction. In a sensitivity analysis, the importance of the prior choice is assessed. Since the posterior results depend greatly on unknown hyperparameters, we propose to use the Monte Carlo Expectation-Maximization (MCEM) algorithm to get the maximum likelihood estimates (MLE) of the hyperparameters. Finally, we demonstrate that the procedure satisfactorily offers predictive performance in a real data set.

2. STATISTICAL MODEL

Let $Z(\cdot) = \{Z(t), \ t \in D\}$ be a Gaussian random field with linear mean $E[Z(t)] = f'(t)\beta$ and covariance function $Cov[Z(u), Z(t)] = \sigma^2 \rho(u, t)$. Here, $f(t) = (f_1(t), \ldots, f_p(t))'$ is a vector of covariates associated with vector t, $\beta = (\beta_1, \ldots, \beta_p)$ is a vector of regression parameters, $\sigma^2 = Var[Z(t)]$ is the fixed variance of $Z(\cdot)$ and $\rho(u, t)$ is the spatial correlation function. For simplicity we assume that $\rho(u, t)$ only depends on the distance $h = \| u - t \|$. Moreover, we use a correlation function from the Matern class formulated as $\rho(h; \vartheta) = \{2^{\nu-1} \Gamma(\nu)\}^{-1} (h/\theta)^{\nu} K_{\nu}(h/\theta)$ where $\nu > 0$ and $\theta > 0$ denote the smoothness parameter and the range parameter, respectively, $\vartheta = (\theta, \nu)$ and $K_{\nu}(\cdot)$ is the modified Bessel function of order ν. Moreover, it is assumed that the random vector $Z = (Z(t_1), \ldots, Z(t_n))$ represents the data measured at the sampling locations $t_1, \ldots, t_n \in D$. By the stated assumptions, we have $Z = (Z(t_1), \ldots, Z(t_n)) \sim N(X\beta, \sigma^2 \Sigma_{\vartheta})$, where $X = (f_j(t_i))$ is the known full column rank $n \times p$ matrix, $n > p$, and $\Sigma_{\vartheta} = (\rho(\| t_i - t_j \|; \vartheta))$ is a positive definite $n \times n$ matrix. For Bayesian analysis it is necessary to specify prior distribution for the model parameters $\eta = (\beta, \sigma^2, \vartheta)$. Moreover, we assume that, a priori, the vector ϑ to be independent of the parameters (β, σ^2) and $\pi(\beta, \sigma^2) \propto \dfrac{1}{\sigma^2}$, which is the asymptotic case of the conjugate normal-inverse gamma prior. Really, if in priors $\beta \mid \sigma^2 \sim N(\beta_0, \sigma^2 \Gamma_0)$ and $\sigma^2 \sim IG(c, d)$, where IG means Inverse Gamma, we set $\Gamma_0^{-1} \longrightarrow 0, c \longrightarrow 0$ then $\pi(\beta, \sigma^2) \propto \dfrac{1}{\sigma^2}$. Also, this joint prior is the Jeffreys' prior for (β, σ^2). It must be noted that $\int\int_{R^+ R^p} \dfrac{1}{\sigma^2} d\beta d\sigma^2 = \infty$, so this joint prior is clearly improper. Then, the prior densities satisfy $\pi(\eta) \propto \dfrac{\pi(\theta \mid \nu)\pi(\nu)}{\sigma^2}$. We consider the reference prior for θ developed by Berger *et al.* [4] as

$$\pi^R(\theta \mid \nu) \propto \left\{ tr[W_{\vartheta}^2] - \frac{1}{n-p}(tr[W_{\vartheta}])^2 \right\}^{\frac{1}{2}}, \tag{1}$$

where $W_{\theta} = \dfrac{\partial \Sigma_{\vartheta}}{\partial \theta} \Sigma_{\vartheta}^{-1}(I - X(X' \Sigma_{\vartheta}^{-1} X)^{-1} X' \Sigma_{\vartheta}^{-1})$. Here $\dfrac{\partial \Sigma_{\vartheta}}{\partial \theta}$ denotes the matrix obtained by differentiating Σ_{ϑ} with respect to θ element by element and I is an identity matrix of order *n*. We have

$$\frac{\partial \rho(h; \vartheta)}{\partial \theta} = \{2^{\nu-1} \Gamma(\nu)\}^{-1} \frac{\partial((h/\theta)^{\nu} K_{\nu}(h/\theta))}{\partial \theta}$$

$$= -\frac{h^{\nu}}{2^{\nu-1} \Gamma(\nu) \theta^{\nu+1}} [\nu K_{\nu}(h/\theta) + (h/\theta) K'_{\nu}(h/\theta)]$$

where according to Abramowitz and Stegun [7], the derivative formula of modified Bessel function is $K'_\nu(x) = -(K_{\nu-1}(x) + \frac{\nu}{x}K_\nu(x))$. As observed, the reference prior (1) depends on the smoothness parameter, and Berger *et al.* [4] showed that it is also proper, i.e. $\int_{R^+} \pi^R(\theta|\nu)\,d\theta = 1$. Further, to avoid having an improper posterior distribution, we consider a proper inverse gamma prior for ν as $IG(a,b)$, where a and b are unknown hyperparameters.

Based on the prior distribution, the joint posterior distribution of the model parameters is given by

$$\pi(\eta|z,a,b) = \pi(\beta|z,\sigma^2,\vartheta)\pi(\sigma^2|z,\vartheta)\pi(\vartheta|z,a,b).$$

Since

$$\pi(\beta|z,\sigma^2,\vartheta) \propto f(z|\beta,\sigma^2,\vartheta)\pi(\beta)$$
$$\propto \exp\{-\frac{1}{2\sigma^2}(z-X\beta)'\Sigma_\vartheta^{-1}(z-X\beta)\}$$
$$\propto \exp\{-\frac{1}{2\sigma^2}(\beta'\hat{V}^{-1}\beta - 2\beta'\hat{V}^{-1}\hat{\beta})\}$$
$$\propto \exp\{-\frac{1}{2\sigma^2}(\beta-\hat{\beta})'\hat{V}^{-1}(\beta-\hat{\beta})\},$$

where $\hat{V} = (X'\Sigma_\vartheta^{-1}X)^{-1}$ and $\hat{\beta} = \hat{V}(X'\Sigma_\vartheta^{-1}z)$, then

$$[\beta|z,\sigma^2,\vartheta] \sim N(\hat{\beta},\sigma^2\hat{V}).$$

Also, we have

$$\pi(\sigma^2|z,\vartheta) \propto f(z|\sigma^2,\vartheta)\pi(\sigma^2)$$
$$\propto \frac{f(z|\beta,\sigma^2,\vartheta)\pi(\beta)}{\pi(\beta|z,\sigma^2,\vartheta)} \times \frac{1}{\sigma^2}$$
$$\propto (\sigma^2)^{-(\frac{n-p}{2}+1)}\exp\{-\frac{1}{2\sigma^2}(z'\Sigma_\vartheta^{-1}z - \hat{\beta}'\hat{V}^{-1}\hat{\beta})\}.$$

Then, $[\sigma^2|z,\vartheta] \sim Inv\chi^2(n-p,S^2)$, where $S^2 = \dfrac{z'\Sigma_\vartheta^{-1}z - \hat{\beta}'\hat{V}^{-1}\hat{\beta}}{n-p}$.

We will use common notation for an inverse chi square, i.e. $X \sim Inv\chi^2(\lambda,\delta)$ means $f(x;\lambda,\delta) = \dfrac{(\delta\frac{\lambda}{2})^{\frac{\lambda}{2}}}{\Gamma(\frac{\lambda}{2})} \dfrac{\exp[-\frac{\lambda\delta}{2x}]}{x^{1+\frac{\lambda}{2}}}$, where $x > 0$, $\lambda > 0$ and $\delta > 0$.

Therefore, the joint posterior distribution of (β,σ^2), given by $\pi(\beta,\sigma^2|z,\vartheta) = N(\hat{\beta},\sigma^2\hat{V}) \times Inv\chi^2(n-p,S^2)$, is a proper distribution. Also, the marginal posterior distribution $\pi(\vartheta|z,a,b)$ is in proportion to

$$\pi(\vartheta|z,a,b) \propto f(z|\vartheta)\pi(\vartheta|a,b)$$
$$\propto \frac{f(z|\beta,\sigma^2,\vartheta)\pi(\beta,\sigma^2)}{\pi(\beta|z,\sigma^2,\vartheta)\pi(\sigma^2|z,\vartheta)}\pi(\vartheta|a,b)$$
$$\propto h(z|\vartheta)\pi(\vartheta|a,b),$$

where $h(z|\vartheta) = |\hat{V}|^{\frac{1}{2}}|\Sigma_\vartheta|^{-\frac{1}{2}}(S^2)^{-\frac{n-p}{2}}$. However, this expression does not define a standard probability distribution. In order to have a feasible computation of the posterior distribution $\pi(\vartheta|z,a,b)$, we

M. Jafari Khaledi

consider a 2 dimensional grid of values belonging to the support set of ϑ with $m = m_1 \times m_2$ nodes. Then, the approximated posterior distribution of the ϑ is

$$\pi(\vartheta_i \mid z, a, b) = \frac{h(z \mid \vartheta_i)\pi(\vartheta_i \mid a, b)}{\sum_{j=1}^{m} h(z \mid \vartheta_j)\pi(\vartheta_j \mid a, b)} \quad \text{for } i = 1, \cdots, m.$$

In order to sample from the posterior distribution, we compute the posterior probabilities $\pi(\vartheta \mid z, a, b)$ on the discrete support set and sample from this distribution. One advantage of the above method is that we can generate independent samples from the posterior distribution, which is much more efficient for spatial prediction. Furthermore, it is fast and easy to implement.

We now consider the prediction of $Z(t_0)$ based on the Bayesian predictive distribution, defined by $f(z_0 \mid z, a, b) = \int f(z_0 \mid z, \beta, \sigma^2, \vartheta)\pi(\beta, \sigma^2, \vartheta \mid z, a, b)d\vartheta$. Since the random field is Gaussian, then $[Z(t_0), Z \mid \beta, \sigma^2, \theta] \sim N(\mu^*, \Gamma^*)$, where

$$\mu^* = \begin{pmatrix} f'(t_0)\beta \\ X\beta \end{pmatrix} \text{ and } \Gamma^* = \sigma^2 \begin{pmatrix} 1 & r_\vartheta' \\ r_\vartheta & \Sigma_\vartheta \end{pmatrix}, \text{ with } r_\vartheta = (\rho(\|t_i - t_0\|; \vartheta).$$

Therefore,

$$[Z(t_0) \mid z, \beta, \sigma^2, \vartheta] \sim N(f'(t_0)\beta + r_\vartheta'\Sigma_\vartheta^{-1}(z - X\beta), \sigma^2(1 - r_\vartheta'\Sigma_\vartheta^{-1}r_\vartheta)).$$

The distributions of $[\beta \mid z, \sigma^2, \vartheta]$ and $[Z(t_0) \mid z, \eta]$ are normal, so $[Z(t_0) \mid z, \sigma^2, \vartheta]$ is also normal with

mean

$$\begin{aligned}
\mu_1 &= E(Z(t_0) \mid z, \sigma^2, \vartheta) = E(E(Z(t_0) \mid z, \beta, \sigma^2, \vartheta) \mid z) \\
&= E[f'(t_0)\beta + r'_\vartheta \Sigma_\vartheta^{-1}(z - X\beta) \mid z] \\
&= f'(t_0)E(\beta \mid z, \sigma^2, \vartheta) + r'_\vartheta \Sigma_\vartheta^{-1}(z - XE(\beta \mid z, \sigma^2, \vartheta)) \\
&= f'(t_0)\hat{\beta} + r'_\vartheta \Sigma_\vartheta^{-1}(z - X\hat{\beta}) \\
&= (f'(t_0) - r'_\vartheta \Sigma_\vartheta^{-1}X)\hat{\beta} + r'_\vartheta \Sigma_\vartheta^{-1}z
\end{aligned}$$

and variance

$$\begin{aligned}
Var(Z(t_0) \mid z, \sigma^2, \vartheta) &= Var[E(Z(t_0) \mid z, \beta, \sigma^2, \vartheta) \mid z] \\
&\quad + E[Var(Z(t_0) \mid z, \beta, \sigma^2, \vartheta) \mid z] \\
&= Var[f'(t_0)\beta + r'_\vartheta \Sigma_\vartheta^{-1}(z - X\beta) \mid z] \\
&\quad + E[\sigma^2(1 - r'_\vartheta \Sigma_\vartheta^{-1}r_\vartheta) \mid z] \\
&= f'(t_0)Var(\beta \mid z, \sigma^2, \vartheta)f(t_0) + Var(r'_\vartheta \Sigma_\vartheta^{-1}X\beta \mid z, \sigma^2, \vartheta) \\
&\quad + \sigma^2(1 - r'_\vartheta \Sigma_\vartheta^{-1}r_\vartheta) \\
&= \sigma^2 f'(t_0)\hat{V}f(t_0) + \sigma^2(r'_\vartheta \Sigma_\vartheta^{-1}X)\hat{V}(r'_\vartheta \Sigma_\vartheta^{-1}X)' + \sigma^2(1 - r'_\vartheta \Sigma_\vartheta^{-1}r_\vartheta) \\
&= \sigma^2 \rho_1
\end{aligned}$$

where $\rho_1 = f'(t_0)\hat{V}f(t_0) + (r'_\vartheta \Sigma_\vartheta^{-1}X)\hat{V}(r'_\vartheta \Sigma_\vartheta^{-1}X)' + (1 - r'_\vartheta \Sigma_\vartheta^{-1}r_\vartheta)$. Also, we have

$$f(z_0 \mid z, \vartheta) = \int f(z_0 \mid z, \sigma^2, \vartheta) \pi(\sigma^2 \mid z) d\sigma^2$$

$$\propto \int (\sigma^2)^{-(\frac{n-p+1}{2}+1)} \exp\{-[\frac{(z_0 - \mu_1)^2}{2\sigma^2 \rho_1} + \frac{(n-p)S^2}{2\sigma^2}]\}d\sigma^2$$

$$\propto \int (\sigma^2)^{-(\frac{n-p+1}{2}+1)} \exp\{-\frac{(n-p)S^2}{2\sigma^2}[\frac{(z_0 - \mu_1)^2}{(n-p)S^2 \rho_1} + 1]\}d\sigma^2$$

$$\propto [\frac{(z_0 - \mu_1)^2}{(n-p)S^2 \rho_1} + 1]^{-\frac{n-p+1}{2}}$$

Then, $[Z(t_0) \mid z, \vartheta] \sim T_{n-p}(\mu_1, \sigma^2 \rho_1)$. In order to sample the Bayesian Predictor distribution $f(z_0 \mid z, a, b) = \int f(z_0 \mid z, \vartheta) \pi(\vartheta \mid z, a, b) d\vartheta$, we proceed as follows. Attaching the sampled values of ϑ from the posterior distribution to $f(z_0 \mid z, \vartheta)$ and sampling from it, we obtain realizations from the Bayesian predictive distribution. Therefore, the mean and variance of these realizations are the Bayesian spatial predictor and prediction variance, respectively.

To carry out Bayesian prediction, the hyperparameters a and b must be specified. Adopting the empirical Bayes method, we determine $\hat{a}$ and $\hat{b}$ which maximize the marginal likelihood, given as $l(a, b \mid z) = \int f(z \mid \vartheta) \pi(\vartheta \mid a, b) d\vartheta$. As there is no explicit solution for the corresponding likelihood equation, we use the EM algorithm (Dempster *et al.* [8]) to approximate $\hat{a}$ and $\hat{b}$. At the $j+1$th iteration, the E-step involves the calculation of

$$Q(a, b \mid a^{(j)}, b^{(j)}) := E^*(\ln \pi(v \mid a, b))$$

$$= a \ln b - \ln \Gamma(a) - (a+1)E^*(\ln v) - bE^*(\frac{1}{v})$$

where E^* denotes the mathematical expectation under the posterior distribution $\pi(\vartheta \mid z, a^{(j)}, b^{(j)})$ with the current estimates of the corresponding hyperparameters used at iteration j, i.e. $a^{(j)}$ and $b^{(j)}$. The M-step consists of maximizing $Q(a, b \mid a^{(j)}, b^{(j)})$ in a and b to yield the new update $a^{(j+1)}$ and $b^{(j+1)}$. The process is iterated from a starting value $a^{(0)}$ and $b^{(0)}$ to convergence. Obtaining a closed form for $Q(a, b \mid a^{(j)}, b^{(j)})$ is not possible, as it requires knowledge of the conditional distribution of v given z evaluated at $a^{(j)}$ and $b^{(j)}$. In this case, we approximate the expectation applying the sampled values from the posterior distribution $\pi(\vartheta \mid z, a^{(j)}, b^{(j)})$. Finally, we replace $\hat{a}$ and $\hat{b}$ obtained by this algorithm, the so-called the Monte Carlo EM algorithm, in the Bayesian predictive distribution.

3. NUMERICAL EXAMPLE

The data set analyzed in this section is comprised of the carbon monoxide (CO) concentration, accumulated for a year beginning from 21 March 2003, in n=11 monitoring sites distributed geographically across Tehran, Iran. It is of interest to provide an annual prediction surface for the entire region. Taking the square root of this annual average concentration made the data distribution nearer to the Gaussian distribution. Thus, in the sequel, we model on the square root of CO concentration. The exploratory analysis of the transformed data shows that a second order polynomial of spatial covariates would be a reasonable structure for spatial trend. In the sequence, following Banerjee [5] we consider the interval (0, 2) as the effective range of v. Also, based on profile log-likelihood, we consider the interval (0, 10) as the effective range of θ. Then, we discretize the effective ranges of the correlation parameters by considering a regular grid with $m_1 \times m_2$ nodes. To determine a sensible value for the number of nodes, we choose five different locations at the four main directions and one at the center of Tehran. The sensitivity analysis showed that for $m_1 = 1000$ and $m_2 = 20$, spatial predictions in these locations are almost robust.

In order to assess how sensitive our posterior inference is to the choice of the prior, we consider some values for the hyperparameters. Table 1 displays the various values of the prior hyperparameters, as well as the posterior mean and 95% credible interval of ν. We observe that the posterior results are sensitive to prior changes. Then, there is relatively little direct information in the data on ν, so the prior is quite important. To elicit reasonable prior distribution, we apply the empirical Bayes method as described in section 2. In this way, we obtain $\hat{a} = 6.3$ and $\hat{b} = 2.8$.

Table 1. Prior sensitivity analysis

Hyperparameters	Mean	95% Interval
a=1, b=1	21.17	(16.33,24.69)
a=1, b=2	10.23	(3.91, 15.06)
a=2, b=1	11.48	(5.78,16.26)
a=2, b=2	17.79	(13.22,23.76)

We now compare the predictive performance of the proposed prediction procedure with the Kriging method. The mean square prediction residual of the proposed and Kriging methods are determined as 1.386 and 2.431, respectively. Likewise, we obtained the 95% Bayesian predictive intervals. Of the 11 95% predictive intervals, which we compared with actual observations, 1 and 4 failed to include the observed value corresponding to the two methods. As a result, we have concluded that the proposed method has a more superior performance than that of the Kriging method. Finally, we plot the map of the Bayesian spatial predictions and the standard deviations of the annual predictions (Fig. 1). According to this figure, the predictions of CO concentration are high in the central part of the city, mainly caused by heavy traffic. Further, as expected, standard deviations are higher at locations far away from the monitoring stations.
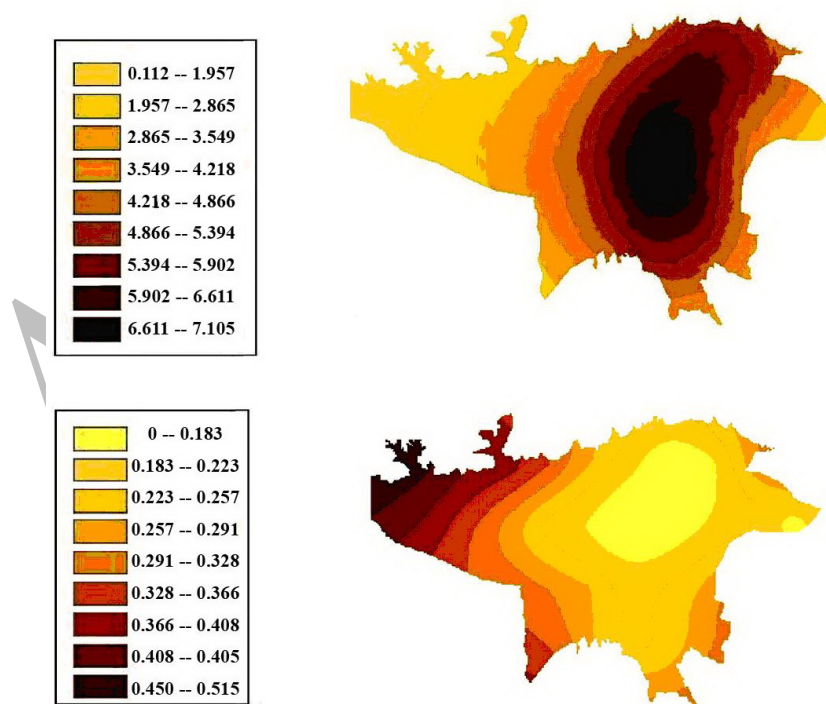

Fig. 1. Predicted CO concentration (up) and Standard deviations predictions (down)

REFERENCES

1. Stein, M. (1999). *Spatial Interpolation, Some Theory for Kriging.* Springer-Verlag, New York.
2. Diggle, P. J. & Ribeiro, P. J. (2007). *Model Based Geostatistics.* Springer-Verlag, New York.
3. Ecker, M. & De Oliveira, V. (2008), Bayesian spatial modeling of housing prices subject to a localized externality. *Communications in Statistics: Theory & Methods. 37,* 2066-2078.
4. Berger, J. O., De Oliveira, V. & Sanso, B. (2001). Objective bayesian analysis of spatially correlated data. *Journal of the American Statistical Association. 96,* 1361-1374.
5. Banerjee, S. (2005). On geodetic distance computations in spatial modeling. *Biometrics. 61,* 617-625.
6. Palacios, M. B. & Steel, M. J. F. (2006) Non-Gaussian Bayesian Geostatistical Modeling. *Journal of the American Statistical Association. 101,* 604-618.
7. Abramowitz, M. & Stegun, I. (1965). *Handbook of Mathematical Functions.* New York: Dover.
8. Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society* B. *96,* 1361-1374.