Journal of Mathematical Sciences and Informatics Vol.1, No.1 (2006), pp 33-46

Bayesian Prediction in Spatial Data Analysis

Khaledi, M.J. & Mohammadzadeh, M.*

Department of Statistics, Faculty of Science,
Tarbiat Modarres University, Tehran, Iran.
Email:mohsen_m@modares.ac.ir

Abstract. When spatial data are realizations of a Gaussian model with parametric mean and covariance functions, then the function of observations that minimizes mean square prediction error depends on some unknown parameters. Usually, these parameters are replaced by their estimates to obtain the plug-in predictor. But, this method has some problems in estimation of the parameters and the optimality and mean square error of the spatial predictor. In this paper, the problems related to plug-in method are discussed and to avoid them, the Bayesian approach for spatial prediction is proposed. Then the Bayesian spatial prediction for Gaussian and trans Gaussian models according to observations, that may contain noise, are derived. Next, in a simulation study, the adequacy of Bayesian prediction is compared with plug-in prediction. Finally, a numerical example illustrates the Bayesian spatial prediction of rainfall in a region at the north of Iran.

1 Introduction

Usually, the spatial data collected from some applied disciplines such as petroleum engineering, civil engineering, geography, geology, meteorology and epidemiology, are thought as realizations of a random field $S(\cdot) = \{S(t)\ ,\ t \in D\}$, where D is an index set in R^d, $d \geq 1$. A common scientific purpose in spatial data analysis is prediction of the random field $S(\cdot)$ in an unmeasured site t_0, say $S(t_0)$, based on measured data in some sampled sites $t_1, \cdots, t_n$ in D. If the random field $S(\cdot)$ is a Gaussian model with parametric mean and covariance functions, then the function of observations that minimizes

AMS: 62H11
Keywords: Spatial Data, Gaussian Model, Trans Gaussian Model, Noisy Observation, Bayesian Spatial Prediction.

* Corresponding author.

mean square prediction error depends on parameters. In the classical spatial prediction, the parameter estimates are replaced in this function to obtain the plug-in predictor, [5] and [4]. In this paper, to avoid problems related to plug-in method, the Bayesian approach is proposed to predict Gaussian spatial models. Also, if, because of measurement error, $S(\cdot)$ is not directly observable and instead of $S = (S(t_1), ..., S(t_n))$, the noisy measurements $Z = (Z(t_1), ..., Z(t_n))$ are taken at sample locations $t_1, \cdots, t_n$, the Bayesian spatial prediction and variance prediction are determined.
In some applications such as rainfall and mining studies, data give evidence of non-Gaussian features, a specific transformation (e.g. a logarithm transformation) may be successful in representing this type of data. [11] considered a Noisy Log-Gaussian model for Bayesian spatial prediction. When a logarithm transformation may not be appropriate, we would apply an estimate of the normalizing transformation in a Bayesian framework. In this matter, the posterior distributions and the Bayesian predictive distribution are analytically determined by a discretisation method. Also, a Bayesian spatial predictor, based on the absolute error loss function, and a measure of prediction uncertainty is derived.

The plug in method for a Gaussian spatial model is described in section 2. Section 3 deals with the determination of Bayesian spatial prediction for the Gaussian model and extension of the prediction problem of this model to the noisy Gaussian and trans Gaussian spatial models. We perform some simulation studies to compare two spatial prediction methods according to cross validation mean square error criterion in section 4. A numerical example, in section 5, illustrates the Bayesian spatial prediction of rainfall in a north region of Iran. Finally, results and discussion are given in section 5.

2 The Spatial Model

Let $\{S(x), x \in D\}$ be a Gaussian random field with mean and covariance functions

$$E[S(t)] = f'(t)\beta = \Sigma_{j=1}^{p} \beta_j f_j(t)$$
$$Cov[S(u), S(t)] = \sigma^2 \rho(u, t; \theta),$$

where $\beta = (\beta_1, ..., \beta_p)' \in R^p$ are unknown regression parameters, $f(t) = (f_1(t), ...f_p(t))'$ are known location-dependent covariates, $\sigma^2 = Var[S(t)]$ is the fixed variance of $S(\cdot)$, $\rho(u, t; \theta)$ is a spatial correlation function and $\theta = (\theta_1, ..., \theta_q)' \in \Theta \subseteq R^q$ are parameters that control geometric aspects of the transformed random field, such as the range and smoothness, as well as the other aspects of the spatial data association structure. If the random vector $S = (S(t_1), ..., S(t_n))$ represents the data measured at the sampling locations

$t_1, ..., t_n \in D$, then we have

$$S \sim N_n(X\beta, \sigma^2 \Sigma_\theta)$$

where $X = (f_j(t_i))$ is the known full rank $n \times p$ matrix, $n > p$, $\Sigma_\theta = (\rho(t_i, t_j; \theta))$ is a positive definite $n \times n$ matrix. The likelihood function of the model parameters $\eta = (\beta, \sigma^2, \theta)$ based on the observed data $s = (s(t_1), ..., s(t_n))$ is given by

$$L(\eta; s) = (\frac{1}{2\pi\sigma^2})^{n/2} |\Sigma_\theta|^{-1/2} \exp\{-\frac{1}{2\sigma^2}(s - X\beta)' \Sigma_\theta^{-1}(s - X\beta)\}$$

Since the joint distribution of $(S(t_0), S)$ is given by

$$N_{n+1}(\begin{pmatrix} f'(t_0)\beta \\ X\beta \end{pmatrix}, \sigma^2 \begin{pmatrix} 1 & r_\theta \\ r'_\theta & \Sigma_\theta \end{pmatrix})$$

where $r_\theta = (\rho_\theta(t_i, t_0))$ is a $n \times 1$ vector, it can be shown that $S(t_0)$ given s and η is normally distributed as

$$(S(t_0)|s, \eta) \sim N(\mu_1, \sigma^2 \rho_1)$$

where

$$\mu_1 = f'(t_0)\beta + r'_\theta \Sigma_\theta^{-1}(s - X\beta), \qquad \rho_1 = (1 + \alpha^2 - r'_\theta \Sigma_\theta^{-1} r_\theta). \quad (10)$$

If the parameters are known, the optimal predictor corresponding to the squared error loss function and the prediction variance, are given by

$$\hat{S}_\eta(t_0) = E(S(t_0)|s, \eta) = \mu_1$$
$$Var(S(t_0)|s, \eta) = \sigma^2 \rho_1 \quad (11)$$

When η is unknown, its maximum likelihood estimates is usually replaced in $\hat{S}_\eta(t_0)$ to obtain the plug-in predictor. There are some problems related to this prediction method. First, [12], [14] and [16] have noted that the likelihood function is often multimodal, making the numerical identification of the maximum a difficult numerical problem, and often the maximum is away from the true value. Second, the MSPE of $\hat{S}_{\hat{\eta}}(t_0)$, namely $E[S(t_0) - \hat{S}_{\hat{\eta}}(t_0)]^2$, has no closed form. [17] proposed an estimate of MSPE, but [15] showed that under some conditions, it underestimate the MSPE. This would result into an overestimation of plug-in predictor precision, and in consequence the prediction intervals of $S(t_0)$ tend to be too optimistic. Third, since a predictor that uniformly for each η, minimises mean square prediction error does not exist, the plug-in predictor is not optimal. To avoid the mentioned problems related to the plug-in method, the Bayesian approach is considered in the next section.

3 Bayesian Spatial Prediction

In Bayesian analysis, the parameters are thought as random variables, therefore we need to specify prior distributions for the unknown model parameters. Here, we consider proper priors which assure proper posteriors. We, also, assume that the vector θ to be independent of the parameters β and σ^2, so that the prior densities satisfy

$$\pi(\eta) = \pi(\beta, \sigma^2, \theta) = \pi(\beta|\sigma^2)\pi(\sigma^2)\pi(\theta)$$

Because of analytical convenience, we choose the following conjugate priors for β and σ^2

$$\pi(\beta|\sigma^2) \sim N_p(\beta_0, \sigma^2 V_0) \tag{12}$$

$$\pi(\sigma^2) \sim \chi^2_{Inv}(a, b) \qquad (i.e.\ \frac{ab}{\sigma^2} \sim \chi^2(a)) \tag{13}$$

also $\pi(\theta)$ is an arbitrary proper priors. Now we consider the prediction of $S(t_0)$ based on the Bayesian predictive distribution, defined by

$$\begin{aligned} f(s_o|s) &= \int_\Omega f(s_o, \eta|s)d\eta \\ &= \int_\Omega f(s_o|s, \eta)\pi(\eta|s)d\eta \end{aligned} \tag{14}$$

where $\Omega = R^p \times (0, \infty) \times \Theta$ and $\pi(\eta|s)$ is the joint posterior distribution of the model parameters. Since the analytical solution of the (p+q+1) dimension integration in (14) is very difficult, the Bayesian predictive distribution can be written as

$$f(s_o|s) = \int_\Theta f(s_o|s, \theta)\pi(\theta|s)d\theta \tag{15}$$

where Θ is the parameter space of θ. Khaledi and Mohammadzadeh (2004) showed that $(s_o|s, \theta)$ has a t distribution with $n+a$ degree of freedom denoted by $t_{n+a}(\mu_2, S_1^2\rho_2)$, where

$$\begin{aligned} \mu_2 &= (f'(t_0) - r'_\theta \Sigma_\theta^{-1} X)(V_0^{-1} + X'\Sigma_\theta^{-1} X)^{-1} V_0^{-1}\beta_0 \\ &\quad + (r'_\theta \Sigma_\theta^{-1} + (f'(t_0) - r'_\theta \Sigma_\theta^{-1} X)(\Sigma_0^{-1} + X'\Sigma_\theta^{-1} X)^{-1} X'\Sigma_\theta^{-1})s \\ S_1^2 &= \frac{ab + nS_2^2 + \beta_2' V_2 \beta_2 + \beta_0' V_0^{-1}\beta_0 - m'V_1 m}{a+n} \\ \rho_2 &= (1 - r'_\theta \Sigma_\theta^{-1} r_\theta) \\ &\quad + (f'(t_0) - r'_\theta \Sigma_\theta^{-1} X)(V_0^{-1} + X'\Sigma_\theta^{-1} X)^{-1}(f'(t_0) - r'_\theta \Sigma_\theta^{-1} X)' \end{aligned}$$

with

$$S_2^2 = \frac{1}{n}(s - X\beta_2)'\Sigma_\theta^{-1}(s - X\beta_2)$$
$$\beta_2 = (X'\Sigma_\theta^{-1}X)^{-1}X'\Sigma_\theta^{-1}s$$
$$V_2 = (X'\Sigma_\theta^{-1}X)^{-1}$$
$$m = (V_2^{-1}\beta_2 + V_0^{-1}\beta_0)$$

and the posterior distribution is proportion to

$$\pi(\theta|s) \propto \frac{f(s|\beta,\sigma^2,\theta)\pi(\beta|\sigma^2)\pi(\sigma^2)}{\pi(\beta|s,\sigma^2,\theta)\pi(\sigma^2|s,\theta)}\pi(\theta)$$
$$\propto h(\theta;s)\pi(\theta) \tag{16}$$

where

$$h(\theta;s) = |V_0^{-1} + X'\Sigma_\theta^{-1}X|^{-\frac{1}{2}}|\Sigma_\theta|^{-\frac{1}{2}}(S_1^2)^{-\frac{a+n}{2}} \tag{17}$$

Some of the improper priors proposed in the literature for spatial correlation structure, such as $\pi(\theta) = 1$ which is used in [11] give rise to improper posterior distributions [13]. For the vague proper priors proposed in [11] the spatial analysis may be extremely sensitive to the hyper parameters chosen for the vague prior. Therefore, we use a discretisation method by choosing a set of values $A = \{\theta_i\}_{i=1}^m$, with $\theta_i \in \Theta$ and assigning a discrete joint prior on A, say $\pi(\theta_i)$. Then we have

$$\pi(\theta_i|z) = \frac{h(\theta_i;s)\pi(\theta_i)}{\sum_{j=1}^m h(\theta_j;s)\pi(\theta_j)}$$

Next, we obtain

$$f(s_o|s) = \sum_{i=1}^m f(s_o|s,\theta_i)\pi(\theta_i|s)$$
$$= \frac{\sum_{i=1}^m f(s_o|s,\theta_i)h(\theta_i;s)\pi(\theta_i)}{\sum_{i=1}^m h(\theta_i;s)\pi(\theta_i)} \tag{18}$$

Now, using (18), we can obtain the Bayesian spatial predictor $\hat{S}(t_0) = E(S(t_0)|s)$ and variance prediction $Var(S(t_0)|s)$.
If instead of $S = (S(t_1), ..., S(t_n))$, the noisy measurements $Z = (Z(t_1), ..., Z(t_n))$ are taken at sample locations $t_1, \cdots, t_n$ such that

$$Z(t_i) = S(t_i) + \varepsilon(t_i) \qquad i = 1, ..., n$$

where $\varepsilon(\cdot)$, representing the noise, is a zero mean Gaussian white noise random field with $Var[\varepsilon(t)] = \sigma^2\alpha^2$, $\alpha^2 \geq 0$, and independent of $S(\cdot)$. In the conventional geostatistical terminology, the parameter $\tau^2 = \sigma^2\alpha^2$ is called nugget effect in the transformed scale. By the stated assumptions, we have

$$Z \sim N_n(X\beta, \sigma^2 V_\theta)$$

where $V_\phi = \Sigma_\theta + \alpha^2 I$ is a positive definite $n \times n$ matrix with $\phi = (\theta, \alpha^2)$ and the identity matrix I. The likelihood function of the model parameters $\eta = (\beta, \sigma^2, \phi)$ based on the positive observed data $z = (z(t_1), ..., z(t_n))$ is given by

$$L(\eta; z) = (\frac{1}{2\pi\sigma^2})^{n/2}|V_\phi|^{-1/2}\exp\{-\frac{1}{2\sigma^2}(z - X\beta)'V_\phi^{-1}(z - X\beta)\}$$

If the parameters are known, the optimal predictor and the spatial variance are respectively given by

$$\hat{S}(t_0) = E(S(t_0)|z, \eta) = f'(t_0)\beta + r'_\theta V_\phi^{-1}(z - X\beta),$$
$$Var(S(t_0)|z, \eta) = \sigma^2(1 + \alpha^2 - r'_\theta V_\phi^{-1} r_\theta).$$

When η is unknown, we assume that the vector ϕ to be independent of the parameters β and σ^2, and θ to be independent of α^2, i.e. the prior densities satisfy

$$\pi(\eta) = \pi(\beta, \sigma^2, \theta, \alpha^2) = \pi(\beta|\sigma^2)\pi(\sigma^2)\pi(\theta)\pi(\alpha^2)$$

If we choose the conjugate priors for β and σ^2 and a discrete joint prior for ϕ on $A = \{\phi_i\}_{i=1}^m \in \Theta \times [0, \infty)$, say $\pi(\phi_i)$, in [11] has Showed that

$$\pi(\phi_i|z) = \frac{h(\phi_i; z)\pi(\theta_i)\pi(\alpha_i^2)}{\sum_{j=1}^m h(\phi_j; z)\pi(\theta_j)\pi(\alpha_j^2)} \qquad i = 1, \cdots, n$$

where

$$h(\phi; z) = |V_1|^{\frac{1}{2}}|V_\phi|^{-\frac{1}{2}}(S_1^2)^{-\frac{a+n}{2}}$$

and $(s_o|z, \phi)$ has a t distribution with $n + a$ degree of freedom denoted by $t_{n+a}(\mu_2, S_2^2\rho_2)$. Then we obtain

$$f(s_o|z) = \sum_{i=1}^m f(s_o|z, \phi_i)\pi(\phi_i|z)$$
$$= \frac{\sum_{i=1}^m f(s_o|z, \phi_i)h(\phi_i; z)\pi(\theta_i)\pi(\alpha_i^2)}{\sum_{i=1}^m h(\phi_i; z)\pi(\theta_i)\pi(\alpha_i^2)} \quad (19)$$

Now, using (19), the Bayesian spatial predictor $\hat{S}(t_0) = E(S(t_0)|z)$ and variance prediction $Var(S(t_0)|z)$, can be obtained.

When data give evidence of non-Gaussian features, let $\{S(x), x \in D\}$ be a trans Gaussian random field of interest such that $\{Y_s(t) = g(S(t))\ ,\ t \in D\}$

is a (nearly) Gaussian random field with parametric mean and covariance functions, where transformation $g(\cdot)$ is monotone, differentiable and continues in R. Alternatively, it is assumed that the random field $S(\cdot)$ is not directly observable. Instead, the random vector $Z = (Z(t_1), ..., Z(t_n))$ represents the data measured at the sampling locations $t_1, ..., t_n \in D$ such that

$$g(Z(t_i)) = g(S(t_i)) + \varepsilon(t_i), \qquad i = 1, ..., n.$$

By the stated assumptions, we have

$$Y_z = g(Z) = (g(Z(t_1)), ..., g(Z(t_n))) \sim N_n(X\beta, \sigma^2 V_\phi).$$

The likelihood function of the model parameters $\eta = (\beta, \sigma^2, \phi)$ based on the observed data $z = (z(t_1), ..., z(t_n))$ is given by

$$L(\eta; z) = (\frac{1}{2\pi\sigma^2})^{n/2} |V_\phi|^{-1/2} \exp\{-\frac{1}{2\sigma^2}(g(z) - X\beta)' V_\phi^{-1} (g(z) - X\beta)\} J,$$

where $J = \prod_{i=1}^n |g'(z_i)|$ is the Jacobian of the transformation. Given the transformation g and the parameters η are known, the predictor distribution is a trans Normal distribution, $TN(\mu_1, \sigma^2\rho_1)$ with density function

$$f(s_o|z, \eta) = (\frac{1}{2\pi\sigma^2\rho_1})^{1/2} \exp\{-\frac{(g(s_0) - \mu_1)^2}{2\sigma^2\rho_1}\} |g'(s_0)|, \qquad (20)$$

where s_0 belongs to the range of $S(t_0)$. Now, the optimal predictor corresponding to the absolute error loss function, is given by

$$\hat{S}(t_0) = Median\ of\ (S(t_0)|z, \eta) = g^{-1}(\mu_1). \qquad (21)$$

To define a measure of uncertainty of $\hat{S}(t_0)$, let

$$L = g^{-1}(\mu_1 - 2\sigma\sqrt{\rho_1})$$
$$U = g^{-1}(\mu_1 + 2\sigma\sqrt{\rho_1})$$

where μ_1 and ρ_1 are given in (1), then (L, U) is an approximately 95% conditional prediction interval for $S(t_0)$. Therefore, we can use

$$\sigma(t_0) = \frac{U - L}{4},$$

as a measure of prediction uncertainty at location t_0. For example, if $S(\cdot)$ is a log Gaussian random field, then

$$\hat{S}(t_0) = e^{\mu_1},$$
$$\sigma(t_0) = \frac{\hat{S}(t_0)(e^{2\sigma\sqrt{\rho_1}} - e^{-2\sigma\sqrt{\rho_1}})}{4}.$$

Note that the measure of prediction uncertainty $\sigma(t_0)$ is large at any location that the prediction $\hat{S}(t_0)$ is large. When the normalizing transformation g is unknown, it is usually assumed that g belongs to a parametric family of transformations $G = \{g_\lambda, \lambda \in \Lambda\}$. Then, η and λ are unknown model parameters. In a special case, [4] determined the maximum likelihood estimates of the model parameters, and proposed the plug-in approach for spatial prediction, in which, the estimates of the parameters are replaced in (21). Due to the fact that in the Bayesian approach, when a logarithm transformation can be successful in representing this type of data, [11] considered a Bayesian Log-Gaussian model for spatial prediction. While a logarithm transformation may not be appropriate, we would propose an empirical Bayes approach by estimating the normalizing transformation and then, treating this transformation as known for Bayesian spatial prediction. In this case, it can be shown that

$$\pi(\phi_i|z) = \frac{h(\phi_i;z)\pi(\theta_i)\pi(\alpha_i^2)}{\sum_{j=1}^m h(\phi_j;z)\pi(\theta_j)\pi(\alpha_j^2)},$$

and $(s_o|z,\phi)$ has a trans t distribution with $n+a$ degree of freedom denoted by $TT_{n+a}(\mu_2, S_1^2\rho_2)$, whose density function is given by

$$f(s_o|z,\phi) = \frac{\Gamma(\frac{n+a+1}{2})|g'(s_0)|}{\Gamma(\frac{n+a}{2})(\pi S_1^2\rho_2)^{1/2}}[1+\frac{(g(s_0)-\mu_2)^2}{S_1^2\rho_2}]^{-\frac{n+a+1}{2}}.$$

Next, based on the discretisation method, the Bayesian predictive distribution can be obtained as

$$\begin{aligned} f(s_o|z) &= \sum_{i=1}^m f(s_o|z,\phi_i)\pi(\phi_i|z) \\ &= \frac{\sum_{i=1}^m f(s_o|z,\phi_i)h(\phi_i;z)\pi(\theta_i)\pi(\alpha_i^2)}{\sum_{i=1}^m h(\phi_i;z)\pi(\theta_i)\pi(\alpha_i^2)} \end{aligned} \tag{22}$$

Since the exponential of a log t distribution does not have finite moment, the minimum square error predictor does not exist. To circumvent this, we consider the absolute error loss function. Now, using (22), the Bayesian predictive distribution, we can obtain the Bayesian spatial predictor of $S(t_0)$ as

$$\hat{S}(t_0) = Median\ of\ (S(t_0)|z) \tag{23}$$

Because the second moment of the Bayesian predictive distribution may be infinite, we can use

$$\sigma(t_0) = \frac{u_{0.975}(t_0) - u_{0.025}(t_0)}{4} \tag{24}$$

as a measure of prediction uncertainty at location t_0, where $u_\alpha(t_0)$ is the 100αth percentile of the predictive distribution and the numerator of (24) is the length of a 95% conditional prediction interval for $S(t_0)$.
The choice of a correlation function family for the transformed random field is often rather arbitrary. Sometimes, the cross-validation criterion is used. But, this method is not applicable for noisy spatial data because there is no exact observation of $S(\cdot)$. Therefore, we address this problem using a model selection procedure based on observations. For the considered discrete prior, the marginal density of observations for an arbitrary model M^*, corresponding to a correlation function ρ^*, is proportion to

$$\begin{aligned} m^{(*)}(z) &= \sum f(z|\phi_i, M^*)\pi(\phi_i) \\ &\propto \sum_i h^{(*)}(\phi_i, z)\pi(\phi_i) \end{aligned}$$

The comparison of r different model M_k, $k = 1, \cdots, r$, typically can be done via computing the posterior probabilities given by

$$\begin{aligned} P_r(M_k|z) &= \frac{m_k(z)}{\sum_{j=1}^r m_j(z)} \\ &= \frac{\sum_i h^{(k)}(\phi_i, z)\pi(\phi_i)}{\sum_{j=1}^r \sum_i h^{(j)}(\phi_i, z)\pi(\phi_i)} \end{aligned}$$

Now, a model with the largest posterior probabilities would be chosen as the best model in terms of posterior probability criterion.

Table 1. Simulated data for three spatial models.

i	t_i	$s_1(t_i)$	$z_1(t_i)$	$z_2(t_i)$
1	(0.5,0.5)	-1.24	-1.37	0.254
2	(0.5,1)	-1.17	-1.02	0.361
3	(0.5,1.5)	-0.78	-0.69	0.502
4	(1,0.5)	-0.62	-0.67	0.512
5	(1,1)	-0.12	0.02	1.02
6	(1,2)	-0.06	-0.09	0.914
7	(1.5,0.5)	0.04	0.15	1.161
8	(1.5,1)	0.24	0.21	1.234
9	(1.5,1.5)	-0.27	-0.35	0.705
10	(2,0)	-0.1	-0.06	0.942

Table 2. Cross validated MSE for three spatial model in two spatial prediction methods.

Method	Gaussian	Noisy Gaussian	Noisy Trans Gaussian
Plug-in	0.28465	.041762	0.48632
Bayesian	0.09339	0.1721	0.22359

4 Simulation Study

In order to check the adequacy of Bayesian spatial prediction, we perform a simulation study. For $p = 1$, $\beta = 0$, $\sigma^2 = 1$ and $\rho(h, \theta) = (\frac{1}{2})^{|h|}$, a Gaussian, a noisy Gaussian with $\tau^2 = 0.05$ and a log Gaussian random field are generated at $n = 10$ locations in the region $D = [0, 2] \times [0, 2]$. Table 1 shows the simulated data, where the third column denotes locations, the second column denotes the Gaussian simulated data, the forth column denotes the noisy Gaussian simulated data and the fifth column denotes the noisy log-Gaussian simulated data. To compare the accuracy of two spatial prediction methods, a cross-validation criterion is used based on single point deletion predictive distributions as described by Gelfand *et.al* (1992). For each entertained model, the mean square error of Bayesian and plug-in predictor are computed and presented in table 2. For all the cases, the cross validated MSE of the Bayesian predictor are clearly lower than those of plug-in predictor, the Bayesian method is more accurate than plug-in prediction.

5 Numerical Example

To illustrate the proposed spatial prediction method in section 3, we applied it on a rainfall data set. Table 3 shows the rainfall at 22 sites in a region with 192×160 squared kilometers in north of Iran on the last month of winter 2002. The histogram of the observations shows the skewness of the data distribution. So, we assume that the normalizing transformation g belongs to the Box-Cox family given by

$$g_\lambda(u) = \begin{cases} \frac{u^\lambda - 1}{\lambda} & if \lambda \neq 0 \\ \log(u) & if \lambda = 0 \end{cases}$$

which offers a great deal of flexibility in normalizing positive data. The value of λ, using trial and error method, can be chosen such that the skewness and kurtosis of the transformed data be near 0 and 3, respectively. For the obtained value $\hat{\lambda} = -0.3$ using this method, the Bayesian spatial prediction of rainfall at any given site t_0 is considered. We assume that the data are

Table 3. Rainfall at 20 stations in north of Iran.

t_i	$z(t_i)$	t_i	$z(t_i)$	t_i	$z(t_i)$	t_i	$z(t_i)$
(1 , 1.5)	18.24	(2 , 6)	29.95	(4 ,8)	32.08	(8 , 2)	41.75
(1 , 5)	21.17	(3, 1)	17.72	(11 , 1)	46.06	(8 , 6)	33.98
(1 ,7.5)	33.1	(3 ,9)	35.27	(6 , 1)	24.35	(9 , 3)	57.4
(2 ,2.5)	19.31	(4 , 2)	24.74	(6 , 5)	38.24	(10 ,6)	23.6
(2 ,5)	22.82	(4 , 7.5)	31.14	(7 , 9)	23.69		
(11 ,5)	26.83	(6.5,7)	26.76	(5 , 3.5)	24.35		

Table 4. The posterior probability for three models.

	M_1 $(\nu = 0.5)$	M_2 $(\nu = 1)$	M_3 $(\nu = 2)$
$P_r(M_i\|z)$	0.5271	0.1894	0.2835

realizations of a noisy g-Gaussian model with fixed mean $E[Y_s(t)] = \beta^*$ and the Matérn isotropic correlation function (Matérn, 1986)

$$\rho(u; \theta, \nu) = \{2^{\nu-1}\Gamma(\nu)\}^{-1}(u/\theta)^{\nu} K_{\nu}(u/\theta), \quad \theta > 0$$

where $K_{\nu}(\cdot)$ denotes the modified Bessel function of order ν. The parameter ν controls the mean square differentiability of the transformed random field. For $\nu = 0.5$, the Matérn family reduces to the exponential model

$$\rho(u; \theta) = \exp^{-\frac{u}{\theta}}$$

For further details on the Matérn family, Stein (1999) is a desire reference. Three families corresponding to Matérn correlation function $\nu = 0.5$, $\nu = 1$ and $\nu = 2$ are considered. Since the hyper parameters β_0, V_0, a and b of the prior distributions (12) and (13) are unknown, we use the limit prior distributions (as V_0^{-1} and a tend to zero in (3) and (4))

$$(\beta|z, \sigma^2, \phi) \sim N_p(\beta_2, \sigma^2 V_2)$$
$$(\sigma^2|z, \phi) \sim \chi^2_{Inv}(n-1, S_2^2)$$

where $V_2 = \frac{1}{\mathbf{1}' V_\phi^{-1} \mathbf{1}}$, $\beta_2 = V_2(\mathbf{1}' V_\phi^{-1} g(z))$ and $S_2^2 = \frac{1}{n}(g(z) - \beta_2 \mathbf{1})' V_\phi^{-1}(g(z) - \beta_2 \mathbf{1})$. For this case (17) becomes

$$h(\phi; z) = |V_2|^{\frac{1}{2}} |V_\phi|^{-\frac{1}{2}} (S_2^2)^{-\frac{n-1}{2}}$$

Now, we generate $m = 1500$ independent random values for $\phi = (\theta, \alpha^2)$ in $(0, 100) \times [0, 0.5)$. For Uniform prior distributions $\pi(\theta_i) = \pi(\alpha_i^2) = \frac{1}{m}$, $i =$

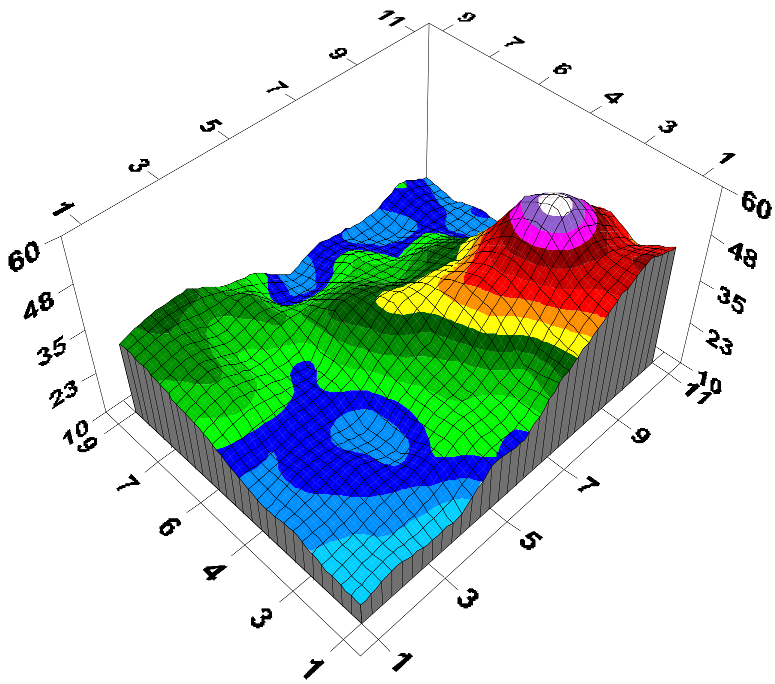

Fig. 1. The map of the Bayesian spatial predictor

$1, \ldots, m$, the posterior probability of each entertained model is determined and presented in table 4. The data clearly support exponential covariance function. Based on this model, $f(s_o|z, \phi)$ is a trans t distribution

$$TT_{n-1}(\mu_3, \frac{n}{n-1} S_2^2 \rho_3)$$

where

$$\mu_3 = (r'_\theta V_\phi^{-1} + \frac{(1 - r'_\theta V_\phi^{-1} \mathbf{1}) \mathbf{1}' V_\phi^{-1}}{\mathbf{1}' V_\phi^{-1} \mathbf{1}}) g(z)$$

$$\rho_3 = (1 - r'_\theta V_\phi^{-1} r_\theta) + \frac{(1 - r'_\theta V_\phi^{-1} \mathbf{1})^2}{\mathbf{1}' V_\phi^{-1} \mathbf{1}}.$$

Now the Bayesian predictive distribution is simplified as

$$f(s_o|z) = \frac{\sum_{i=1}^m f(s_o|z, \phi_i) h(\phi_i; z)}{\sum_{i=1}^m h(\phi_i; z)}.$$

Figure 1 shows the map of the Bayesian spatial predictor, derived by computing $\hat{S}(t_0)$ given by (23). According to figure 1, the rainfall predictions are high in the south east region, low in the south west and north east regions, and moderate in other regions.

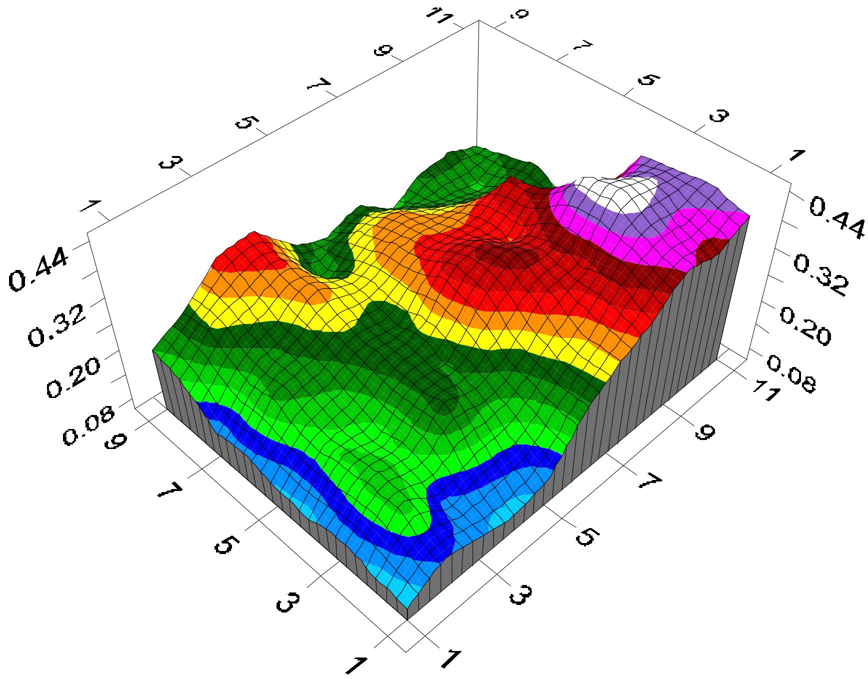

Fig. 2. Map of the measure of prediction uncertainty

Figure 2 shows the measure of prediction uncertainty measure, obtained by computing of $\sigma(t_0)$, in (24). It is observable that the measure of prediction uncertainty $\sigma(t_0)$ is high at locations with big $\hat{S}(t_0)$. This can be a consequent of using the trans Gaussian model, as described in section 3.

6 Discussion

When a normalizing transformation is desirable in representing the spatial data that may contain noise, this work provides a Bayesian model for spatial prediction. Furthermore, this approach is avoided about problems related to the plug in method. As simulation study shows, the Bayesian approach provides a more accurate spatial prediction than plug-in method. In addition, it can be used to determine the Bayesian predictive distribution, deriving the predictive uncertainty measure of predictor at any given site based on the absolute error loss function.

References

1. Banerjee, S., Carlin,B. P. and Gelfand, A. E. (2004). Hierchical Modelling and Analysis for Spatial Data. Chapman and Hall.

2. Berger, J. O., De Oliveira, V., and Sanso, B. (2001). Objective Bayesian analysis of spatially correlated data. *Journal of the American Statistical Association* **96**: 1361-1374.

3. Chiles, J. P. and Delfiner, P.(1999). Geostatistics; Modelling Spatial Uncertainty. John Weily, New York.

4. Christensen, O. F., Diggle, P. J. and Ribeiro Jr, P. J. (2001). Analysing Positive-Valued Spatial Data: the Transformed Gaussian Model. *Geostatistics for Environmental Applications.* **11**: 287-298.

5. Cressie, N. (1993). Statistics for Spatial Data. John Weily, New York.

6. De Oliveira, V. Fokianos and K. Kedem, B. (2003). Bayesian Transformed Gaussian Random Field : a Review. *Japanese Journal of Applied Statistics.* (to appear).

7. De Oliveira, V. and Ecker, M. D. (2002). Bayesian Hot Spot Detection in the presence of Spatial Trend: Application to totorial Nitrogen Cocentration in the Chesapeake Bay. *Environmetrics* . **13**: 85-101.

8. Diggle, P. J. and Ribeiro Jr, P. J. (2003). Bayesian Inference in Gaussian Model Based Geostatistics. *Geographical and Environmental Modelling.* (to appear).

9. Diggle, P. J., Tawn, J. A. and Moyeed, R. A. (1998). Model Based Geostatistics. *Applied Statistics.* **47**: 299-350.

10. Gaudard, M., Karson, M., Linder, E. Sinha, D. (1999). Bayesian Spatial Prediction. *Environmental and Ecological Statistics.* **6**: 147-171.

11. Handcock, M. S. and Stein, M. L. (1993). A Bayesian Analysis of Kriging. *Technometrics.* **35**: 403-410.

12. Mardia, K. V. and Watkins, A. J. (1989). On Multimodality of the Likelihood in the Spatial Linear Model. *Biometrika.* **76**: 289-295.

13. Mohammadzadeh, M., and Jafari Khaledi, M., (2004), Baysian Prediction for a Noisy Log-Gaussian Spatial Model, COMPSTAT, Proceedings in Computational Statistics, Prauge, Czeck Repoblic. 1511-1518, Edited by Jaromir Antoch, Physica-Verlag Heidelberg.

14. Ripley, B. P. (1988). Stochastic Simulation. John Weily, New York.

15. Todini, E. (2001). Influence of Parameter Estimation Uncertainty in Kriging. *Hydrology and Earth System Sciences.* **5**: 215-223.

16. Warnes, J. J. and Ripley, B. P. (1987). Problems with Likelihood Estimation of Covariance Function of Spatial Gaussian Process. *Biometrika.* **74**: 640-642.

17. Zimmerman, D. L. and Cressie, N. (1992). Mean Square Error Prediction in Spatial Linear Model with Estimated Covariance Parameters. *Annals of the Institute of Statistical Mathematics.* **44**: 27-43.