Environmental and Ecological Statistics **4**, 1–29 (1997)

Estimating the area affected by phosphorus runoff in an Everglades wetland: a comparison of universal kriging and Bayesian kriging

SONG S. QIAN

Environmental Sciences and Resources, Portland State University, Portland, OR 97207–0751, USA

Received October 1995. Revised July 1996

Phosphorus-enriched agriculture runoff is believed to be the leading cause of ecosystem changes of Everglades wetlands. To study this effect, it is necessary to estimate the area of the affected region. In this study, Bayesian kriging and universal kriging were used to estimate the area by analysing the data collected by Reddy *et al*. (1991). The background level of the soil's total phosphorus concentration is used to determine whether the region is affected by the agriculture runoff, through an indicator function. The area of the affected region was represented by the integration of the indicator function over the entire wetland. The expected value of the affected area was calculated using the results derived from Bayesian and universal kriging. The outcome indicates that universal kriging is sensitive to specification of the covariance model. It was observed that universal kriging and Bayesian kriging yield comparable results, if the specified covariance structures are of similar nature.

Keywords: geostatistics, nutrients, soils, water quality

1. Introduction

Historically, the Water Conservation Area – 2A (WCA2A) is part of the Everglades wetlands in south Florida, USA (Fig. 1). It receives agricultural runoff from the Everglades Agriculture Area. Phosphorus-enriched agriculture runoff caused some significant changes in this phosphorus-limited wetland ecosystem. The most obvious change was the conversion of cattail (*Typha clomingensis*) as the dominant species, near inlets of the runoff, instead of sawgrass (*Cladium jamaicense*). The significant increase of the phosphorus level in the water of WCA2A is considered a major threat to the Everglades National Park. In order to protect the park, the South Florida Water Management District has proposed the use of constructed wetlands, marsh-like buffer areas, to remove excess phosphorus before the water enters WCA2A.

Constructed wetlands have been used for wastewater treatment since the early 1950s. However, most constructed wetlands are designed to remove organic pollutants rather than phosphorus. Studies show that constructed wetlands are not always effective for phosphorus removal (Howard-Williams, 1985).

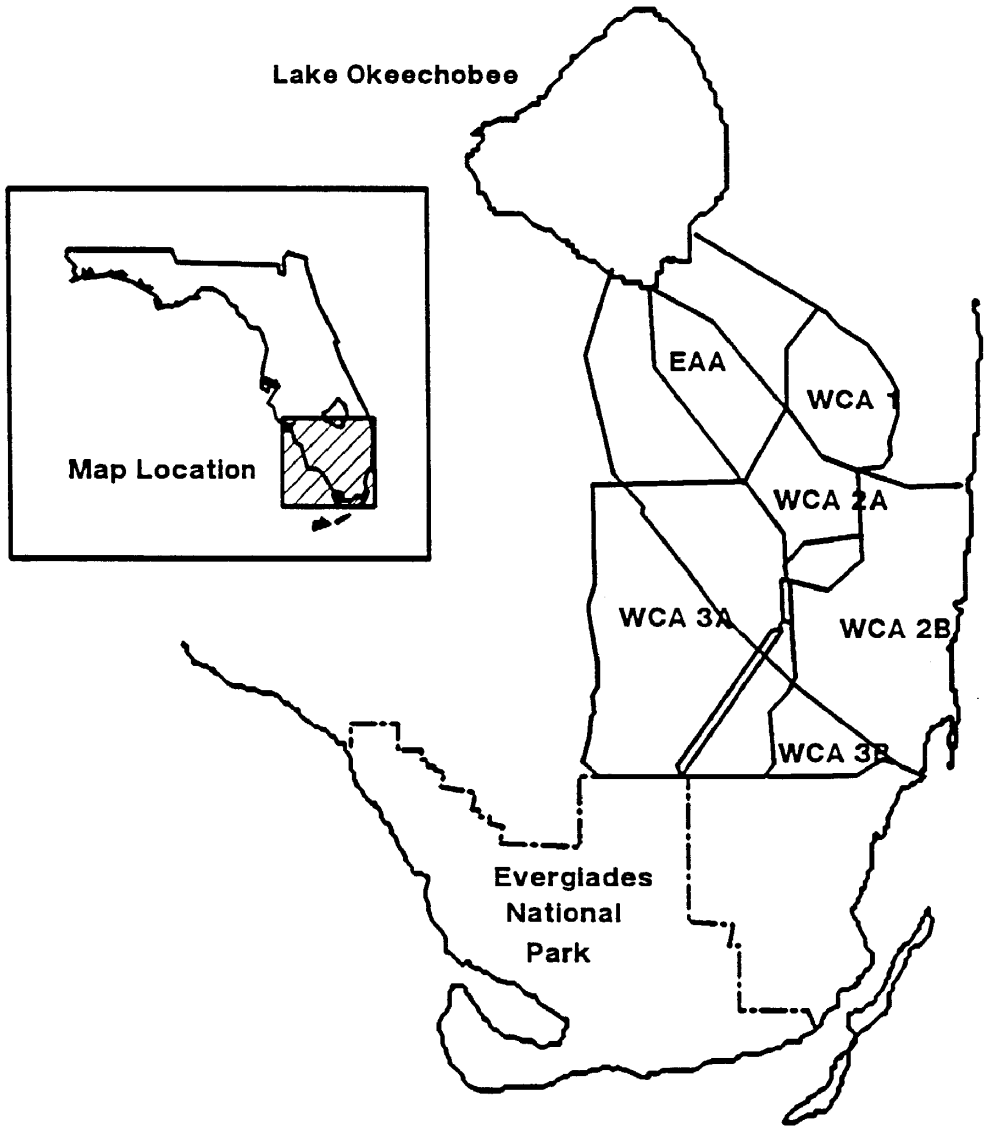

Fig. 1. Map of south Florida and location of the study region.

Estimating the area of the phosphorus-affected region in WCA2A can help us in understanding how the Everglades wetlands respond to added phosphorus. Many studies (such as that by Reckhow and Qian, 1994) have found that the unit area phosphorus mass loading rate (in mass of phosphorus per unit area per unit time) is an important parameter in modelling phosphorus in wetlands and lakes. The unit mass loading rate can be used to determine the size of a constructed wetland. The unit area loading rate is usually calculated by measuring the influent mass loading rate (in mass of phosphorus per unit time) divided by the area of the receiving wetland. Because WCA2A covers such a large region, not all parts are effective in removing phosphorus carried by the agricultural runoff. (In other words, not all parts are

affected.) If the entire WCA2A were taken as affected, we might underestimate the phosphorus-assimilating capacity of the wetland.

In this paper, the area of the affected region was calculated by using universal kriging (Cressie, 1991) and Bayesian kriging (Handcock and Stein, 1993). Some relevant references for the basic theory of kriging are Matheron (1965), Journel and Huijbregts (1978), and Ripley (1981). From a Bayesian perspective, the difference between universal kriging and Bayesian kriging lies in the methods of estimating the unknown spatial correlation structure. To shed light on how to select the appropriate approach to this type of study, results of both universal and Bayesian kriging are compared.

Applications of spatial statistics (or geostatistics) in environmental related areas are abundant. For example, Verly *et al*. (1984) in natural resources; Brus and de Gruijter (1993), Laslett *et al*. (1987), and Burgess *et al*. (1981) in soil science; Ahmed and Marsily (1987) and Black and Freyberg (1987) in groundwater modelling; Cox *et al*. (1995) and Bogardi *et al*. (1985) in environmental spatial sampling design. Cressie and ver Hoef (1993) presented an overview of environmental and ecological applications of spatial statistical analysis.

A brief review of universal and Bayesian kriging is first presented in Section 2, with special emphasis on the estimation of the spatial correlation structure. Section 3 presents the data used in this paper and the method used for calculating the area of the affected region. The results are presented in Section 4. The paper concludes with discussions on other estimations of the affected area, additional sources of uncertainty of the methods, and future work.

2. Universal and Bayesian kriging

Suppose we are interested in modelling a real-valued Gaussian random field $Z(x)$ using the following model:

$$Z = X\beta + \eta \tag{1}$$

where X is the known design matrix containing the spatial coordinates and other predictor variables, β is a vector of unknown regression coefficients, and η is the error term. η is assumed to be second-order stationary and normally distributed with mean 0 and covariance Σ. The covariance function is represented by $\text{cov}\{\eta(s_1), \eta(s_2)\} = \alpha K_\theta(s_1, s_2)$ for any pair of spatial coordinate points (s_1, s_2) of interest, where $\alpha > 0$ is a scale parameter, and θ is a vector of structural parameters which specify the shape of the covariance function. In the case of kriging, the objective is to predict $Z(s_0)$ for a location with no observed data (s_0) based on observations: $Z = \{Z(s_1), Z(s_2), \ldots, Z(s_n)\}'$.

The universal kriging predictor, of the form $\hat{Z}_\theta(s_0) = \lambda(\theta)'Z$, is the best linear unbiased predictor (BLUP, Ripley, 1981), where $\lambda(\theta)' = b_\theta'(X'K_\theta^{-1}X)^{-1}X'K_\theta^{-1} + k_\theta'K_\theta^{-1}$, K_θ is the correlation matrix of the model residuals, k_θ is the vector of correlations of the residuals between site s_0 and sites s_i $(i = 1, \ldots, n)$, $b_\theta = X(s_0) - X'K_\theta^{-1}k_\theta$, and $X(s_0)$ is the vector of predictor variables for s_0. Under the normality assumption, the distribution of the prediction error $e(s_0)$ is $\text{N}(0, \alpha V_\theta)$, and $V_\theta = K_\theta(s_0, s_0) - k_\theta'K_\theta^{-1}k_\theta + b_\theta'(X'K_\theta^{-1}X)^{-1}b_\theta$. In other words, the distribution of $Z(s_0)$ conditional on α, θ, and Z is a normal distribution with mean of $\hat{Z}_\theta(s_0)$, and variance of αV_θ:

$$Z(s_0) \mid \alpha, \theta, Z \sim \text{N}(\hat{Z}_\theta(s_0), \alpha V_\theta)\,. \tag{2}$$

The covariance matrix αK_θ is usually unknown. It is often estimated by using the variogram, $\mathrm{var}\{\eta(s_1) - \eta(s_2)\}$, which is a function of h (the distance between the two sites) under the assumption of second-order stationary and isotropic, i.e.

$$\mathrm{var}\,\{\eta(s_1) - \eta(s_2)\} = 2\,\gamma(h)$$

The empirical variogram, $2\hat{\gamma}(h)$, is a non-parametric estimator of $2\gamma(h)$, given by:

$$2\hat{\gamma}(h) = \frac{\sum_{S(h)}[\eta(s_i) - \eta(s_j)]^2}{\mathrm{N}(h)}$$

where $S(h) = \{(s_i, s_j): |s_i - s_j| = h\}$ is the set of all pairs of data points separated by distance h, and $\mathrm{N}(h)$ is the number of distinct pairs in $S(h)$.

The covariance function is usually estimated by fitting the empirical variograms as a function of distance to a specific parametric form (or variogram model). The three commonly used variogram models are:

1. Spherical with nugget:

$$\gamma(h) = \begin{cases} 0 & \text{if } h = 0 \\ c_0 + c_s\{\tfrac{3}{2}(h/a_s) - \tfrac{1}{2}(h/a_s)^3\} & \text{if } 0 < h \leqslant a_s \\ c_0 + c_s & \text{if } h > a_s \end{cases}$$

2. Exponential with nugget:

$$\gamma(h) = \begin{cases} 0 & \text{if } h = 0 \\ c_0 + c_e\,(1 - \mathrm{e}^{-h/a_e}) & \text{if } h > 0 \end{cases}$$

3. Gaussian with nugget:

$$\gamma(h) = \begin{cases} 0 & \text{if } h = 0 \\ c_0 + c_g\,(1 - \mathrm{e}^{-h^2/a_g^2}) & \text{if } h > 0 \end{cases}$$

Under a second-order stationary assumption, the variogram is:

$$2\gamma(h) = 2(C(0) - C(h))$$

where $C(h) = \mathrm{cov}\{\eta(s + h), \eta(s)\}$, and $C(0) = \lim_{h\to\infty}\gamma(h)$ when the three variogram models above are used.

As one referee pointed out, the empirical variogram only estimates values of the variograms which is not the same as estimating the variogram function; if the residuals are obtained by fitting model (1) with a trend surface, then a bias is induced. The optimal kriging predictor, $\lambda(\theta)'Z$, is a kriging estimator which requires knowing the variogram. However, there are advantages in estimating the variogram instead of the covariance function directly. One advantage is that one need not separately estimate the (constant) mean; a second reason is that the empirical covariance will not give evidence of a non-stationarity.

Selecting a variogram model from the above-mentioned three candidate models will certainly introduce uncertainty to the optimal kriging predictor, in addition to the uncertainty of estimating parameters using empirical variograms. Many studies have addressed the effect of using

an incorrect covariance functions. For example, Warnes (1986) studied the sensitivity of the fitted surface to perturbations in the covariance model. He found that the fitted surface is not sensitive to the perturbation of covariance model parameter when the exponential model is used, and is sensitive when the Gaussian model is used. Diamond and Armstrong (1984) studied the robustness of variograms by defining the 'closeness' of two variograms. They found that minor perturbations in the data may give rise to major effects in the final kriged values, and a 'robust' procedure for the estimation of experimental variogram is not a guarantee of the robustness of the entire predictive algorithm. Stein (1988) and Stein and Handcock (1989) showed that the impact on the kriging predictor from using the incorrect covariance function is asymptotically negligible, as long as the estimated variogram function is 'compatible' to the true variogram. Yakowitz and Szidarovszky (1985) discussed some of the considerations of variogram model selection, and they suggested that one should not select a variogram entirely algorithmically, but should pay attention to past experience with similar geostatistical data. A method was proposed by Zimmerman and Cressie (1992) for accounting for the uncertainty in the estimated covariance parameters. Brooker (1986) studied the robustness of the spherical variogram. However, in many environmental applications, once the variogram model is chosen, the covariance function is taken as known, and the uncertainty in the selected model is often ignored.

To include this uncertainty in the analysis, Handcock and Stein (1993) introduced Bayesian kriging using the Matérn class of covariance functions, characterized by a parameter $\theta = (\theta_1, \theta_2)$. $\theta_1 > 0$ is a scale parameter controlling the spatial range of correlation and $\theta_2 > 0$ is a smoothness parameter controlling the smoothness of the random field. Many commonly used covariance functions are special cases of the Matérn class (*e.g.*, $\theta_2 = 0.5$ corresponds to the exponential model and $\theta_2 \to \infty$ represents the Gaussian model). The general form of K_θ for the Matérn class is

$$K_\theta(h) = \frac{1}{2^{\theta_2 - 1}\Gamma(\theta_2)} \left(\frac{h}{\theta'_1}\right)^{\theta_2} \mathsf{K}_{\theta_2}\left(\frac{h}{\theta'_1}\right)$$

where h is the distance between two points, $\theta_1' = \theta_1(2\sqrt{\theta_2})$, and K_{θ_2} is the modified Bessel function of order θ_2 discussed by Abramowitz and Stegun (1964). Handcock and Stein (1993) and Handcock and Wallis (1994) have demonstrated its flexibility in handling a variety of spatial data sets.

The posterior predictive distribution for $Z(s_0)$ is derived by adapting a non-informative prior:

$$\pi(\alpha, \beta, \theta) \propto \frac{\pi(\theta)}{\alpha} \tag{3}$$

where π represents a probability density function.

Combining Equations (2) and (3), Handcock and Stein (1993) showed that the conditional posterior distribution of $Z(s_0)$ is

$$Z(s_0) | \theta, Z \sim t_{n-q}\left(\hat{Z}_\theta(s_0), \frac{n}{n-q}\hat{\alpha}(\theta)V_0(\theta)\right) \tag{4}$$

where q is the number of regression coefficients in β.

The marginal posterior distribution of θ is

$$\pi(\theta \mid Z) \propto \pi(\theta) \mid K_\theta \mid^{-1/2} \mid X'K_\theta^{-1}X \mid^{-1/2} \hat{\alpha}(\theta)^{-(n-q)/2} \tag{5}$$

Therefore, the Bayesian predictive distribution for $Z(s_0)$ is

$$\pi(Z(s_0) \mid Z) \propto \int_\theta \pi(Z(s_0) \mid \theta, Z) \times \pi(\theta \mid Z)\, \mathrm{d}\theta \tag{6}$$

where $\hat{Z}_\theta(s_0)$ is the usual kriging point predictor,

$$\hat{\alpha}(\theta) = \frac{1}{n-q}(Z - X\hat{\beta})' K_\theta^{-1}(Z - X\hat{\beta}), \quad \alpha V_0(\theta)$$

is the usual kriging prediction error variance, and t_{n-q} is a Student's t distribution with $(n-q)$ degrees of freedom. See Handcock and Stein (1993) for details.

Comparing expressions (2) and (6), it is easy to see that the difference between universal and Bayesian kriging lies in the method used for estimating the unknown covariance function, parametrized by α and θ. In universal kriging, when the covariance function is unknown, it is approximated by a specific variogram model and the parameters of the selected model are estimated from the empirical variograms; in Bayesian kriging, the unknown covariance function is defined through the joint posterior distribution of the parameters of the Matérn class of covariance function (expression 5). Theoretically, Bayesian kriging is 'better' since the uncertainty of the covariance function is considered. However, because of a significantly increased computation intensity, it is necessary to justify the use of Bayesian kriging over universal kriging in the light of empirical evidence of improved performance.

3. WCA2A data and the area estimation

The data used in this paper were from Reddy *et al.* (1991), who collected soil core samples at 74 stations on seven north–south transects, spaced at 2 mile (3.22 km) intervals (Fig. 2). Soil core samples were sectioned in the laboratory into four increments of 10 cm in depth, and 16 different parameters were measured for each increment. The soil's total phosphorus (STP) content, one of the 16 parameters measured by the Reddy study, was used in this paper. Dr C.J. Richardson (personal communication, 1993), a leading expert in Everglades wetlands, believes that the STP content would be fairly stable in the Everglades if there were no agricultural runoff problem, and the background level of the STP content is about 500 μg of phosphorus per gram of soil (μg/g) in the top 20 cm.

Accordingly, the mean STP contents in the first two increments (top 20 cm layer) of the Reddy samples were used. The STP values were log-transformed to stabilize the variance. The spatial coordinates were converted from latitude and longitude to the Universal Transverse Mercator (UTM) grid system; therefore, the distances calculated from the data are in metres. Figure 3 shows the data. The data used in this paper are available from the author upon request (send e-mail to: bwsq@odin.cc.pdx.edu).

Bayesian and universal kriging were used to predict the phosphorus concentration over the region. When the concentrations are larger than the background level mentioned above, the corresponding regions were assumed to be affected by the agriculture runoff.

The area of the affected region A_p can be represented by

$$A_\mathrm{p} = \int I(Z(X) > t \mid Z)\, \mathrm{d}X \tag{7}$$

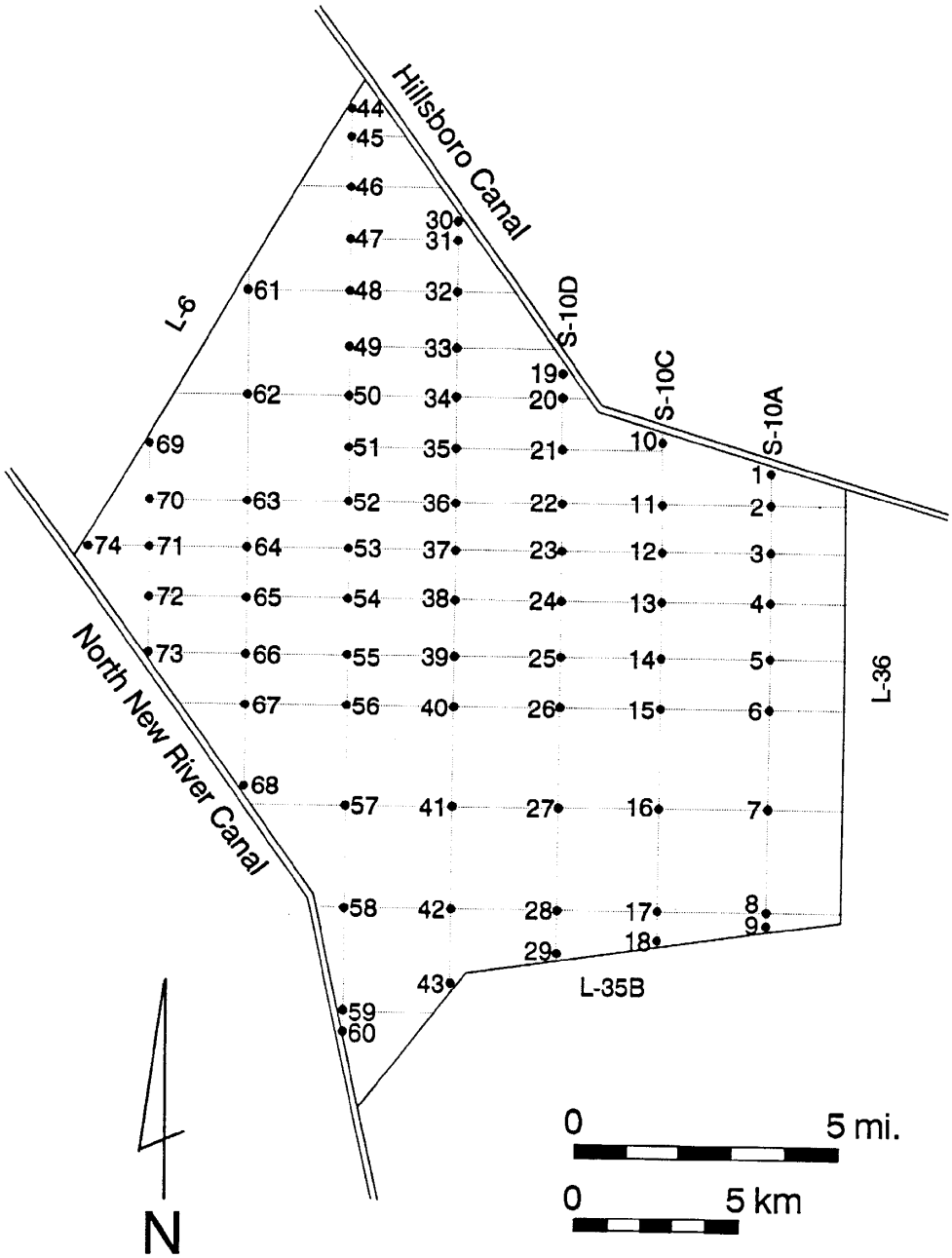

Fig. 2. Locations of sampling sites in WCA2A.

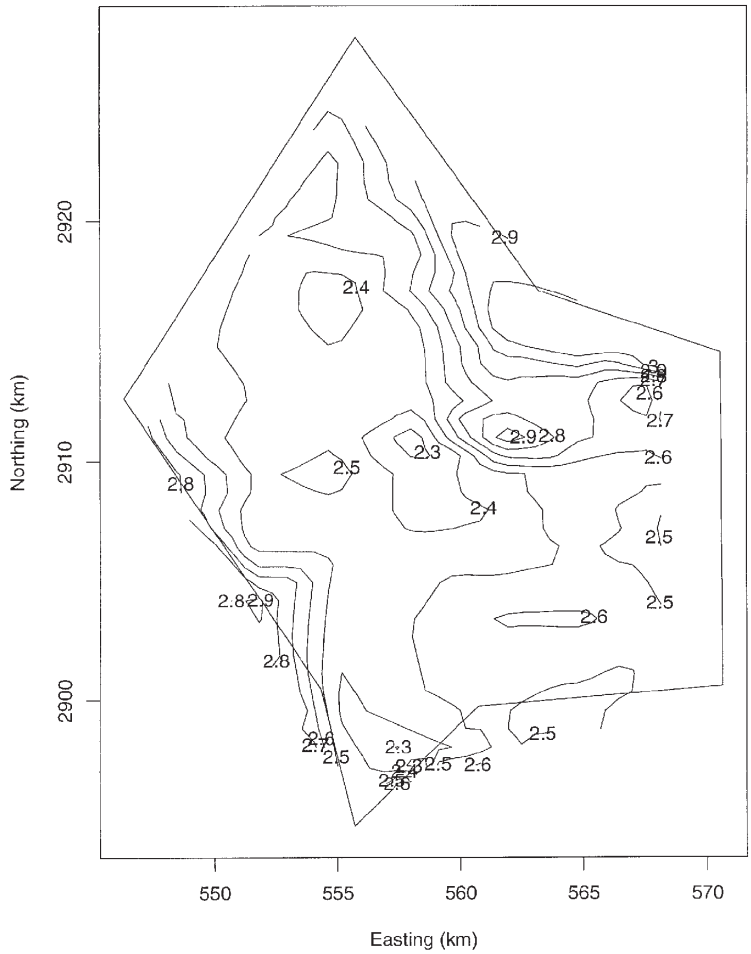

Fig. 3. The log-transformed data.

where $I(\cdot)$ is an indicator function, X is the coordinate vector, and $t = \log_{10}(500)$ is the background level of STP content.

Within a Bayesian framework, A_p is evaluated as a random variable. Therefore, the expected value of the area is estimated:

$$\mathrm{E}(A_p) = \int\left[\int I(Z(X) > t)\,\mathrm{d}X\right] \times \pi(Z(X) \mid Z)\,\mathrm{d}Z = \int \mathrm{prob}(Z(X) > t \mid Z)\,\mathrm{d}X \quad (8)$$

where prob represents probability.

When universal kriging is used, the integrand in the right-hand side (RHS) of Equation (8) is the upper tail of the predictive distribution, and it can be evaluated using the CDF of a normal density defined in expression (2).

When Bayesian kriging is used,

$$\mathrm{E}(A_p) = \int_{X \epsilon R^2}\int_{\theta} \mathrm{prob}(Z(X) > t \mid \theta, Z) \times \pi(\theta \mid Z)\,\mathrm{d}\theta\,\mathrm{d}X \quad (9)$$

Note that the integrand in the RHS of Equation (9) is a product of two density functions: the first one, $\text{prob}(Z(X) > t \mid \theta, Z)$, is the upper tail of a t-distribution, and the second one, $\pi(\theta \mid Z)$, is not a function of X. Therefore, switching the order of integration will improve the speed of computing. (In fact, it took about 20 hours to integrate Equation (9) on a DEC5000 workstation, and about 40 minutes to integrate Equation (10), using the same Gaussian quadrature program written in C.) This switch results in

$$\mathrm{E}(A_{\mathrm{p}}) = \int_{\theta} \pi(\theta \mid Z) \left\{ \int_{X \epsilon R^2} \text{prob}(Z(X) > t \mid \theta, Z)\, \mathrm{d}X \right\} \mathrm{d}\theta \tag{10}$$

The uncertainty involved in the estimation of the affected area using Bayesian kriging is evaluated by using a Monte Carlo simulation method. The procedure is based on the fact that given θ, the area is

$$A_{\mathrm{p}} = \int I(Z(X) > t \mid Z, \theta)\, \mathrm{d}X \tag{11}$$

It can be approximated by sampling $Z(X)$ and evaluating the integrand in Equation (11), and the mean of these values is approximately equal to the proportion of the affected area over the entire WCA2A. Samples of Z were taken from the joint predictive distribution of Z from some unobserved locations $\boldsymbol{s}_0 = \{s_0^1, \ldots, s_0^n\}$ randomly selected over entire WCA2A. The joint predictive distribution is a multivariate t-distribution (see Appendix for details).

The following procedure is used:

1. Generate a sample of θ according to Equation (5).
2. Using the θ generated in (1), sample a value of $Z(X)$ according to Equation (4), call it $Z(X_1)$.
3. Generate $Z(X) \mid Z(X_1)$ according to Equations (A5) and (A8) in the Appendix, call it $Z(X_2)$.
4. Generate $Z(X) \mid Z(X_1), Z(X_2)$ according to Equations (A5) and (A8), call it $Z(X_3)$.
5. Continue to sample $Z(X) \mid Z(X_1), Z(X_2), \ldots, Z(X_i)$ for n times, we have $Z(X_1), Z(X_2), \ldots, Z(X_n)$, samples from the conditional predictive distribution defined in Equation (A5).
6. Compute the mean of $I(Z(X_i) > t)$, $i = 1, 2, \ldots, n$. This gives one value of A_{p}, as the percentage of the affected region.
7. Repeat (1) – (6) m times to generate m samples of θ and A_{p}.

Due to the cost of computing, m in this study is limited to 200 and n to 150. The sample size may not be large enough to give a precise estimation of the variance; it is, nevertheless, indicative of the quality of the estimate of the mean value of the area.

For Bayesian kriging, the prior distribution of covariance function parameters was chosen as:

$$\pi(\theta) = \pi(\theta_1) \times \pi(\theta_2) = \frac{1}{(1 + \theta_1)^2 (1 + \theta_2)^2} \tag{12}$$

This prior distribution assumes that smaller values are more likely to occur than larger ones. This assumption is reasonable for both parameters. For the smoothness parameter θ_2, this prior poses, for example, a random field is more likely to be 2 or 3 times differentiable than, say, 99 times differentiable. For the range parameter θ_1, the prior indicates that two closely-located points are more likely to be correlated than two points farther apart.

4. Results

The distribution of the residuals of model (1) is found to be close to Gaussian. Z in Equation (1) is the log-transformed STP content; the predictor variables are the latitude, longitude, and the distance to the nearest source of agricultural runoff. The residuals of the model (1) are found to be intrinsically stationary and isotropic, through visual inspection of the variograms in both the north–south and east–west directions (not shown in this paper). The results are organized in three groups: covariance function, predictive distribution, and area estimation. For universal kriging, four different covariance functions are used. One of these functions is the Matérn covariance function using the maximum posterior estimate of θ, hereby referred to as UK-Matérn. The other three covariance functions correspond to the three commonly used variogram models: the exponential, the spherical, and the Gaussian. They are referred to as UK-exponential, UK-spherical, and UK-Gaussian, respectively.

4.1 Covariance function

The covariance function used in Bayesian kriging can be studied through the posterior distribution presented in Equation (5). The log-transformed posterior, as a function of the parameters, θ_1 and θ_2, are computed based on the non-informative prior shown in Equation (12), and plotted against the log-transformed (base 10) θ_1 and θ_2 as shown in Fig. 4. The values of the posterior

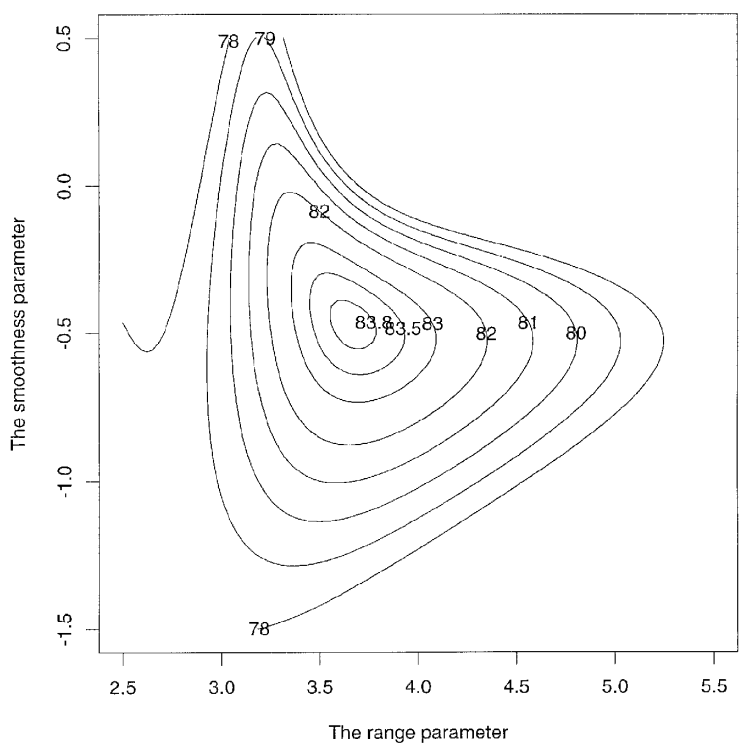

Fig. 4. The posterior distribution of the parameters of the Matérn covariance function.

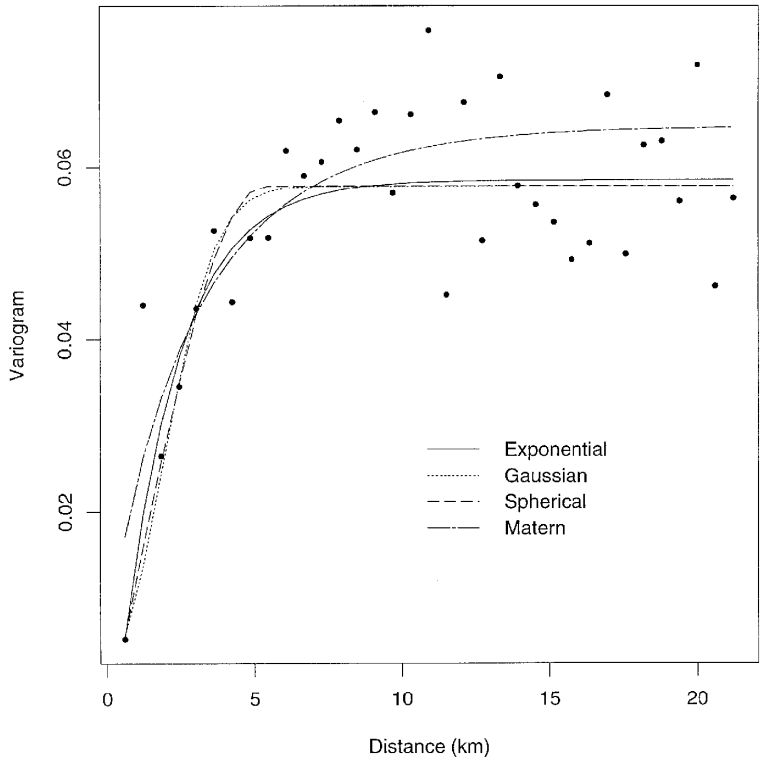

Fig. 5. The variogram models.

shown here are proportional to the log-posterior up to a constant. The posterior mode is at (3.6515, –0.4697); i.e., the Maximum Posterior Estimates (MPE) of the parameters are (4482, 0.3391). The estimated α given the MPE of θ is 0.032363.

The empirical variograms ($2\hat{\gamma}$) were calculated from the residuals of model (1) by using the S-Plus program in Venables and Ripley (1994). The three commonly used variogram models are fitted to the empirical variograms as a function of the distance, utilizing the weighted least squares method (Cressie, 1991). The exponential model best fits the empirical variograms (based on $\mathbb{R}^2$); however, all three variogram models fit well, where the line labelled 'Matérn' (as shown in Fig. 5) is the Matérn covariance function using the MPE of θ. Since the Matérn function is not fitted to the empirical variograms, it is understandable that the estimated Matérn covariance function does not seem to fit the empirical variograms very well. Fig. 6 compares the covariance functions derived from the above. The Matérn covariance function in Figure 6 is very different from the exponential model, although the MPE of the smoothness parameter θ_2 is very close to 0.5 ($\theta_2 = 0.5$ represents the exponential model.)

4.2 Predictive distributions

The predictive distributions of the log-transformed STP for the 10th and 25th sampling sites are estimated. Site 10 has the largest STP content, and the STP content of site 25 is below

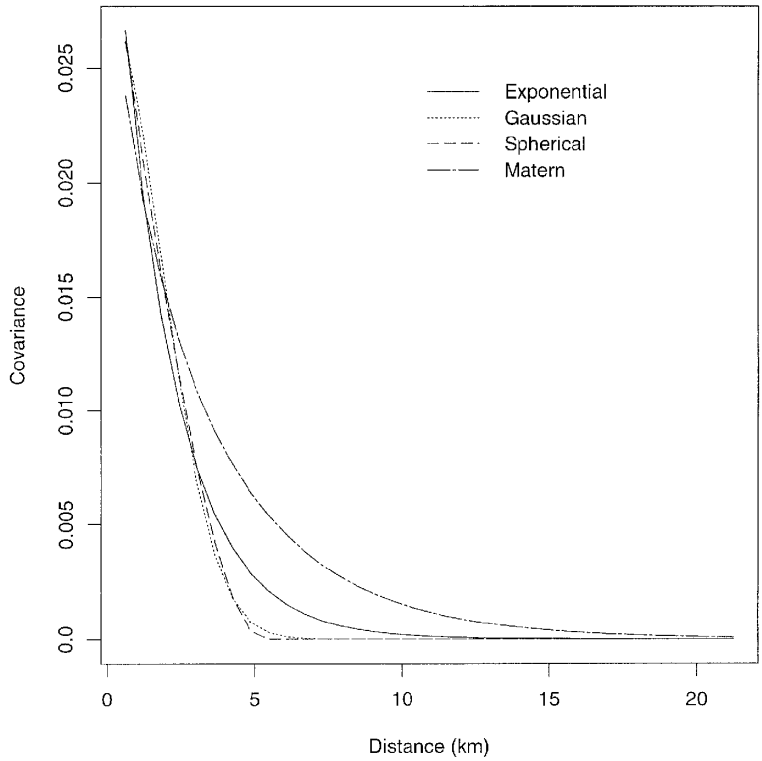

Fig. 6. The covariance functions.

Table 1. Predictive distributions

Model	*Sampling site*	*Predicted mean*	*Prob. > 500 μg/g*	*Measured STP*
BK	25	2.49	0.08	2.4
UK-Matérn	25	2.48	0.05	2.4
UK-exponential	25	2.46	0.03	2.4
UK-Gaussian	25	2.18	0	2.4
UK-spherical	25	2.45	0.02	2.4
BK	10	2.83	0.79	3.1
UK-Matérn	10	2.85	0.85	3.1
UK-exponential	10	2.84	0.82	3.1
UK-Gaussian	10	2.81	0.74	3.1
UK-spherical	10	2.84	0.84	3.1

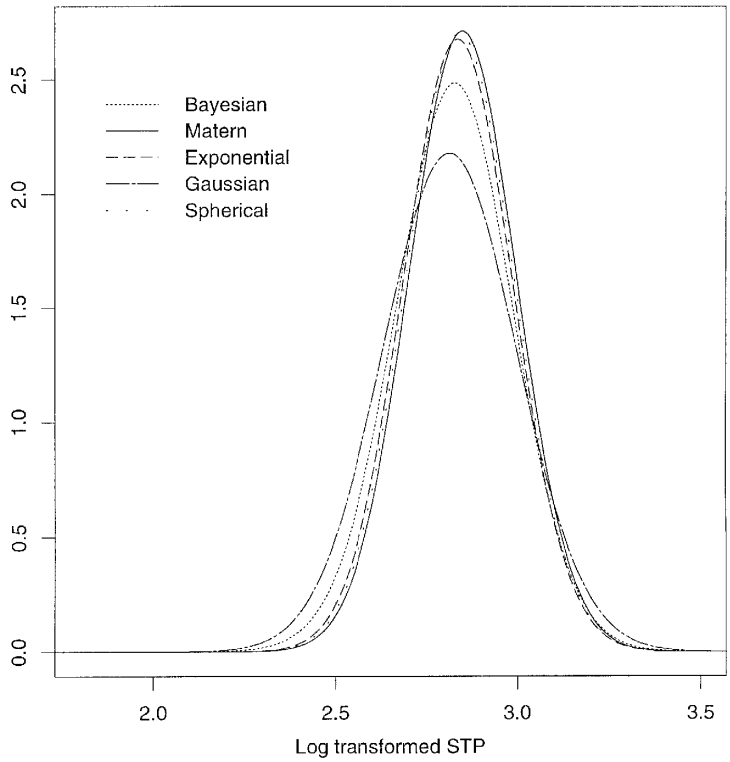

Fig. 7. Predictive distributions for the 10th sampling site.

average. When the predictive distribution for that site was computed, the same was not included in the data set. The density functions are shown in Figs 7 and 8. Table 1 compares the results of both Bayesian kriging and universal kriging.

Comparing five predictive distributions for the 10th sampling site, we note that:

1. UK-exponential, UK-spherical, and UK-Matérn are nearly identical.
2. The Bayesian kriging model yields a similar mean value but slightly larger estimation variance. A larger estimated variance from Bayesian kriging is to be expected owing to the fact that the uncertainty in the covariance function is ignored in the universal kriging analysis whereas it is considered in the Bayesian analysis.
3. The predictive distribution from UK-Gaussian has the largest variance.

For the 25th sampling site, comparable predictive distributions were observed from UK-Matérn, UK-exponential, and UK-spherical. Bayesian kriging yields a similar mean and slightly larger variance. The predictive distribution from UK-Gaussian is notably different from the other four. As shown by Stein and Handcock (1989), Gaussian models with different range parameters are not compatible, therefore a slight estimation error may yield a significantly different predictive distribution. On the other hand, the Gaussian covariance function is equivalent to the Matérn covariance function with $\theta_2 \to \infty$; i.e. realizations from it are infinitely differentiable. However, this representation may not be realistic in the natural world.

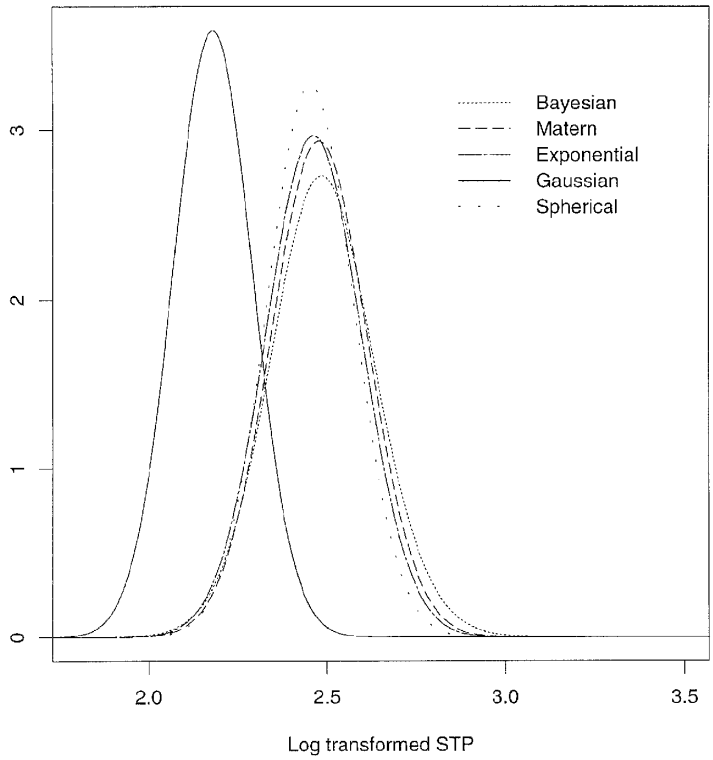

Fig. 8. Predictive distributions for the 25th sampling site.

4.3 Area of the affected region

The expected area is calculated using Equations (8) and (10), where X is the spatial coordinate vector, i.e. $X = (Northing, Easting)$. Equation (8) is used for universal kriging , and Equation (10) is used for Bayesian kriging. In Equation (8), the integrand is the upper tail of a normal distribution. The integral is calculated using a Gauss–Legendre quadrature over a 30×30 grid. In Equation (10), the integrand of the inner integral is the upper tail of a t-distribution, and the integral is calculated using Gauss–Legendre quadrature (Davis and Rabinowitz, 1984) over a 30×30 grid. The outer integral of Equation (10) is calculated using Gauss–Hermite quadrature, over a 20×20 grid. The results are listed in Table 2.

Table 2. Expected area

Model	*Affected area (km^2)*	*% of total area*
BK	97.19	22.39
UK-Matérn	95.77	22.07
UK-spherical	89.4	20.6
UK-exponential	99.61	22.95
UK-Gaussian	76.59	17.65

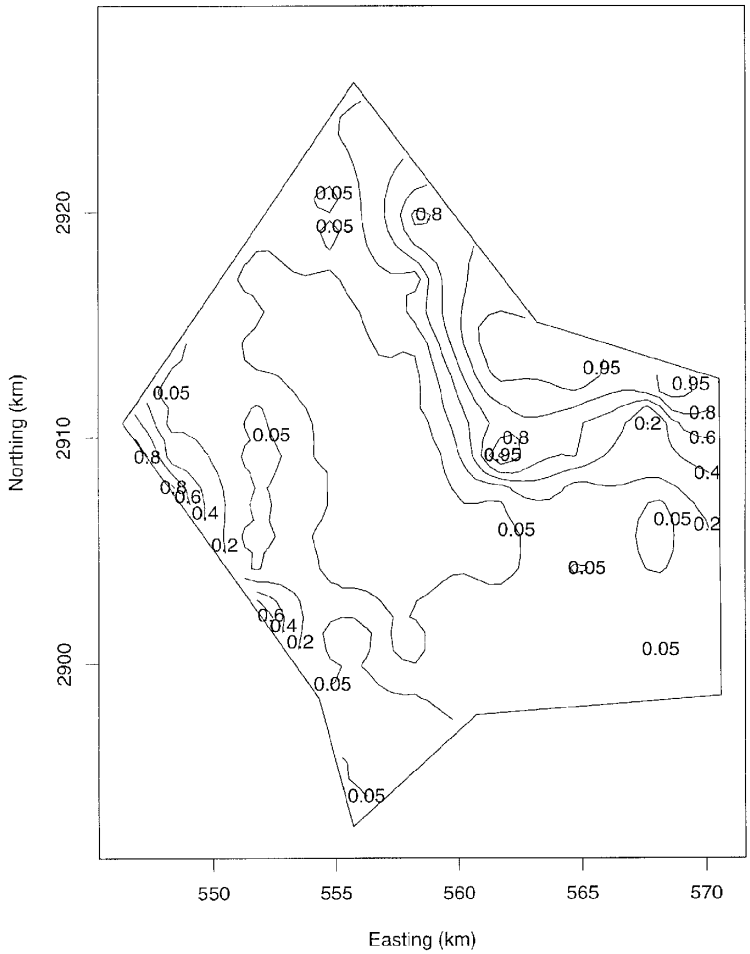

Fig. 9. Predicted probability that STP exceeds the background level using Bayesian kriging.

In this study, the boundary of the affected region was not specified simply because the area of the affected region (A_p) is treated as a random variable. However, the shape of the affected region may be stable; therefore the contour lines of prob(STP > 500 μg/g) were plotted to indicate the possible shape. Figure 9 shows contour lines using Bayesian kriging, and Figs 10–13 show the surfaces from the universal kriging analysis. The surface from the UK-Gaussian is mostly flat but has sudden jumps (large slopes). The reason for this unnaturally shaped surface is due to the fact that the Gaussian covariance function describes 'super smooth' surfaces of the residuals in this case. A model with very smoothed residual surface must have a jumpy mean surface.

From Table 2, we note that the estimated area using Bayesian kriging model is very close to the estimated area using UK-Matérn and UK-exponential. This is expected since the Bayesian kriging is based on the Matérn variogram and the estimated smoothness parameter is very close to the value of 0.5, which is the value it takes for the exponential variogram. The estimate of the area is sensitive to the specification of the covariance function. Bayesian kriging is seen as

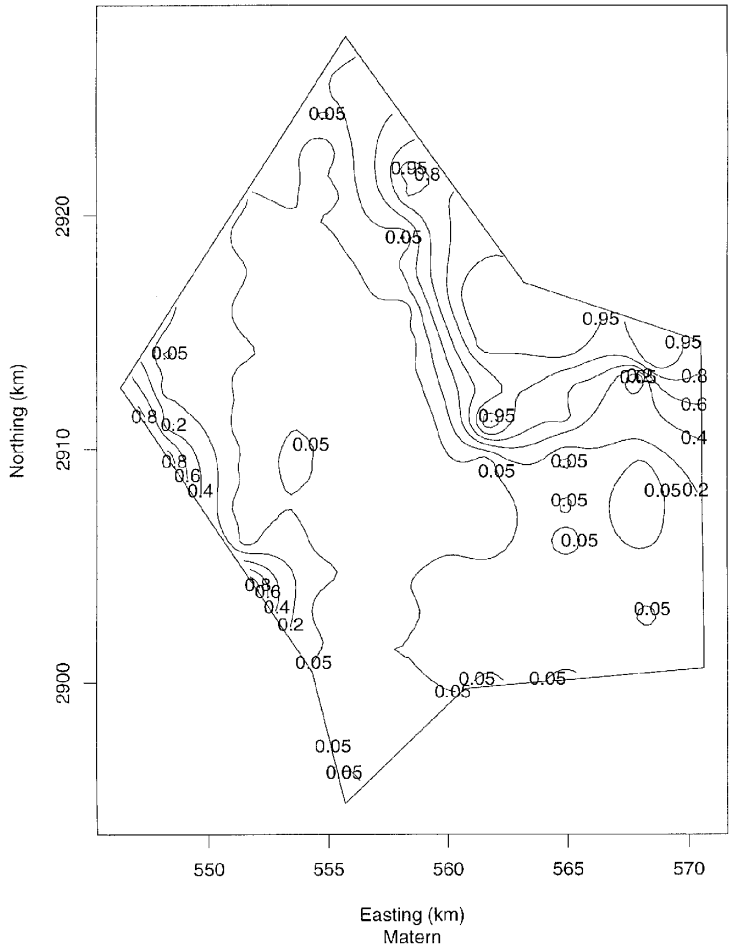

Fig. 10. Predicted probability that STP exceeds the background level using UK-Matérn.

more appropriate in this study, since the uncertainty of the unknown covariance function is taken into consideration.

The uncertainty of the covariance function is evaluated for Bayesian kriging, following the procedure previously described. The random variates of θ are generated from Equation (5), using the adaptive rejection Metropolis sampling method presented in Gilks *et al.* (1994). 200 pairs of θ were generated, and for each pair of θ, a random vector of size 150 was generated from the joint predictive distribution. The percentage of the generated $\boldsymbol{Z}_0 = (Z_0(s^1{}_0), \ldots, Z_0(s^{150}{}_0))$, larger than $\log_{10}(500)$, was recorded for each pair of θ. These percentages are approximately equal to the proportion of the polluted area of WCA2A. Figure 14 represents the histogram of these percentages, which indicates that the variation of the estimated area may be quite large. Since the sample size of the Monte Carlo simulation is small, it is not meaningful in calculating the variance.

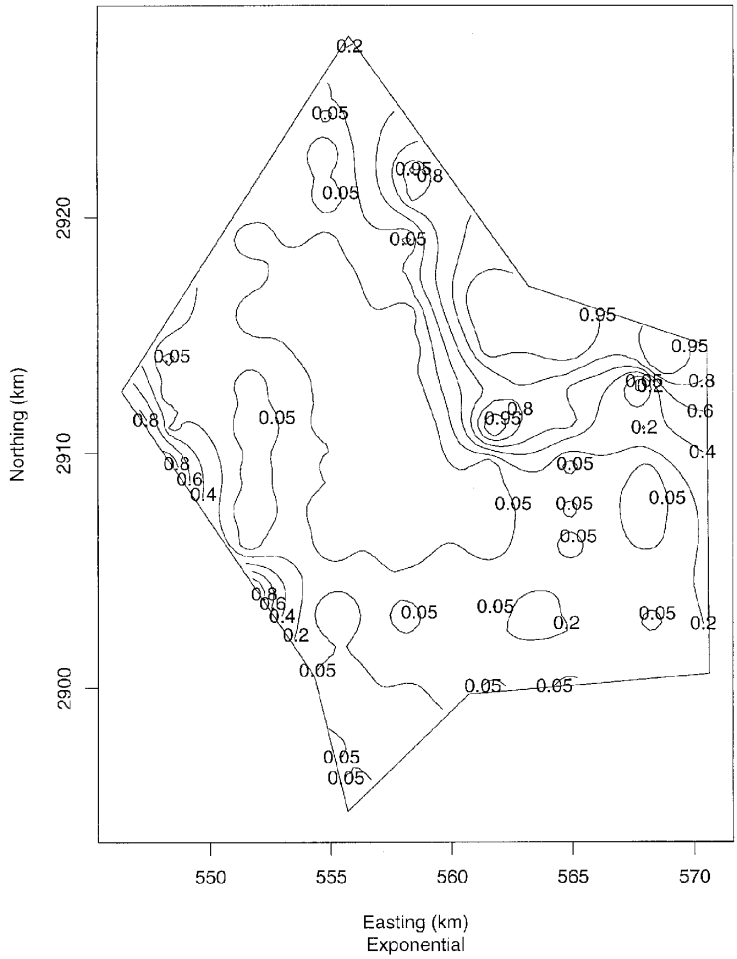

Fig. 11. Predicted probability that STP exceeds the background level using UK-exponential.

5. Discussion and Conclusion

In conclusion, it is seen that one significant difference between Bayesian kriging and universal kriging is an increased computational intensity of Bayesian kriging. On a DEC5000 workstation, it takes about 40 minutes to compute the affected area with Bayesian kriging, and under 1 minute with universal kriging. However, the results indicate that universal kriging is sensitive to the specification of covariance function. The difference between two different covariance functions may be significant even though variogram models of the two are fitted to empirical variograms equally well. In this study the spherical model and the Gaussian model fitted these empirical variograms equally well, but results from the universal kriging models using these two covariance functions are significantly different. This sensitivity to the covariance function justifies the use of Bayesian kriging.

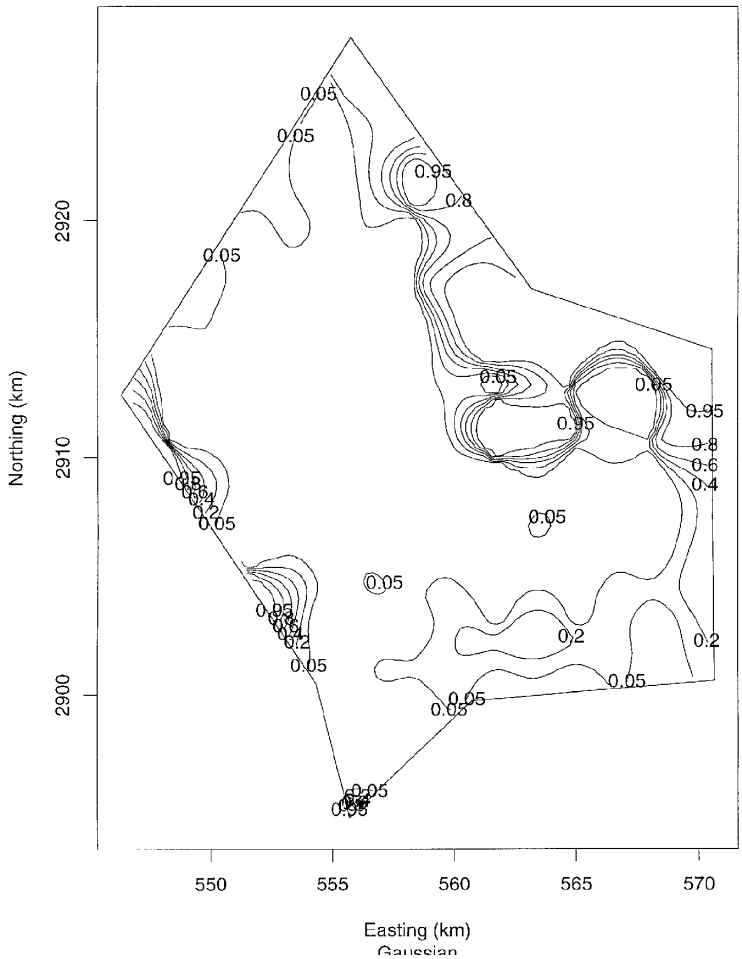

Fig. 12. Predicted probability that STP exceeds the background level using UK-Gaussian.

In terms of compatibility of a covariance function defined by Stein and Handcock (1989), the sensitivity of universal kriging to the specification of covariance functions can be explained. In their paper, it is shown that two exponential variograms with different range parameters are compatible. The same is not true for the Gaussian variogram, which is not compatible with the exponential variogram. Furthermore, Stein and Handcock showed that the spherical covariance function is an inappropriate model for most three-dimensional fields. However, the problem of recognizing the 'true' covariance function still exists; as stated in Stein and Handcock (1989) there is a need for 'a procedure that is able to distinguish strongly between variograms that are not compatible and choose a variogram that is nearly compatible with the true variogram'. On the other hand, although the effect of using an incorrect covariance function is negligible asymptotically, it is not a guarantee that the effect under a limited number of observations is also negligible.

There are very few studies which attempt to estimate the area of the affected region. One of them is by Craft and Richardson (1993). In their study, the boundary of the affected region

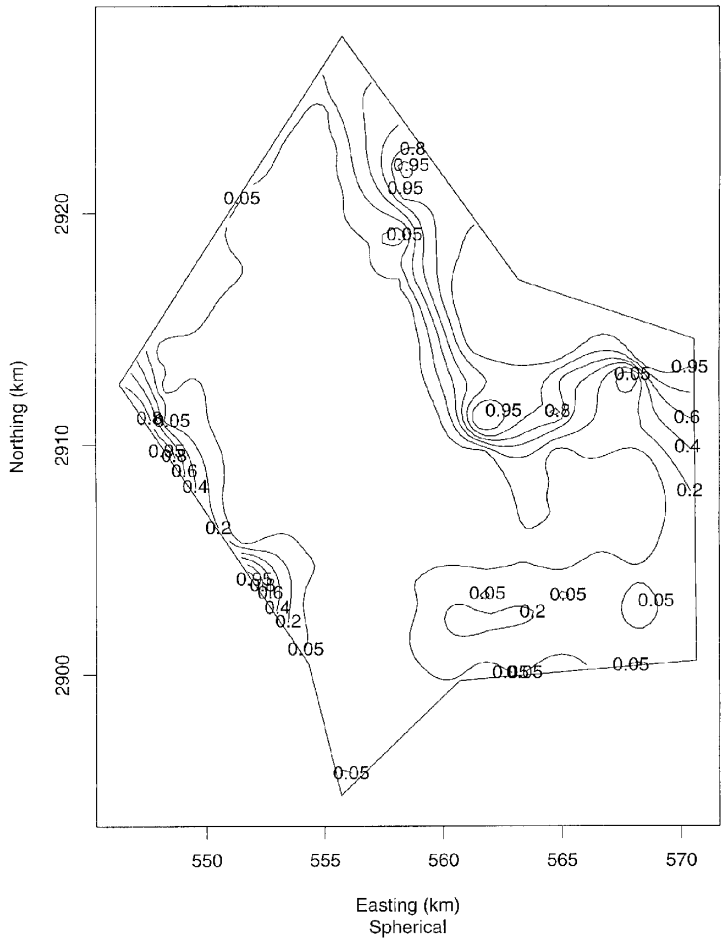

Fig. 13. Predicted probability that STP exceeds the background level using UK-spherical.

was delineated by using both the STP content and the phosphorus accumulation rate. The delineated boundary is plotted on to a USGS topographic map, and the area is thus measured. They estimated the affected area to be 115 km^2. The STP data used in their study were also from Reddy *et al.* (1991). Only the top 10 cm layer was used, which employed a background level of 600 μg/g. The STP contour lines were produced using an ordinary kriging algorithm from a commercial graphics software package (Surfer). The variogram model used was not reported.

Walker (1993) presented another method, largely based on the assumption that the flow of water over the region was plug flow. His estimate was about 70 km^2. However, the plug-flow assumption is not valid. WCA2A is not a simple flat wetland. There are tree islands all over the region, air-boat channels crisscross the densely vegetated area, and deep water sloughs are scattered throughout the wetland.

It is more appropriate to estimate the affected area by using the STP content as an indicator of the influence of the agricultural runoff. A spatial statistical approach accounts for the uncertainty in the sample data. We believe that Walker's estimate is not reliable on account of the

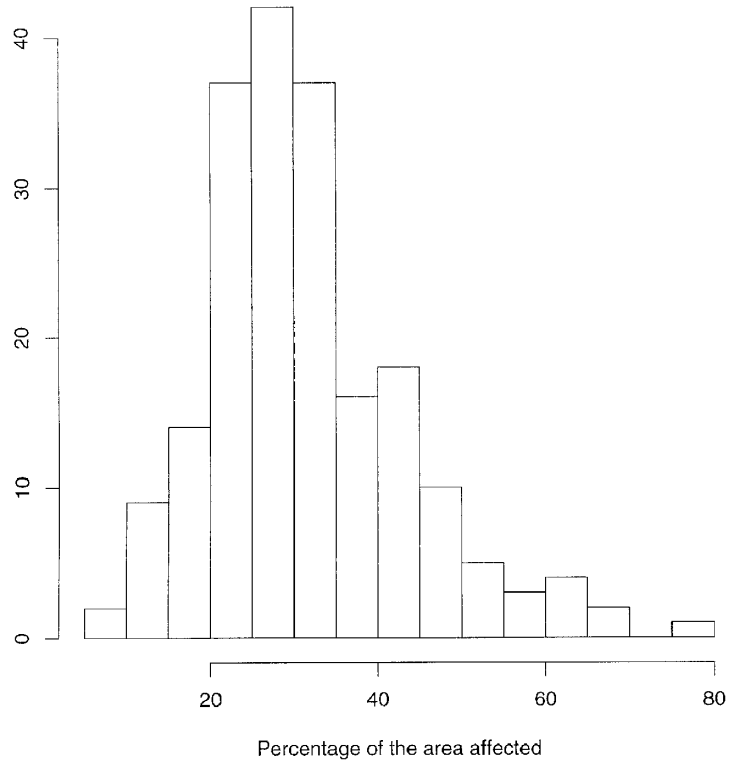

Fig. 14. Distribution of the phosphorus-affected area as percentage of the total area.

unrealistic plug-flow assumption. Considering the high level of uncertainty involved in the estimating process (Fig. 14), we feel that the area calculated by means of Bayesian kriging is comparable to the result in Craft and Richardson (1993). The discrepancy between the estimates may be caused by two factors: their use of a different covariance function and the background level, without considering the error of measuring the area from a map.

The normality assumption of Equation (1) is not necessary for kriging methods in some geostatistics studies. It is adopted in this study for two reasons:

1. Bayesian kriging of Handcock and Stein is based on the normality assumption, therefore, any comparison of the method without normality assumption is not warranted.
2. The prediction of $Z(s_0)$ is not the objective of this study, rather the estimation of the affected area is the objective. The predictive distribution of $Z(s_0)$ is used to calculate the area, derived by using a Bayesian approach. Without normality assumption, it is difficult, if not impossible, to find the distribution. Because of this, the Bayesian kriging model also incorporates uncertainty that is not quantified, namely the validity of the multivariate normal assumption.

It seems impossible to account for all sources of uncertainty using any specific method. The intention of this paper is to address the uncertainty in the covariance function when universal and Bayesian kriging were applied to Gaussian data. This study used the linear form of the universal kriging predictor $\mathrm{E}(Z(s_0) \mid Z)$, which is optimal under squared-error loss only for a

few exceptions (including Gaussian model) (Cressie, 1991, Section 5.1). For many non-Gaussian models, the optimal predictor is not linear, and the linear predictor is only an approximation. The Gaussian model is explicitly assumed for this purpose.

Many studies indicate that distributions of most water quality/resources related variables can be approximated by a log-normal distribution (Helsel and Hirsch, 1992). Therefore, log-transformation of the data will result in normally distributed variables.

There is no reason to believe that the background level is a constant. Field samples from the region that is not affected by agriculture runoff show a considerable variation (Craft and Richardson, 1993). The background level of STP content should be considered as a random variable. If a constant background level is used, uncertainty about background level is not taken into consideration. The author was not able to obtain enough data to evaluate the background level distribution at the time of this project. This source of uncertainty will be ascertained in a future study.

Acknowledgements

Drs Richard Smith and Peter Müller motivated the author to pursue this project. The author thanks Dr Marcia Gumpertz, the editor, and three referees, whose comments and suggestions led to a significantly improved version of this article. The author acknowledges the invaluable help of Dr Lily Chou for her editorial suggestions, which greatly enhanced the readability of this paper. The encouragement and support of Drs Michael Lavine, Curtis Richardson, and Ken Reckhow are greatly appreciated. This work was partially supported by Duke University Wetland Center, Duke University, Durham, North Carolina, USA.

References

Abramowitz, M. and Stegun, I.A. (1965) Handbook of Mathematical Functions. Dover, New York.

Ahmed, S. and de Marsily, G. (1987) Comparison of geostatistical methods for estimating transmissivity using data on transmissivity and specific capacity. *Water Resources Research*, **23**, 1717–37.

Black, T.C. and Freyberg, D.L. (1987) Stochastic modelling of vertically averaged concentration uncertainty in a perfect stratified aquifer. *Water Resources Research*, **23**, 997–1004.

Bogardi, I., Bardossy, A. and Duckstein, L. (1985) Multicriterion network design using geostatistics. *Water Resources Research*, **21**, 199–208.

Brooker, P. (1986) A parametric study of robustness of kriging variance as a function of range and relative nugget effect for a spherical variogram. *Mathematical Geology*, **18**, 477–88.

Brus, D.J. and de Gruijter, J.J. (1993) Design-based versus model-based estimates of spatial means: theory and application in environmental soil science. *Environmetrics*, **4**, 123–52.

Burgess, T.M., Webster, R. and McBratney, A.B. (1981) Optimal interpolation and isarithmic mapping of soil properties – IV: sampling strategies. *Journal of Soil Science*, **32**, 643–59.

Cox, D.D., Cox, L.H. and Ensor, K.B. (1995) Spatial sampling and the environment. Technical Report No. 38, National Institute of Statistical Sciences, Research Triangle Park, NC.

Craft, C.B. and Richardson, C.J. (1993) Peat accretion and phosphorus accumulation along a eutrophication gradient in the northern Everglades. *Biogeochemistry*, **22**, 133–56.

Cressie, N.A.C. (1991) *Statistics for Spatial Data*. Wiley, New York.

Cressie, N.A.C. and ver Hoef, J.M. (1993) Spatial statistical analysis of environmental and ecological data. In *Environmental Modeling with GIS*, M.F. Goodchild, B.O. Parks, and L.T. Steyaert (eds), Oxford University Press, New York, pp. 403–13.

Davis, P. and Rabinowitz, P. (1984) Methods of Numerical Integration. 2nd edn. Academic Press, Orlando, FL.

Diamond, P. and Armstrong, M. (1984) Robustness of variograms and conditioning of kriging matrices. *Mathematical Geology*, **16**(8), 809–22.

Gilks, W.R, Best, N.G. and Tan, K.K.C. (1994) Adaptive rejection Metropolis sampling within Gibbs sampling. Manuscript, Medical Research Council Biostatistics Unit, Institute of Public Health, Cambridge, UK.

Handcock, M.S. and Stein, M.L. (1993) A Bayesian analysis of kriging. *Technometrics*, **35**(4), 403–10.

Handcock, M.S. and Wallis, J.R. (1994) An approach to statistical spatial-temporal modelling of meteorological fields. *Journal of the American Statistical Association*, **89**(426), 368–90.

Helsel D.R. and Hirsch, R.M. (1992) *Statistical Methods in Water Resources*. Elsevier, Amsterdam.

Howard-Williams, C. (1985) Cycling and retention of nitrogen and phosphorus in wetlands: a theoretical and applied perspective. *Freshwater Biology*, **15**, 391–431.

Journel, A.G. and Huijbregts, C.J. (1978) *Mining Geostatistics*. Academic Press, London.

Laslett, G.M., McBratney, A.B., Pahl, P.J. and Hutchinson, M.F. (1987) Comparison of several spatial prediction methods for soil pH. *Journal of Soil Science*, **38**, 325–41.

Le, N.D. and Zidek, J.V. (1992) Interpolation with uncertain spatial covariance: a Bayesian alternative to kriging. *Journal of Multivariate Analysis*, **43**, 351–74.

Matheron, G. (1965) *Les Variables Regionalisées et Leur Estimation*. Masson, Paris.

Reckhow, K.H. and Qian, S.S. (1994) Modelling phosphorus trapping potential of wetlands using generalized additive models. *Water Resources Research*, **30**(11), 3105–14

Reddy, K.R., DeBusk, W.F., Wang, Y., DeLaune, R. and Koch, M. (1991) Physico-chemical properties of soils in the Water Conservation Area 2 of the Everglades. Final report submitted to South Florida Water Management District, West Palm Beach, FL 33416.

Ripley, B.D (1981) *Spatial Statistics*. Wiley, New York.

Stein, M.L. (1988) Asymptotically efficient spatial interpolation with a misspecified covariance function. *Annals of Statistics*, **16**, 55–63.

Stein, M.L. and Handcock, M.S. (1989) Some asymptotic properties of kriging when the covariance function is misspecified. *Mathematical Geology*, **21**(2), 171–90.

Venables, W.N. and Ripley, B.D. (1994) *Modern Applied Statistics with S-Plus*. Springer-Verlag, New York.

Verly, G., David, M. Journel, A.G. and Marechal, M. (eds) (1984) *Geostatistics for Natural Resources*. Reidel, Dordrecht.

Walker, W.W. Jr. (1993) A mass-balance model for estimating phosphorus settling rate in Everglades Water Conservation area – 2A. Prepared for U.S. Department of Justice.

Warnes, J.J. (1986) A sensitivity analysis for universal kriging. *Mathematical Geology*, **18**(7), 653–76.

Yakowitz, S.J. and Szidarovsky, F. (1985) A comparison of kriging with nonparametric regression methods. *Journal of Multivariate Analysis*, **16**, 21–53.

Zimmerman, D.L. and Cressie, N.A.C. (1992) Mean squared prediction error in the spatial linear model with estimated covariance parameters. *Annals of the Institute for Statistical Mathematics*, **44**(1), 27–43.

Appendix

A.1 Conditional predictive distribution for several unobserved points

Le and Zidek (1992) used the conjugate prior distributions for the multivariate normal mean and covariance matrix, and they found that the conditional posterior distribution of the unobserved points is a multivariate *t*-distribution. In this study, we use the non-informative prior (Equation 7) to derive the conditional posterior distribution.

Let $\boldsymbol{s}_0 = (s_0^1, \ldots, s_0^r)$ be the vector of r unobserved points, and $\boldsymbol{Z}_0 = (Z(s_0^1), \ldots, Z(s_0^r))'$. The covariance matrix of the unobserved points is denoted by V_{x0}, the covariance matrix of the unobserved and observed points can be partitioned into four parts:

$$\sum = \alpha \begin{pmatrix} V_{x_0} & \nu_\theta' \\ \nu_\theta & K(\theta) \end{pmatrix} \tag{A1}$$

where the matrix ν_θ contains the covariances between the observed and unobserved points.

Based on the normality assumption, the conditional distribution of $(\boldsymbol{Z}_0 \mid \alpha, \theta, Z)$ is also normal:

$$\boldsymbol{Z}_0 \mid \alpha, \theta, Z \sim \mathrm{N}(\hat{Z}_0(\theta), \alpha V_0(\theta)) \tag{A2}$$

where:

$$\hat{\boldsymbol{Z}}_0(\theta) = \nu_\theta' K_\theta^{-1} Z + b_\theta' \hat{\beta}$$

$$V_0(\theta) = V_{x_0} - \nu_\theta' K_\theta^{-1} \nu_\theta + b_\theta'(X' K_\theta^{-1} X)^{-1} b_\theta$$

$$b_\theta = \boldsymbol{x}_0 - X' K^{-1} \nu_\theta$$

$\boldsymbol{x}_0$ is the matrix of independent variables for s_0

$\mathrm{pr}(\boldsymbol{Z}_0 \mid \theta, Z)$ is obtained by integrating the density of (A2) with respect to α:

$$\mathrm{pr}(\boldsymbol{Z}_0 \mid \theta, Z) = \int \mathrm{pr}(\boldsymbol{Z}_0 \mid \alpha, \theta, Z)\, \mathrm{pr}(\alpha \mid \theta, Z)\, \mathrm{d}\alpha$$

Since $G^2(\theta)/\alpha$ has a χ^2 distribution, we know that

$$\mathrm{pr}(\alpha \mid \theta, Z) \propto [G^2(\theta)]^{(n-q)/2} \alpha^{-(n-q+2)/2} \exp\left\{-\frac{G^2(\theta)}{2\alpha}\right\}$$

Therefore

$$\mathrm{pr}(\boldsymbol{Z}_0 \mid \theta, Z) \propto |V_0(\theta)|^{-1/2} [G^2(\theta)]^{(n-q)/2}$$

$$\times \int \alpha^{-(n-q+r)/2} \exp^{-1}\left\{-\frac{1}{2\alpha}\left[G^2(\theta) + (Z(\boldsymbol{s}_0) - \hat{Z}(\boldsymbol{s}_0))' V_0^{-1}(\theta)(Z(\boldsymbol{s}_0 - \hat{Z}(\boldsymbol{s}_0))\right]\right\} \mathrm{d}\alpha \tag{A3}$$

Since the integrand of (A3) is proportional to an inverse Gamma distribution, (A3) can be easily integrated out:

$$\mathrm{pr}(\mathbf{Z}_0 \mid \theta, Z) \propto |V_0(\theta)|^{-1/2}[G^2(\theta)]^{(n-q)/2}$$

$$\times \left[G^2(\theta) + (Z(\boldsymbol{s}_0) - \hat{Z}(\boldsymbol{s}_0))' V_0^{-1}(\theta)(Z(\boldsymbol{s}_0 - \hat{Z}(\boldsymbol{s}_0)) \right]^{(n-q+r)/2}$$

$$\propto \left[\frac{G^2(\theta)}{n-q} \right]^{-r/2} |V_0(\theta)|^{-1/2} \left\{ 1 + \frac{1}{n-q} \frac{1}{G^2(\theta)/(n-q)} \left(Z(\boldsymbol{s}_0) - \hat{Z}(\boldsymbol{s}_0) \right)' \right.$$

$$\left. \times V_0^{-1}(\theta)\, (Z(\boldsymbol{s}_0 - \hat{Z}(\boldsymbol{s}_0)) \right\}^{-(n-q+r)/2} \tag{A4}$$

It is recognized that $\mathrm{pr}(\mathbf{Z}_0 \mid \theta, Z)$ is a t-distribution:

$$\mathbf{Z}_0 \mid \theta, Z \sim t_{n-q} \left(\hat{Z}_\theta(\boldsymbol{s}_0), \frac{n}{n-q}\ \hat{\alpha}(\theta)\, V_0(\theta), r \right) \tag{A5}$$

A2 Conditional density of a multivariate t-distribution

Let the p-dimensional vector

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

be distributed as $t_\alpha(\mu, \Sigma)$. The density function is:

$$\mathrm{pr}(X) = \frac{\Gamma(\alpha+p)/2}{|\Sigma|^{1/2}(\alpha\pi)^{p/2}\Gamma(\alpha/2)} \left[1 + \frac{1}{\alpha}(X-\mu)'\Sigma^{-1}(X-\mu) \right]^{-(\alpha+p)/2}$$

Let

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

be the mean and

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

the variance. It can be shown that:

$|\Sigma| = |\Sigma_{22}|\, |\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}|$ and

$$\mathrm{pr}(X_1 \mid X_2) = \frac{\mathrm{pr}(X)}{\mathrm{pr}(X_2)},$$

and we know

$$\mathrm{pr}(X_2) = \int \mathrm{pr}(X)\, \mathrm{d}X_1 .$$

Let

$$R = (X_2 - \mu_2)' \, \Sigma_{22}^{-1} (X_2 - \mu_2)$$

$$X_1{}^* = \mu_1 + \Sigma_{12} \Sigma_{22}{}^{-1} (X_2 - \mu_2)$$

$$\Sigma_{11}^* = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$$

We have:

$$\mathrm{pr}(X) \propto |\Sigma|^{-1/2} \left[1 + \frac{R}{\alpha} + \frac{1}{\alpha} (X_1 - X_1^*)' (\Sigma^*{}_{11})^{-1} (X_1 - X_1^*) \right]^{-(\alpha + p_1 + p_2)/2}$$

$$\propto |\Sigma_{22}|^{-1/2} \left[1 + \frac{R}{\alpha} \right]^{(\alpha+p)/2} \left[1 + \frac{1}{\alpha + p_2} \cdot \frac{(X_1 - X_1^*)' (\Sigma_{11}^*)^{-1} (X_1 - X_1^*)}{(\alpha/[\alpha + p_2])(1 + [R/\alpha])} \right]^{-(\alpha +_2 + p_1)/2}$$

$$\mathrm{pr}(X_2) = \int \mathrm{pr}(X)\, \mathrm{d}X_1 \propto \left[1 + \frac{R}{\alpha} \right]^{(\alpha+p)/2} \times$$

$$\int \left[1 + \frac{1}{\alpha + p_2} \cdot \frac{(X_1 - X_1^*)' (\Sigma_{11}^*)^{-1} (X_1 - X_1^*)}{(\alpha/[\alpha + p_2])(1 + (R/\alpha])} \right]^{-(\alpha + p_2 + p_1)/2} \mathrm{d}X_1$$

or

$$\mathrm{pr}(X_2) \propto \left[1 + \frac{R}{\alpha} \right]^{-(\alpha+p)/2} \left| \frac{\alpha}{\alpha + p_1} \left(1 + \frac{R}{\alpha} \right) \Sigma_{11}^* \right|^{1/2} |\Sigma_{22}|^{-1/2} |\Sigma_{11}^*|^{-1/2}$$

(A6) is equivalent to:

$$\mathrm{pr}(X_2) \propto \left[\frac{\alpha}{\alpha + p_2} \right]^{p_1/2} \left[1 + \frac{R}{\alpha} \right]^{p_1/2} |\Sigma_{22}|^{-1/2} \left[1 + \frac{R}{\alpha} \right]^{-(\alpha + p_1 + p_2)/2}$$

$$\propto |\Sigma_{22}|^{-1/2}\left[1+\frac{R}{\alpha}\right]^{-(\alpha+p_2)/2} \quad (\text{or } X_2 \sim t_\alpha(\mu_2, \Sigma_{22}))$$

$$\mathrm{pr}(X_1|X_2) \propto \frac{|\Sigma_{22}|^{-1/2}\,|\Sigma^*_{11}|^{-1/2}\left[1+\frac{R}{\alpha}\right]^{-(\alpha+p)/2}\left[1+\frac{1}{\alpha+p_2}\cdot\frac{(X_1-X_1^*)'(\Sigma^*_{11})^{-1}(X_1-X_1^*)}{(\alpha/(\alpha+p_2))(1+(R/\alpha))}\right]^{-(\alpha+p_2+p_1)/2}}{|\Sigma_{22}|^{-1/2}\left[1+R/\alpha\right]^{-(\alpha+P_2)/2}}$$

$$\mathrm{pr}(X_1|X_2) \propto |\Sigma^*_{11}|^{-1/2}\left[1+\frac{R}{\alpha}\right]^{-p_1/2}\left[1+\frac{1}{\alpha+p_2}\cdot\frac{(X_1-X_1^*)'(\Sigma^*_{11})^{-1}(X_1-X_1^*)}{(\alpha/(\alpha+p_2))(1+R/\alpha)}\right]^{-(\alpha+p_2+p_1)/2} \tag{A7}$$

or

$$X_1|X_2 \sim t_{\alpha+p_2}\left[X_1^*, \frac{\alpha}{\alpha+p_2}\left(1+\frac{R}{\alpha}\right)\Sigma^*_{11}\right] \tag{A8}$$

In case of a bivariate t-distribution, that is

$$\mu = \begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix}$$

and

$$\Sigma = \begin{bmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{21} & \sigma_{22}\end{bmatrix}$$

the conditional density is:

$$X_1|X_2 \sim t_{\alpha+1}\left[\mu_1+\frac{\sigma_{12}^2}{\sigma_{22}}\left(X_2-\mu_2\right), \frac{\alpha}{\alpha+1}\left(1+\frac{(X_2-\mu_2)^2}{\alpha\sigma_{22}}\right)\left(\sigma_{11}-\frac{\sigma_{12}^2}{\sigma_{22}}\right)\right]$$

Biographical sketch

Song S. Qian is an assistant professor of environmental science at Portland State University in Portland, OR, USA. He received his PhD degree in environmental sciences and MS degree in statistics from Duke University. He has a Master's degree in environmental systems engineering from Nanjing University, and a BS degree in environmental engineering from Tsinghua University, both in China.

Discussion

DONALD E. MYERS

Mathematics Department, University of Arizona, Tucson, Arizona 85721, USA

The following comments are not so much a critique of the above paper but rather intended to present some additional perspectives on the problem and some variations on the method of analysis. The author considers the area within Water Conservation Area–2A (WCA2A) where the phosphorus content exceeds a specified level. Determining or estimating the area affected is judged to be useful in understanding how wetlands respond to added phosphorus. As noted by the author, the area of the affected region of WCA2A is the same as the proportion of the WCA2A that is affected. Unlike some soil or water contamination problems, the contemplated remediation is designed to prevent phosphorus from flowing into WCA2A rather than removing it from WCA2A. In particular the environment is not static and changes with time **if** there is agricultural run-off. The data used in the analysis represents a snapshot in time.

The analysis considers only one 'specified level' but even then one additional question is of interest, namely, the mean concentration of phosphorus in the affected area. As noted by the author there is evidence that the concentration of phosphorus is not spatially constant, i.e. not uniform, and hence not only the conditional mean but also the conditional variance would be of interest. Estimating the area of the affected region for multiple cutoffs is similar to but not quite equivalent to estimating the conditional mean. These areas could be estimated in the same way as in the paper by repeating the analysis for each cutoff, however the author notes that the analysis is computationally intensive when using Bayesian kriging. One disadvantage of simply estimating the area of the affected region is that there is no assurance that the affected region is contiguous nor where it is located within WCA2A. If it is not connected then knowing the number of connected subregions would be of interest. One possible approach is to partition WCA2A into subregions and apply the analysis to each subregion separately. In that case one of the crucial questions is the size and shape of the subregions. This is not unrelated to the problem of optimum mining blocks in ore reserve estimation and in turn depends on the partition of the variance, i.e. the variability of the phosphorus content within a subregion vs the variability between subregions. This can be computed using the variogram or covariance function.

The area of the affected region is computed as the integral of an indicator function, $I(Z(X) > 500) \mid Z)$ which would be the same as the integral of $I(\ln Z(X) > \ln t \mid Z)$. However when universal kriging is used the integral is numerically approximated by considering $I(\ln^* Z(X) > \ln t \mid Z)$ where $\ln^* Z(X)$ is the universal kriging estimate of $\ln Z(X)$. However this is not the same as the integral of $I(Z^*(X) > t|Z)$ since it is well known that the lognormal kriging estimator is biased and $\exp[\ln^* Z(X)]$ is not the same as $Z^*(X)$. If $Z(X)$ is not second order stationary, i.e. if the drift term in Equation (1) is not a constant then the log transformation does not transform $Z(X)$ into a similar decomposition of a deterministic component and a random component. This question has not been addressed in the paper.

Finally, a word about 'Bayesian kriging'. This term has been used by a number of authors but the meanings are not all the same. Omre and Halvorsen (1989) used it to refer to using Bayesian methods to estimate and model the drift term. O'Hagan (1992) uses it to model the

drift term, Cui *et al*. (1995) use Bayesian updating of the variogram parameters. Abrahamsen (1992) uses a Bayesian approach for incorporating additional information. Christakos (1990) has an entirely different approach.

Additional references

Abrahamsen, P.(1992) Bayesian kriging for seismic depth conversion of a multi-layer reservoir. In *Geostatistics Troia '92*, A. Soares (ed), Kluwer, Dordrecht, 385–98.

Christakos, G. (1990) A Bayesian/maximum-entropy view of the spatial estimation problem. *Mathematical Geology*, **22**, 763–78

Cui, H., Stein, A. and Myers, D.E. (1995) Extension of spatial information, Bayesian kriging and updating of prior variogram parameters. *Environmetrics*, **6**, 373–84

O'Hagan, A. (1991) Some Bayesian numerical analysis. In *Bayesian Statistics 4*, J.M. Bernado (ed.), Oxford University Press, pp. 1–17

Omre, H. and Halvorsen, K.B. (1989) The Bayesian bridge between simple and universal kriging. *Mathematical Geology*, **21**, 767–86

Rejoinder

SONG S. QIAN

I appreciate the comments of Dr Myers very much. I have only three brief remarks to make in this rejoinder.

First, the data used in the paper is the total phosphorus content in the top 20 cm of the soil, which represents the accumulation in the past 60 or so years. (The annual soil buildup in WCA2A is slightly over 0.3 cm.) Because phosphorus does not have a gaseous phase, it cannot be released into the atmosphere and can only be stored in the soil or remain in the water. For the purpose of this study, that is to study the effects of agricultural runoff from the Everglades Agricultural Area (EAA) which has been in existence for about 50 years, the data represent the effect of the entire EAA history. If the water column phosphorus concentration were used, the data would represent a snapshot in time.

Second, the proposed constructed wetlands are expected to be similar to WCA2A in terms of ecological functions, but with better engineering control of water distribution. However, the question under debate is how large the constructed wetlands should be in order to reduce the phosphorus concentration in agricultural runoff to background level. Since the effluent water column phosphorus concentration from WCA2A is the same as the background level, the affected area of WCA2A indicates what the size of the proposed constructed wetlands should be. This is why only the affected area is of interest. The conditional mean concentration of phosphorus would indeed be of interest, especially in studies of how ecosystems respond to elevated nutrient levels. For example, there are two mono-culture zones where respectively, cattail and sawgrass are the dominant species, and a mixed cattail and sawgrass zone in WCA2A.

Conditional mean phosphorus and other nutrient concentrations in the soil would provide useful information on the optimal environment for different species.

Third, the lognormal kriging estimate is indeed biased. However, in this study the universal kriging estimate is treated as a random variable and only the expected value of the integral of $I(\ln(Z(X)) > \ln(t) \mid Z)$ is estimated. From a Bayesian perspective, the posterior distribution of $Z(X)$ is normal with mean $\hat{Z}_\theta$ and variance αV_θ (Equation 2). Since the integral of $I(\ln(Z(X)) > \ln(t) \mid Z)$ is finite, its expectation with respect to $Z(X)$ is equal to the integral of $\text{prob}(\ln(Z(X)) > \ln(t) \mid Z)$, which is not biased. In other words, Equation (8) is an unbiased estimate of the expected value of the affected area. It is nevertheless important to point out the biased nature of the lognormal kriging estimator, since most variables related to water resources are approximately of lognormal distribution, and log-transformation of data is often a standard routine before data analysis. Inference should be made in terms of median and interquartile range, which are more relevant to biological and environmental sciences than mean and variance.