Journal of the Faculty of Environmental Science and Technology, Okayama University
Vol.4, No.1, pp.37-55, February 1999

Analysis of non-stationary spatial data : A study on the performances of Universal Kriging, Median-Polish Kriging and LOESS

Seung Bae CHOI [†] Yutaka TANAKA [‡]

(Received November 20 , 1998)

Abstract

One of major problems in spatial analysis is to estimate the value $z(s_0)$ at an unknown location s_0 using the information about observations $z(s_\alpha)$, $\alpha = 1, \cdots, n$. In this article, we will perform a numerical study about some methods for this problem. That is, we examine both the tranditional statistical method which does not take into account spatial correlation and the spatial statistical method which takes into account spatial correlation by applying them to a set of non-stationary spatial data. We compare the predictive powers of these methods. More precisely, we choose Universal Kriging(UK) and Median-Polish Kriging(MPK) as spatial statistical methods, and locally weighted regression or LOESS as a traditional method. As the major criterion for comparison, we use the so-called PRESS statistic, and also draw the prediction surface plot and the prediction standard error surface plot as minor criteria. A real numerical example of non-stantionary spatial data is analyzed for the comparison among the above three methods.

1 Introduction

When a data set is analyzed by traditional statistical methods, the analysis is performed assumming the observations are mutually independent. But in the case of time-series or spatial data, as they are correlated with one another, a general assumtion in traditional statistics may be absurd. That is to say, it is practically natural to assume that the closely located data in space/time are often more alike than those that are far apart, and this assumption has been used to model the physical or social phenomenon. Geostatistics, a branch of statistics dealing with spatial data, is different from traditional statistics in some terminology and has been developed isolatedly from the mainstream of statistics. And, because there are innumerable situations in which data are collected at various locations in space, application fields of spatial statistics are extensive. The application fields include geology, soil science, image processing, epidemiology, crop science, ecology, forestry, astronomy, atmospheric science and environmental science. Although both of time and space can be dealed with in these fields, this article will discuss only the space problem.

The major aims of this article are to review the available methods of predicting the values of unobserved points based on the observations at n points in two dimensional space and to compare their performances numerically. This kind of spatial prediction problem is known as

*Keywords: Stationary, Variogram, UK, MPK, LOESS

[†]Graduate School of Natural Science and Technology, Okayama University, Tsushima, Okayama 700, Japan
[‡]Department of Environmental and Mathematical Sciences, Okayama University, Tsushima, Okayama 700, Japan

"Kriging" in spatial statistics. It is mainly studied by Matheron(1965, 1969). In practice, the kriging problem is very important in daily life. For example, persons in the region where they do not have a contamination meter will feel like knowing the degree of contamination in water or air in the their own region.

On the other hand, many methods have been developed for the purpose of smoothing. They can be used also for spatial prediction. The difference between kriging and smoothing is that the former takes into account the spatial correlation while the latter does not. Among various kriging methods, "universal kriging" and "median-polish kriging" can be used for the analysis of non-stationary data. We adopt these two kriging methods. Among many methods of smoothing, we choose LOESS or locally weighted regression for our study. In our expectation, kriging methods are superior to smoothing methods because the former takes into account the spatial correlation but the latter do not. However, since each method has its own merits and demerits, it will be valuable to compare their performances numerically.

At first in Section 2, we introduce a basic ideas of spatial statistics. Universal kriging including a trend in the presence of nonstationarity is described in Section 3 and then a general idea of the median-polish kriging ,one method of removing a trend, is considered in Section 4. LOESS ,one of general smoothing methods, is described in Section 5. And Section 6 presents criteria to compare the performances of these three methods. In the last Section, the results of analysis of the real data are summarized and interpreted.

2 Basic ideas of spatial statistics

Spatial data can be considered as a realization of a stochastic process $Z(s)$, i.e.,

$$\{z(s) : s \in D \subset R^d\} \tag{1}$$

where s indicates a location in D and R^d is a d-dimensional Euclidean space. Most often d is 1, 2, or 3. The basic form of spatial data can be expressed as $(z_i,\ s_i) : i = 1, \cdots, n$, where z_i is the ith observation of a phenomenon of interest at location s_i. In spatial data analysis, it is assumed that the observed data have the following structure;

$$z(s) = m(s) + e(s), \tag{2}$$

where $m(s)$ denotes a large-scale variation called drift or trend and $e(s)$ a small-scale variation. The latter term is a fluctuating random component with zero expectation like random variation or measurement error within region. In most cases, a spatial data set represents a single realization of a random process. As such, some degree of stationarity must be assumed in order to make inferences about the data. Stationarity refers to some form of "location invariance" of the data. It implies that the relationships within any subset of points remain the same no matter where the points reside in space(mathsoft, 1996). In particular, when the mean, variance and covariance of stochastic process $Z(s)$ do not depend on the location, i.e.,

$$E(Z(\mathbf{s}+\mathbf{h}) - Z(\mathbf{s})) = 0, \tag{3}$$

$$Var(Z(\mathbf{s}+\mathbf{h}) - Z(\mathbf{s})) = 2\gamma(\mathbf{h}), \qquad \mathbf{s}, \mathbf{s}+\mathbf{h} \in D, \tag{4}$$

$Z(\cdot)$ is said to be intrinsically stationary. Here, $2\gamma(h)$ is called the variogram and $\gamma(h)$ the semivariogram. Futhermore, if $2\gamma(\mathbf{h}) = 2\gamma(||\mathbf{h}||)$, the variogram $2\gamma(\cdot)$ is called isotropic. If $2\gamma(\mathbf{h})$ depends the direction of $\mathbf{h}$ as well as the distance $||\mathbf{h}||$, it is called anisotropic. Although it is possible to think the covariance function or correlation function as measures of dependence,

variogram is usually used in spatial statistics.

When we can assume that the variogram is isotropic, the empirical variogram, a sample version of the variogram, is computed by

$$\gamma(h) = \frac{1}{2|N(h)|} \sum_{N(h)} (z_i - z_j)^2, \tag{5}$$

where $N(h)$ is the set of all pairs with Euclidean distance h, $|N(h)|$ is the number of distinct pairs in $N(h)$, and z_i and z_j are data values at spatial locations i and j, respectively. When the variogram is anisotropic, the directional empirical variogram is computed using the same fomula by replacing h by vector $\mathbf{h}$. In S-Plus, we can calculate the empirical variogram by using the variogram function, which has some optional arguments such as lag, nlag, lag(distance) tolerance and angle tolerance. The "lag" is the distance of the lags at which the variograms are calculated. If missing, it is automatically calculated as "$maxdist/nlag$". The "nlag" is the maximum number of lags to calculate, and the "lag tolerance(lag.tol)" indicates that pairs with distance within "$h \pm lag.tol$" are regarded as the pairs in $N(h)$. The "angle tolerance" plays the similar role in calculating directional variograms. In choosing lag and nlag, there are two pratical rules that should be considered(Mathstat, P76); Firstly, the empirical variogram should only be considered for distances h for which the number of pairs is greater than 30. Secondly, the distance when the variogram is reliable is $h < D/2$, where D is the maximum distance over the field of data. Usually, variogram is calculated using equation (5) given by Matheron(1963), but sometimes there are situations where it would be better to use robust variogram developed by Cressie and Hawkins(1980) in which the effect of outlier is reduced. It is given as follows.

$$\overline{\gamma}(h) = \frac{[\frac{1}{2|N(h)|} \sum_{N(h)} |z_i - z_j|^{1/2}]^4}{0.457 + 0.494/|N(h)|} \tag{6}$$

The next step of the variogram analysis is to fit a variogram model which explains best the dependence (autocorrelation structure) of the underlying stochastic process. Most variogram models are defined through several parameters; namely, the nugget effect, sill, and range. Theoretical variogram has several models(functions) according to its form; for example, sherical, Gaussian, exponential, power, and linear. S-plus, which is used for this study, provides functions for some theoretical variogram models. They include exponential, spherical and gaussian models as bounded variogram functions, and linear and power models as unbounded variogram models.

Exponential model:

$$\hat{\gamma}(\mathbf{h}; \boldsymbol{\theta}) = \begin{cases} 0, & \mathbf{h} = 0 \\ c_0 + c_e\{1 - exp(-\mathbf{h}/a_e), & \mathbf{h} \neq 0 \end{cases} \tag{7}$$

$\boldsymbol{\theta} = (c_0, c_e, a_e)$, where c_0 is nugget effect, c_e is sill, and a_e is range.

Spherical model:

$$\hat{\gamma}(\mathbf{h}; \boldsymbol{\theta}) = \begin{cases} 0, & \mathbf{h} = 0 \\ c_0 + c_s\{(\frac{3}{2})\frac{\mathbf{h}}{a_s} - (\frac{1}{2})(\frac{\mathbf{h}}{a_s})^3], & 0 < \mathbf{h} \leq a_s \\ c_0 + c_s, & \mathbf{h} > a_s \end{cases} \tag{8}$$

$\boldsymbol{\theta} = (c_0, c_s, a_s)$, where c_0 is nugget effect, c_s is sill, and a_s is range.

Gaussian model:

$$\hat{\gamma}(\mathbf{h}; \boldsymbol{\theta}) = \begin{cases} 0, & \mathbf{h} = 0 \\ c_0 + c_g\{1 - exp(-\mathbf{h}^2/(a_g)^2), & \mathbf{h} \neq 0 \end{cases} \tag{9}$$

$\boldsymbol{\theta} = (c_0, c_g, a_g)$, where c_0 is nugget effect, c_g is sill, and a_g is range.
Linear model:

$$\hat{\gamma}(\mathbf{h}; \boldsymbol{\theta}) = \begin{cases} 0, & \mathbf{h} = 0 \\ c_0 + c_l \mathbf{h}, & \mathbf{h} \neq 0 \end{cases} \tag{10}$$

$\boldsymbol{\theta} = (c_0, c_l)$, where $c_0 \geq 0$ and $c_l \geq 0$.
Power model:

$$\hat{\gamma}(\mathbf{h}; \boldsymbol{\theta}) = \begin{cases} 0, & \mathbf{h} = 0 \\ c_0 + b_p \mathbf{h}^{\lambda}, & \mathbf{h} \neq 0 \end{cases} \tag{11}$$

$\boldsymbol{\theta} = (c_0, b_p, \lambda)$, where $c_0 \geq 0$, $b_p \geq 0$, and $0 \neq \lambda < 2$. In our study those models are fitted to the empirical variogram, omnidirectional or directional, and then the model which fits best is selected.

Our interest is to estimate or predict the value at an arbitrary unobserved position based on the observed data as well as possible. In the case where it can be assumed that the stochastic process underlying the observation is second order stationary, a kriging method called oridinary kriging is widely used. It is a best linear unbiased prediction method, which is based on two assumptions; 1) Model assumption: The mean structure $m(s)$ is an unknown constant. 2) Predictor assumption: Linear predictors in the form $Z^*(s_0) = \sum_{i=1}^n w_i Z(s_i)$ are considered. For unbiasedness, the weights should satisfy $\sum_{i=1}^n w_i = 1$. By minimizing the prediction variance $E(Z(s_0) - Z^*(s_0))^2$ under the equality constraints on the weights, we can obtain the following system of equations;

$$\begin{aligned} &-\sum w_i \gamma(s_i - s_j) + \gamma(s_0 - s_i) - \mu = 0, \quad i, j = 1, \cdots, n, \\ &\sum_{i=1}^{n} w_i = 1, \end{aligned} \tag{12}$$

where μ is a Lagrange mutiplier. If the semivariogram $\gamma(s_i - s_j)$ and $\gamma(s_0 - s_i)$ are known, the optimal weights $\{w_i\}$ can be obtained by solving the above equations. In practical data analysis we usually do not know the variograms. Therefore, before applying this kriging method we have to estimate these variograms. More precisely, at first we calculate empirical variograms, omnidirectional or directional variograms depending on the structure of spatial correlation, and then find a theoretical variogram model which fits best the empirical variogram.

3 Universal Kriging

In case of non-stationary data, it is assumed that mean structure $m(s)$ can be expressed as an unknown linear combination of known functions. In our numerical example, we use a family of polynomial functions up to the second order. In general, the mean function is expressed a linear combination as follows.

$$m(s) = \sum_{j=0}^{p} \lambda_j f_j(s), \tag{13}$$

where $\lambda_1, \cdots, \lambda_p$ are fixed unknown nonzero parameters and f are known p functions of s. In particular, the function $f_0(s)$ is defined as $f_0(s) = 1$. For predicting the value at s_0, we consider a linear combination

$$Z^*(s_0) = \sum_{\alpha=1}^{n} w_\alpha Z(s_\alpha), \tag{14}$$

where w_α are weights. From the condition of unbiasedness

$$E[Z(s_0) - Z^*(s_0)] = 0 \tag{15}$$

which yields

$$m(s_0) - \sum_{\alpha=1}^{n} w_\alpha m(s_\alpha) = 0. \tag{16}$$

Substate equation (13) to equation (19), we obtain

$$\sum_{\alpha=0}^{p} \lambda_j (f_j(s_0) - \sum_{\alpha=1}^{n} w_\alpha f_j(s_\alpha)) = 0. \tag{17}$$

Since λ_j is nonzero,

$$\sum_{\alpha=1}^{n} w_\alpha f_j(s_\alpha) = f_j(s_0), \qquad j = 0, \cdots, p. \tag{18}$$

For the constant function $f_0(s)$ this is the usual condition expressed as

$$\sum_{\alpha=1}^{n} w_\alpha = 1. \tag{19}$$

Developing the expression for the prediction variance, introducing the constraints into the objective function together with Lagrange parameters μ_j and minizing, we obtain the following system of equations called the universal kriging system(Wackernagel, 1995).

$$\begin{aligned} &\sum_{\beta=1}^{n} w_\beta C(s_\alpha - s_\beta) - \sum_{j=0}^{p} \mu_j f_j(s_\alpha) = C(s_\alpha - s_0), \qquad \alpha = 1, \cdots, n, \\ &\sum_{\beta=1}^{n} w_\beta f_j(s_\beta) = f_j(s_0), \qquad j = 0, \cdots, p, \end{aligned} \tag{20}$$

where $C(\mathbf{h})$ is the covariance between $Z(\mathbf{s})$ and $Z(\mathbf{s}+\mathbf{h})$, which has the following relation with the semivariogram

$$\gamma(\mathbf{h}) = C(0) - C(\mathbf{h}). \tag{21}$$

Note that it is necessary to have the values of covariance functions $C(\cdot)$ or the semivariograms $\gamma(\cdot)$ to be able to solve the above UK system. In an actual data analysis, the UK method is applied as follows. 1) Plot the result of an empirical variogram(including the examination whether or not the variogram is different according to the direction). 2) Search for the theoretical variogram which fits best the empirical variogram. It is known that the universal kriging has some shortcomings. 1) The order of polynomial is usually not known and therefore must be estimated 2) Similarly, the variogram is usually not known and must be estmated from the residuals 3) The result of universal kriging is biased. 4) In case the variogram for the error is unknown, it is difficult to calculate. To avoid these thorny problems of Universal kriging, Cressie(1986) fitted variogram in the direction without trend. Following his idea, we fit the variogram in the direction in which there seems no trend in our study. The fitted theoretical variogram is given in Figure 6 of a section 7.2.1.

4 Median-Polish Kriging

Median-Polishing is a resistant method for detrending gridded data and is based on an additive decomposition(Mathsoft, 1996). As in ANOVA models, it is natural to decompose $m(\cdot)$ additively into directional components as

$$m(\mathbf{s}) = a + c(x) + r(y), \qquad \mathbf{s} = (x, y)' \in D, \tag{22}$$

where a is the general mean, c is the column effect and r is the row effect. In particular, if $\{s_i;\ i = 1, \cdots, n\}$ are actually on a grid $\{(x_l, y_k)';\ k = 1, \cdots p;\ \ l = 1, \cdots, q\}$, then, in obvious notation, $\mathbf{s}_i = (x_l, y_k)'$ implies $m(\mathbf{s}_i) = a + r_k + c_l$. Thus, the row effect r_k can be estimated by exploiting replication in the other dimension; that is, r_k can be estimated from $\{z(\mathbf{s}_i);\ 2nd\ coordinate\ of\ \mathbf{s}_i\ is\ y_k\ ;\ i = 1, \cdots, n\}$, where $k = 1, \cdots, p$. Similar considerations allow the column effect c_l to be estimated, $l = 1, \cdots, q$(Cressie, 1991). Miller and Kahn propose a formal two-way analysis of variance and claim to test for nonstationarity by performing the within-rows and within-columns F tests(Cressie, 1991). Unfortunately, the F tests are invalid because the data are in general correlated; however, underlying the two-way analysis of variance is an additive decomposition as above, and it is very useful.

The median-polish requires that the data are aligned in rows and columns, and thus is naturally suited for gridded data. However, the median polishing can be used also on non-gridded data. In such cases, the non-gridded data must be coerced to grids at first. The method is performed as follows. Because median-polish yields large-scale spatial variation, and because these main effects do not necessarilly depend on the data's precise spatial locations, a natural way to extend the approach to nongridded data is to draw a low-resolution map. That is to say, the resolution of the spatial coordinates is often chosen in an ad hoc way so that each (x_i, y_j) combination has approximately one observation $z(x_i, y_j)$ at (x_i, y_j). In practice, this is done by overlaying a grid onto the high-resolution map and assigning data location $\{s_i, i = 1, \cdots, n\}$ to the nearest nodes of the grid $\{(x_i, y_j)'\ , i = 1, \cdots p;\ \ j = 1, \cdots, q\}$(Cressie, p. 193). Therefore, $z(s_i)$ can be expressed by $z(x_i, y_j)$ and the median-polish is carried out on the data $\{z(x_i, y_j)\}$.

The median-polish residuals can be considered to be stationary. Therefore, the residuals can be analyzed by using the ordinary kriging. Taking into account both of the median polishing and the ordinary kriging, we can obtain the predicted value $Z^*(s_0)$ and its variance.

5 LOESS

Smoothers can be classified broadly as linear and nonlinear. Linear smoothers are composed of linear combinations of the data values, where the weights depend upon the Euclidean distances between the point to be smoothed and the points used for smoothing. Nonlinear smoothers often rely on combinations of medians and nearby data values. In linear smoothing methods, there are disk and weighted averages, empirical Bayes, LOESS etc. Nonlinear smoothing methods include headbanging, resmoothed medians and median polish(Kafadar, 1993). There are many smoothing methods including polynomial regression surface, spline and kriging. We adopt LOESS method among other possibility as the representative of smoothing methods which do not take into account the spatial correlation. Locally weighted regression, or LOESS, is a method to fit a polynomial surface a each point to be predicted using only the nearby data points. It is explained as follows(Cleveland and Devlin, 1988). Let $z_i (i = 1, \cdots, n)$ be measurements of the dependent variable, and let $s_i = (x_{i1}, \cdots, x_{ip})$, $i = 1, \cdots, n$ be n measurements of p independent variables. It is assumed that the structure is expressed as

$$z_i = g(s_i) + \epsilon_i, \tag{23}$$

where function g is a smoothing function of the independent variables and ϵ_i is an *i.i.d.* normal variable with mean 0 and variance σ^2. In this article, we consider the case of p=2, two densional smoother LOESS. The aim of the LOESS is to estimate the regression surface $g(s)$ at any point s in the 2-dimensional space of the independent variables. For a given fraction f of the data points, let R be the set of the nearest $q = [f \cdot n]$ points to z_i, the value to be smoothed. And $\mathbf{d}_R$ be the vector of the (q+1) distances from any s_i(one is 0 for the point itself) and let $d(s)$ be the maximum distance in the elements of $\mathbf{d}_R$. $\rho(s_k, s_i)$ is the Euclidean distance from s_i to s_k as a distance function in the space of the independent variables. The smoothed value of z_i, $\hat{z}_i$, is the predicted value from the weighted regression of z_R on $(\mathbf{x}_{1R}, \mathbf{x}_{2R})$, where $\mathbf{x}_{1R}$, $\mathbf{x}_{2R}$ and $\mathbf{z}_R$ are the three vectors of length $(q+1)$ corresponding to the points in the set R, and the weight for the observation (z_k, s_k) is given by $w_k(s_i) = [1 - (\rho(s_k, s_i)/d(s))^3]^3, \quad k \in R$. Notice that the set R changes for $i = 1, \cdots, n$ as in the case of moving averages in a time series(Kafadar, 1993).

6 Criteria for comparison

We wish to compare the performences of the three prediction methods. As a major measure, we adopt the so-called PRESS statistic defined as

$$PRESS = \sum_{i=1}^{n} \left\{ Z(s_i) - \hat{Z}(s_{-i}) \right\}^2, \tag{24}$$

where $Z(s_i)$ is an observation at the ith location and $\hat{Z}(s_{-i})$ is the predicted value of the ith location using the observed values excepting the ith one. The prediction surfaces and the prediction standard error surfaces are also compared.

7 Numerical example

7.1 Data

Rainfall data observed at 80 observation station are extracted from the chronological table of science('97) of Japan. The raw data are shown in Appendix A and the histograms are given in `Figure 1`. Histogram(1) is based on the original data themselves. It is noted that this distribution is highly skewed to the right. Then, we applied log transformation to the rainfall data. The histogram(2) shows the distribution of the log-transformed data. It is noted that the distribution of the transformed data is approximately normal and that there exist some outliers.

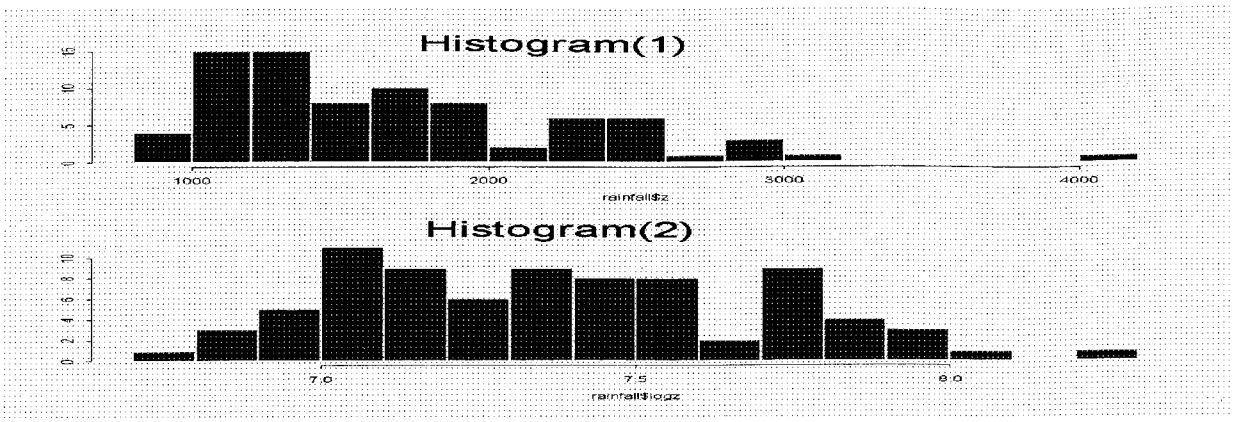

Figure 1 Histogram of the rainfall data: histogram 1 is of original data and histogram 2 is of log-transformed data

Then the coordinates are transformed as follows. At first, the original coordinates corresponding to latitude and longitude are scaled so that all points are located in the interval (-1, 1) for both axes. Figure 2(a) shows the data loctions along with contour curves for this scaled data.

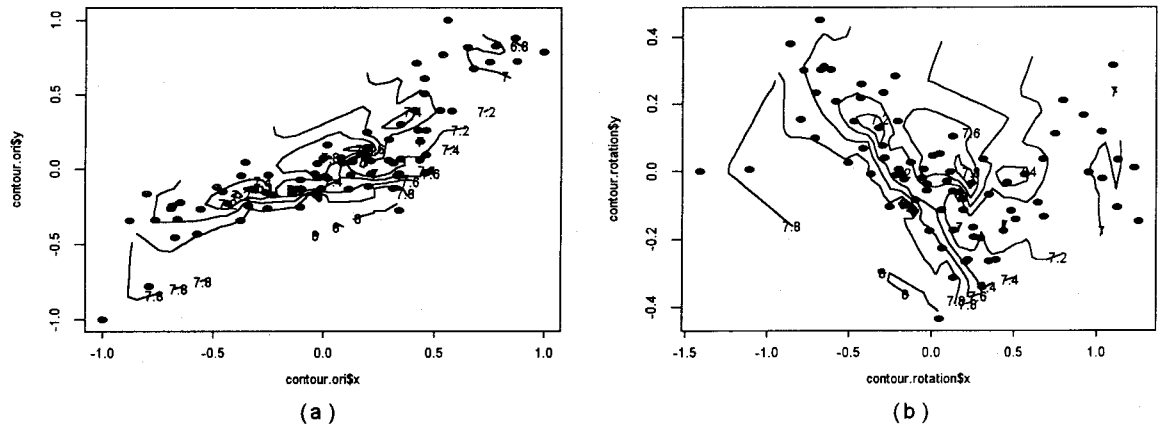

Figure 2 (a) Map of rainfall locations along with contour curves for original data. (b) Map of rainfall locations for 45° rotated data.

The contours in Figure 2(a) show a ridge of high rainfall values running northeast to southwest but do not show an evident trend. We shall investigate this spatial trend more carefully, since it will affect the modeling of our data. To show explicitly a trend and to make it easy to interpret in a direction of the eastwest and southnorth, the axes are rotated 45 degrees clockwisely from Figure 2(a) to Figure 2(b). We can see that the trend is clearer in (b) than in (a) of Figure 2. Thus, in this study, the data shown in Figure 2(b) are used for analysis, that is, these data are analyzed by the order given in analysis flow chart in Figure 3 using S-Plus package.

7.2 Prediction analysis

7.2.1 Universal kriging

In general, kriging is making use of a spatial correlation measure for describing the sample data variations with a distance and direction. In the spatial analysis, variogram is almost used

as a standard of the spatial correlation. The delineations of prediction surface and prediction standard error surface are fitted over a grid of 20 × 20 in this article.

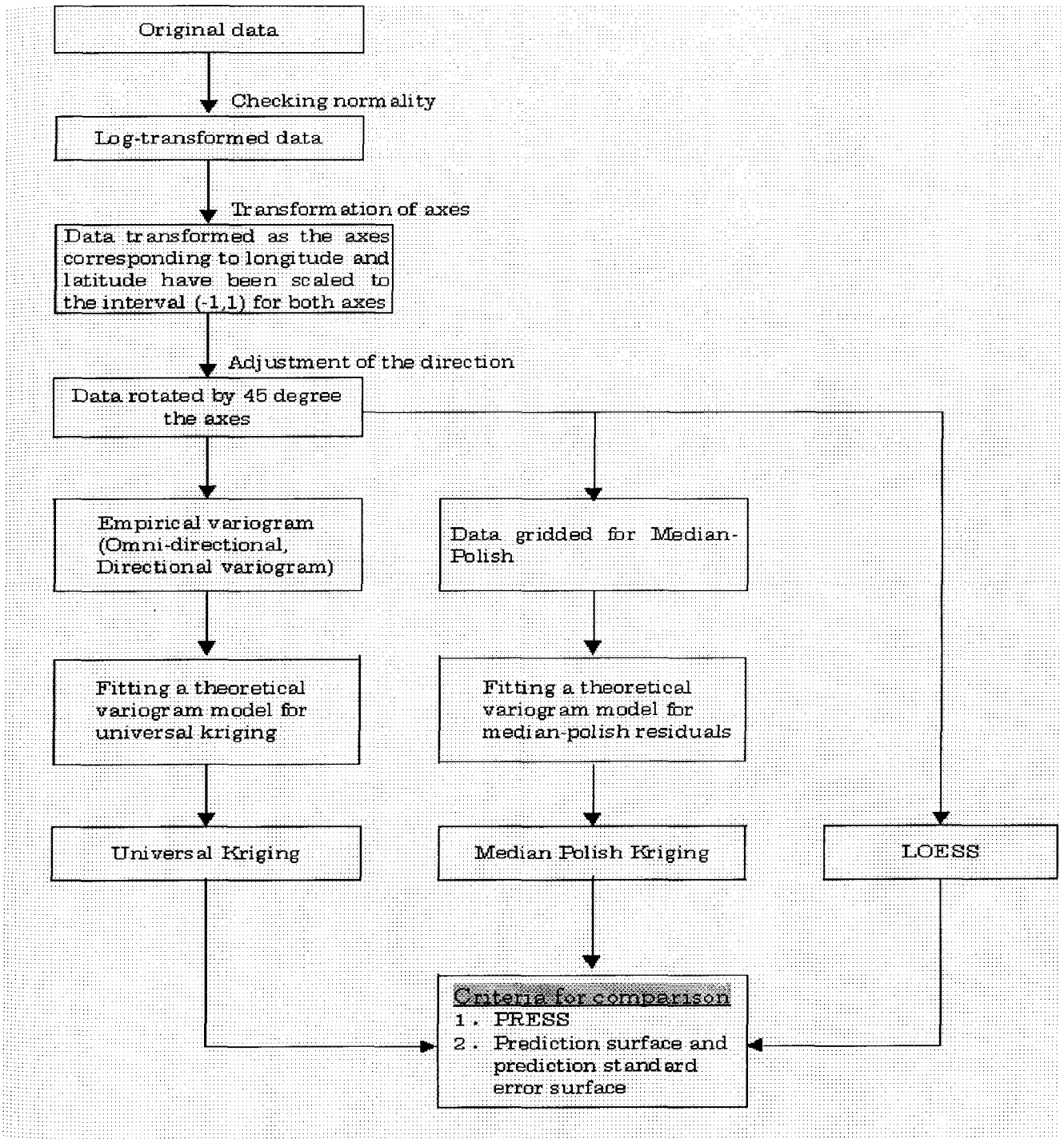

Figure 3 Analysis flow chart

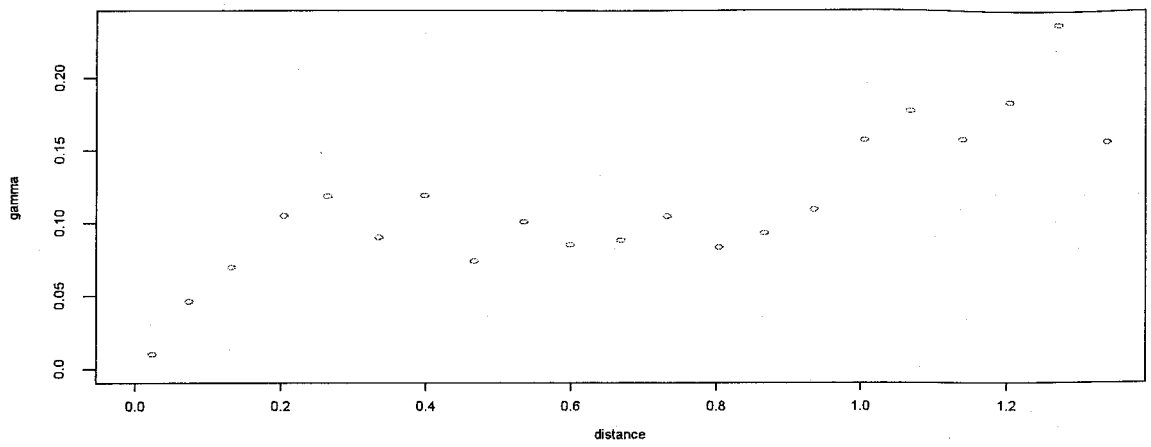

Figure 4 Robust omnidirectional empirical variogram of rainfall data.

To estimate spatial parameters of rainfall data, we fit a theoretical variogram model to the empirical variogram. Fitting a theoretical variogram model to the empirical variogram is often done by eye to decide the initial values of variogram parameters. Because there are some outliers as shown in the `Figure 1`, variogram models are fitted using the robust variogram estimation method developed by Cressie and Hawkins(1980). `Figure 4` shows the sample omnidirectional variogram for the given data set.

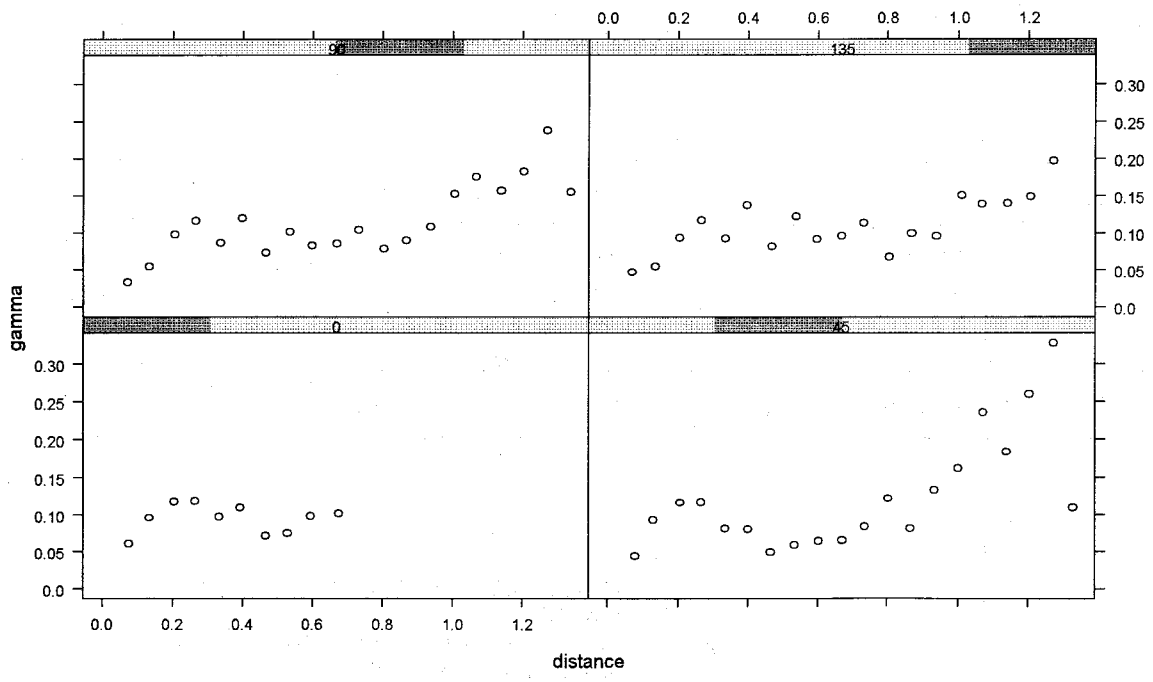

Figure 5 Robust directional empirical variogram for rainfall data.

As shown, the omnidirectional variogram is generally increasing with distance. This suggests that large-scale trend may exist or, in other words, stochastic process may be nonstationary. The directional variogram is based on both the magnitude and direction of **h**. It is calculated for the four principle directions using "azimuth" argument of S-plus pakage. Also, we used the "tol.azimuth" argument setted to 45 so that each directional variogram is based on all pairs of points that fall with the specified azimuth $\pm 45°$(`Figure 5`). In the plot of `Figure 5`, the

direction of the southnorth is represented by 0°, whereas that of the eastwest is represented by 90°. The variograms of the two direction is different. For the most part, the variogram of the southnorth direction is displayed as an even form, indicating little or no autocorrelation. The variograms in the other direction are generally increasing with distance, which could be caused by the existance of trend or anisotropy. To avoid the thorny problems of UK, theoretical variogram is fitted in the direction that has no trend(Cressie, 1984). Accordingly, it is fitted in the southnorth direction. Spherical model is fitted to the estimated variogram(`Figure 6`) and the estimated parameters are $c_0(nugget) = 0.07875724$, $c_s(sill) = 0.028771$ and $a_s(range) = 0.226253$. The estimated spherical model is given by

$$\hat{\gamma}(h; c_0, c_s, a_s) = \begin{cases} 0, & h = 0 \\ 0.07875724 + 0.028771[(\frac{3}{2})\frac{h}{0.226253} - (\frac{1}{2})(\frac{h}{0.226253})^3], & 0 < h \leq 0.226253 \\ 0.07875724 + 0.028771, & h > 0.226253 \end{cases} \tag{25}$$

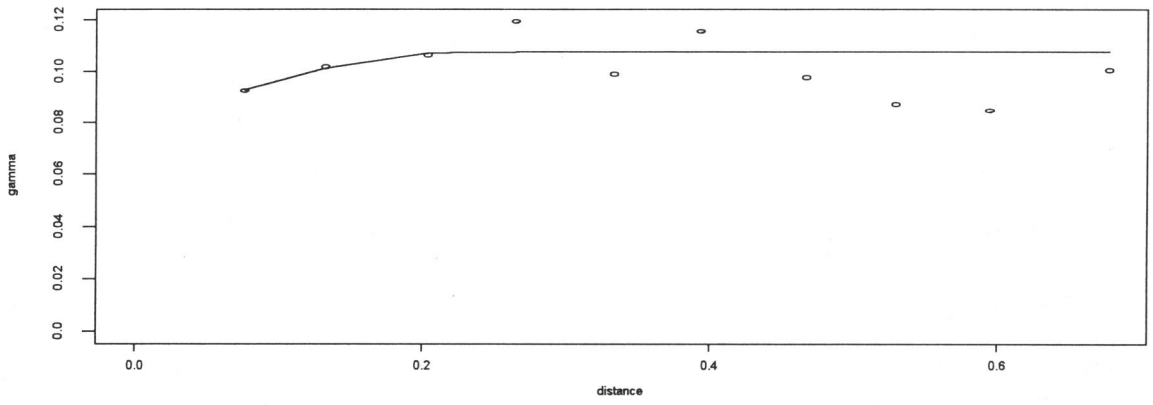

Figure 6 Estimated variogram in the southnorth direction and the superimposed line is the fitted theoretical variogram.

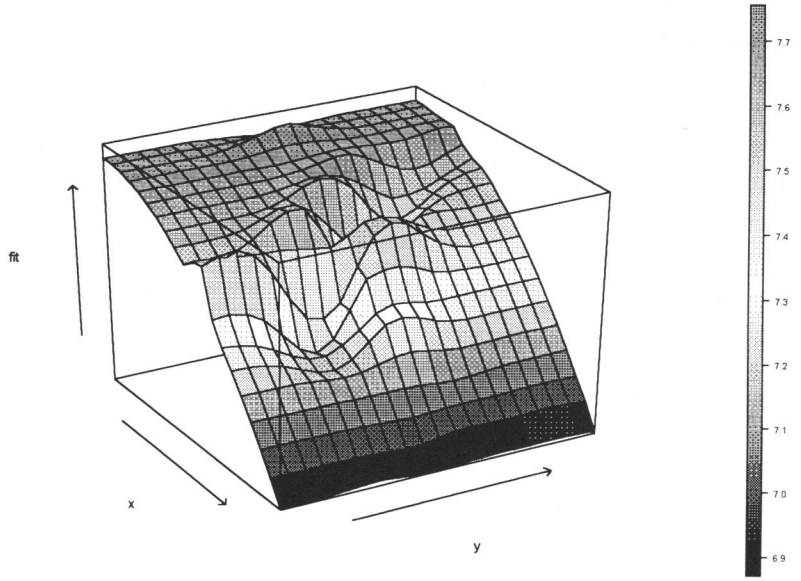

Figure 7 Surface plot of a rainfall based on universal kriging predictions.

The resulting surface plot of the UK obtained by utilizing the fitted variogram model is shown in Figure 7 and the prediction standard errors plot of that is shown in Figure 8. Figure 7 clearly demonstrates that there is a trend in the eastwest direction and Figure 8 shows that the standard error increases as the location is par from the observed points.

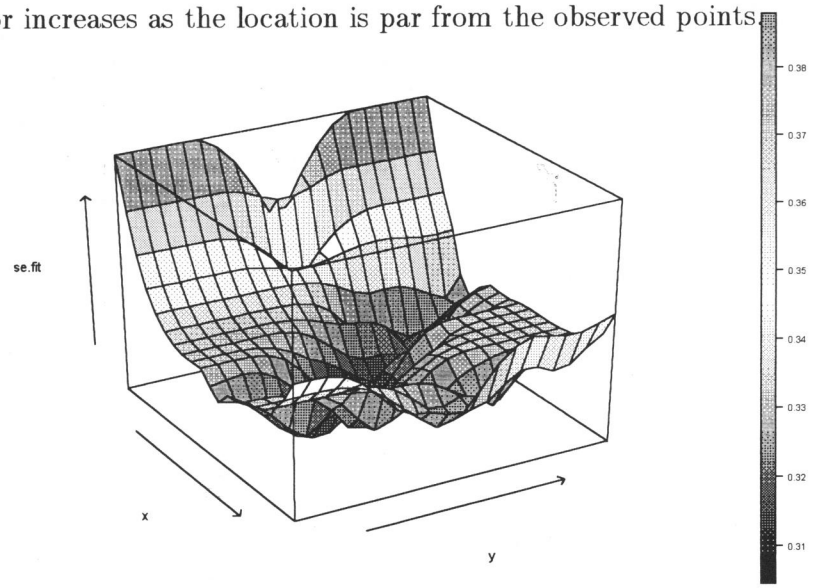

Figure 8 Surface plot of universal kriging prediction standard error.

7.2.2 Median-polish kriging

We have seen the presence of a trend in the eastwest direction from the exploratory data analysis presented in the section 7.1. First of all, we can identify the effect of Median-polish method by seeing the variogram of the results in the eastwest direction as shown in Figure 9. This figure shows that the trend of the original data set is eleminated in the eastwest variogram with the use of the median-polish residuals.

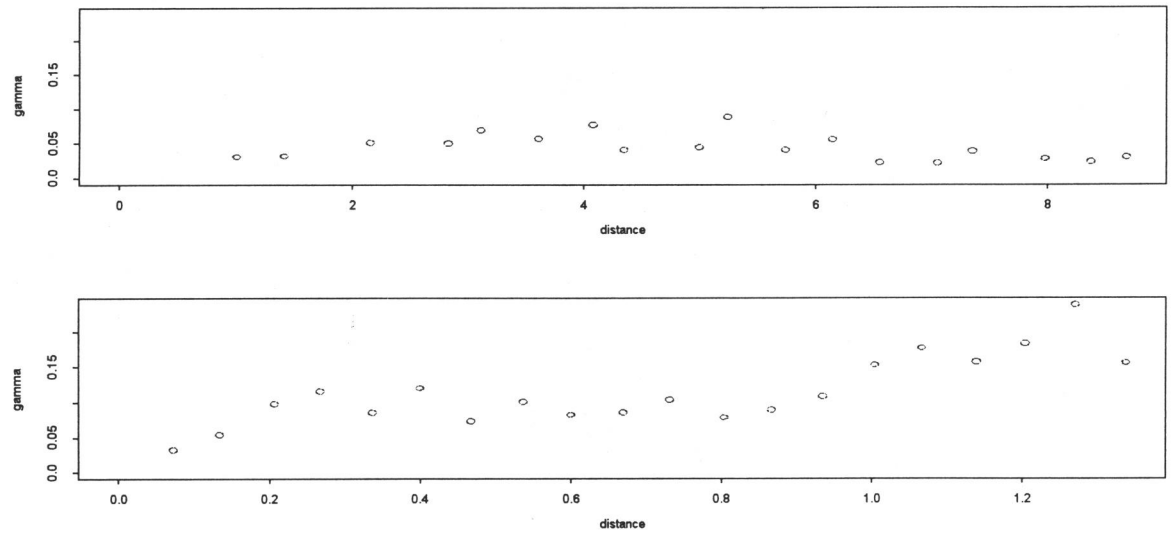

Figure 9 Eastwest variogram of the rainfall data calculated from median-polish residuals(top) and original data(bottom).

Then the variogram models in section 5 are fitted to the empirical variogram of the median polish residuals. The best fitted variogram model, which is shown in `Figure 10`, is a spherical model with estimated parameters; $c_0(nugget) = 0.006358832$, $c_s(sill) = 0.04862538$ and $a_s(range) = 3.168378$. The estimated spherical model is given by

$$\hat{\gamma}(h; c_0, c_s, a_s) = \begin{cases} 0, & h = 0 \\ 0.0063588 + 0.04862538[(\frac{3}{2})\frac{h}{3.168378} - (\frac{1}{2})(\frac{h}{3.168378})^3], & 0 < h \leq 3.168378 \\ 0.0063588 + 0.04862538, & h > 3.168378 \end{cases} \tag{26}$$

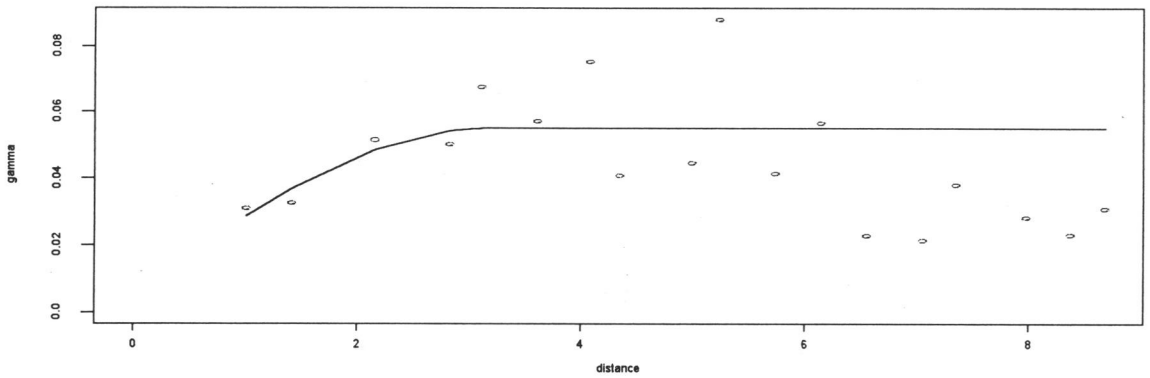

Figure 10 Estimated isotropy variogram based on residuals obteined from MP and the superimposed line is the fitted theoretical variogram.

The median-polish surface is estimated by adding median-polish estimate $\hat{m}(s_0)$ and the predicted value obtained by ordinary kriging of the median-polish residuals. Since the median-polish residuals can be considered as stationary, they can be analyzed by using oridinary kriging. The surface of a MPK is shown in `Figure 11` and it is clear that there is a trend in the eastwest as the surface of UK.

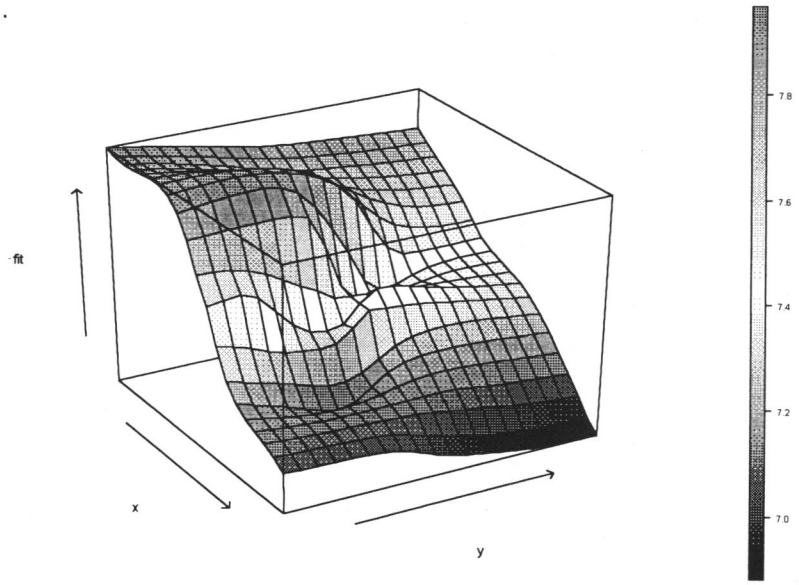

Figure 11 Surface plot of a rainfall based on median polish kriging predictions.

The prediction standard error surface is shown in Figure 12. This figure shows that the standard error is high as far as away from the location of observed points.

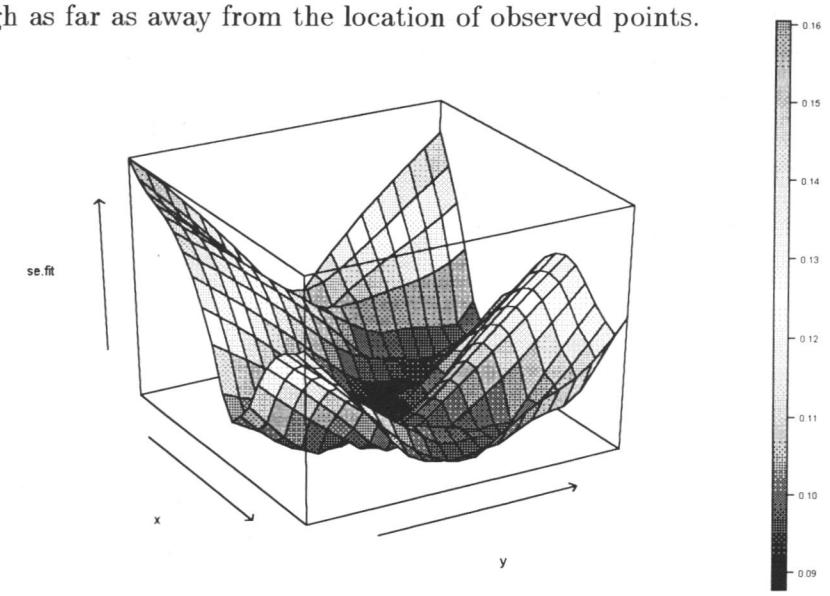

Figure 12 Surface plot of median polish kriging prediction standard error.

7.2.3 LOESS

As explained in section 5, the LOESS is one of traditional methods which fit regression surfaces locally without taking into account the spatial correlation. The resulting surface plot of a LOESS is shown in Figure 13 and its prediction standard errors plot is shown in Figure 14. We can see that prediction surface in Figure 13 is quite different from those of UK and MPK, in particular in the corners where observations are sparse. Also it is noticed that the prediction standard error surface has different shape from those of UK and MPK especially near the boundaries of the region.

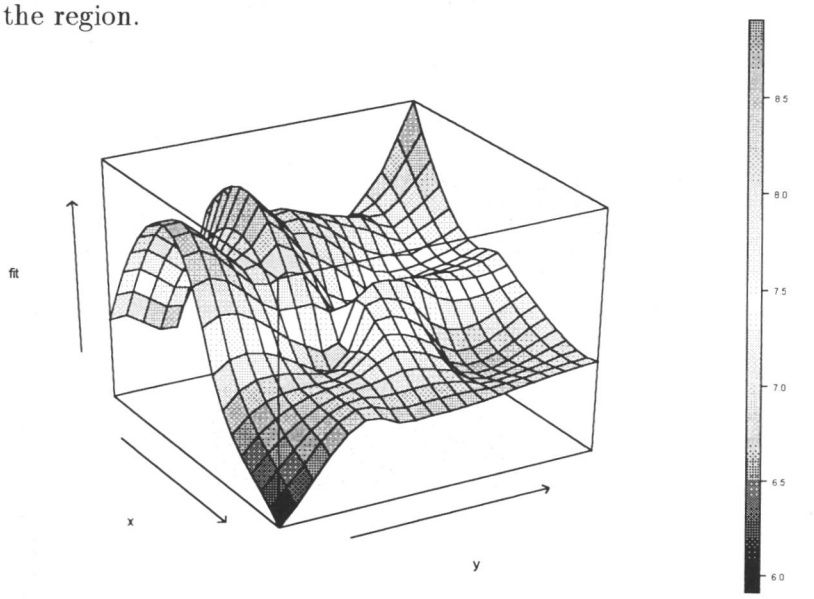

Figure 13 Surface plot of the LOESS prediction.

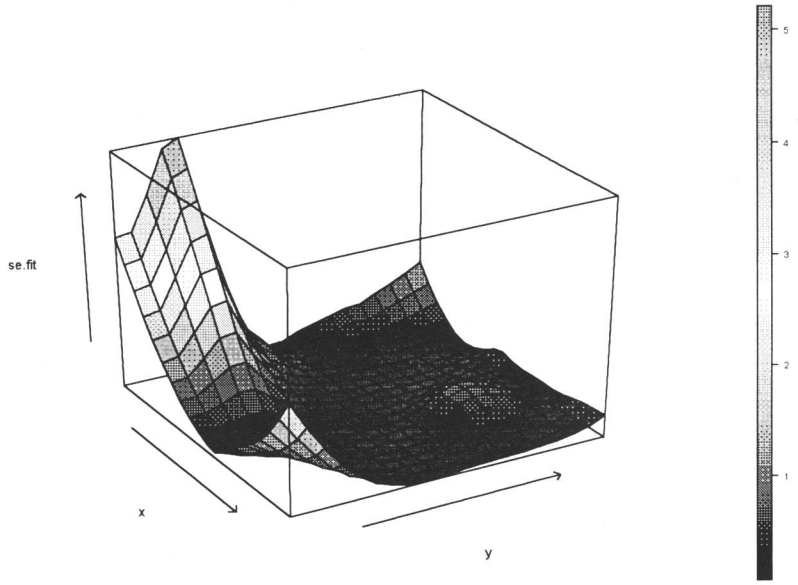

Figure 14 Surface plot of the LOESS prediction standard error.

7.3 Comparison of results

The PRESS(Predicted REsidual Sum of Square) statistic is computed to compare the performances of the three methods. The PRESS values are obtained as the results of predicting each observation, after removing one at a time, using the other observations. Since there are 4 missing values in the data set of 80 observations, values given in the table below are obtained using the other 76 observations.

	UK	MPK	LOESS
PRESS	4.588384	6.028596	6.00258

This table shows that there is no difference between the PRESS values for MPK and LOESS but that the PRESS value of UK is smaller than the others. Therefore, it can be concluded that, as expected, spatial analysis method which takes into account spatial correlation is better than the traditional method which does not take into account spatial correlation. `Figure 15`, `Figure 16`, and `Figure 17` show scatter plots between original observations and their predicted values at the 76 locations. In this scatter plots, it can be thought that as the points being closer to 45° line, its predicting model is performing better. However, we can see that the vertical axis's scale of LOESS is different from those of UK and MPK. Therefore, caution should be given when we interprete about three scatter plots. We may be thought that the result of MPK is similar to that of LOESS in the figure, but the result of UK shows a little difference compared with them. It is difficult to conclude from the figures which model is well performing. However, it might be concluded that UK model is performing better than the other two models. This result coincides with the conclusion based on the PRESS statistic.

It seems that UK and MPK in the fitted prediction surfaces and prediction standard errors surfaces given in the section 7.2 are similar with each other. However, the fitted prediction standard errors of LOESS is generally higher than those of UK and MPK. But the result of numerical analysis is not necessary. The result of LOESS in the fitting of prediction standard

errors may be considered as a result of predicting the value of par position from the observed point.

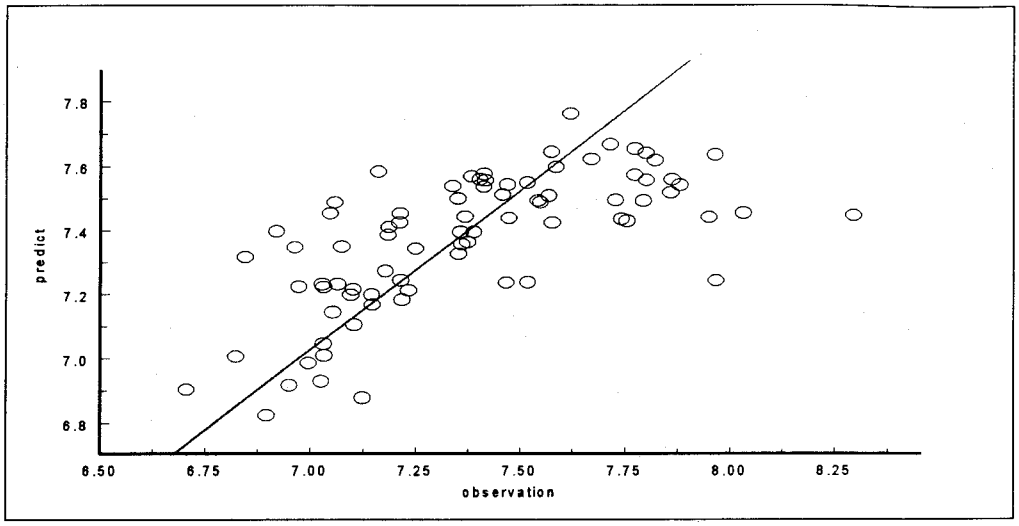

Figure 15 Plot of $(\tilde{Z}_{-i}(s_{0,i}), Z(s_{0,i}))$ for UK.

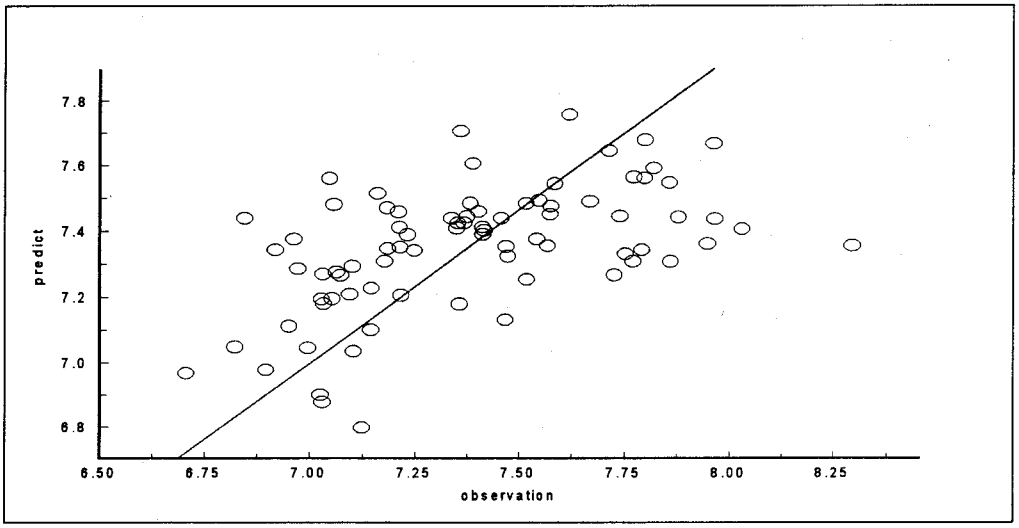

Figure 16 Plot of $(\tilde{Z}_{-i}(s_{0,i}), Z(s_{0,i}))$ for MPK.

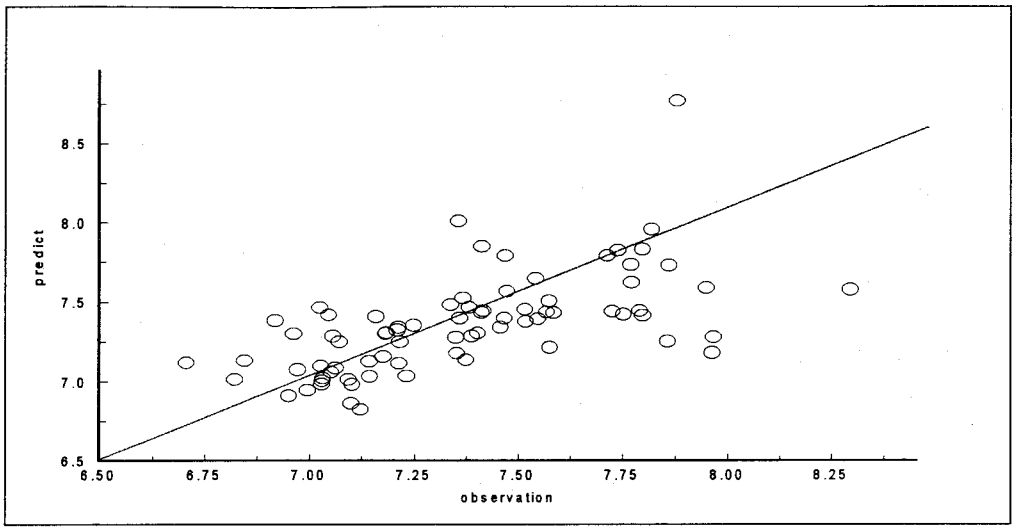

Figure 17 Plot of $(\tilde{Z}_{-i}(s_{0,i}), Z(s_{0,i}))$ for LOWESS.

8 Discussion

In this study, non-stationary data was analyzed with UK and MPK as spatial analysis method and LOESS as a traditional analysis method. The results of these analyses show that the method which accomodates spatial correlation structure performs better than any other method which does not accomodate such spatial correlation structure. However, this is not always the case. The results may depend on the selected grid pattern of the observation data. Although the variogram estimation is the Keystone in any spatial analysis, it is not given much attention in this study. Caution should be given when dealing with such spatial analysis.

References

[1] Cleveland, W. S.(1979), "Robust locally Weighted Regression and Smoothing Scatterplots," *Journal of the American Statistical Association* Vol. 74, No. 368, pp. 829-836.

[2] Cleveland, W. S. and Devlin, S. J.(1988), "Locally Weighted Regression: An approach to regression analysis by local fitting," *Journal of the American Statistical Association* Vol. 83, No. 403, pp. 596-610.

[3] Cressie, N. A. C.(1986), "Kriging Nonstationary Data," *Journal of the American Statistical Association* Vol. 81, No. 395, pp. 625-634.

[4] Cressie, N. A. C.(1991), "Statistics for Spatial Data," New York:*John Wiley and Sons.*

[5] Cook, D. G. and Pocock, S. J.(1983), "Multiple Regression in Geographical Mortality Studies with Allowance for Spatially Correlated Errors," *Biometrics*, Vol. 39, No. 2, pp. 361-371.

[6] Cothern, C. R. and Ross N. P.(1994), "Environmental Statistics, Assessment and Forecasting," *CRC Press.*, pp. 131-146.

[7] Deutsch, C. V. and Journel, A. G.(1998), "Applied Geostatistics Series(GLSIB:Geostatistical Software Library and User's Guide)," New York:*Oxford University Press.*

[8] Donnelly, C. A., Ware, J. H. and Laird, N. M.(1992), "Regression Analysis of Spatially Correlated Data : The Kanawha Country Health Study," *Handbook of Statistics*, Vol. 12. pp. 643-659.

[9] Isaaks, E. H. and Srivastava, R. M.(1989), "An Introduction to Applied Geostatistics," New York:*Oxford University Press.*

[10] Kafadar, K.(1994), "Choosing among two-dimensional smoothers in practice," *Computational Statistics and Data Analysis*, Vol. 18, pp. 419-439.

[11] Mardia, K. V. and Marshall, R. J.(1984), "Maximum Likelihood Estimation of Models for Residual Covariance in Spatial Regression," *Biometrika*, Vol. 71, pp. 135-146.

[12] Mardia, K. V. and Watkins, A. J.(1989), "On Multimodality of the Likelihood in the Spatial Linear Model," *Biometrika*, Vol. 76, No. 2, pp. 289-295.

[13] Richard, F. G.(1995), "Estimating Spatial Correlations from Spatial-Temporal Meteorological Data," *Journal of Climate*, Vol. 8, No. 10, pp. 2454-2470.

[14] Stein, A. and Corsten, L. C. A.(1991), "Universal Kriging and Cokriging as a Regression Procedure," *Biometrics*, Vol. 47, No. 2, pp. 575-587.

[15] Venables, W. N. and Ripley, B. D.(1994), "Modern Applied Statistics with S-plus," *Springer-Verlag*, pp. 383-396.

[16] Wackernagel, Hans(1995), "Multivariate Geostatistics:An Introduction with Application," *Springer-Verlag.*

[17] Warnes, J. J. and Ripley, B. D.(1987), "Problem with Likelihood Estimation of Covariance Functions of Spatial Gaussian Processes," *Biometrika*, Vol. 74, pp. 640-42.

[18] Christensen, R.(1991), "Linear Model for Multivariate, Time Series and Spatial Data," *Springer-Verlag.*

Appendix A: Original data

obs	longitude	latitude	rainfall	obs	longitude	latitude	rainfall
1	141.41	45.25	1123.8	41	140.52	35.44	1557.6
2	143.38	43.57	1239.4	42	136.31	34.44	1654.8
3	142.22	43.46	1090.8	43	137.43	34.42	1884.0
4	144.17	44.07	815.4	44	138.24	34.58	2326.9
5	141.20	43.03	1129.6	45	139.46	35.41	1405.3
6	143.13	42.55	917.2	46	136.12	34.04	4001.9
7	144.24	42.59	1042.6	47	139.39	35.26	1568.9
8	145.35	43.20	987.5	48	139.22	34.45	2831.1
9	140.14	42.48	1214.0	49	139.47	33.06	3073.2
10	142.47	42.10	1131.5	50	133.20	36.12	1751.0
11	140.45	41.49	1155.0	51	133.04	35.27	1894.8
12	140.46	40.49	1360.6	52	134.14	35.29	1949.5
13	140.06	39.43	1746.4	53	132.04	34.54	1730.6
14	141.10	39.42	1265.4	54	135.44	35.01	1581.1
15	141.58	39.39	1267.4	55	136.15	35.16	1653.7
16	139.51	38.54	1839.5	56	130.56	33.57	1659.9
17	140.21	38.15	1126.3	57	132.28	34.24	1554.6
18	140.54	38.16	1204.4	58	133.55	34.39	1159.7
19	140.29	37.45	1065.8	59	135.11	34.41	1315.5
20	140.54	36.57	1356.8	60	135.31	34.41	1318.0
21	136.54	37.23	2264.8	61	135.10	34.14	1352.6
22	138.15	38.02	1563.2	62	135.46	33.27	2640.9
23	139.03	37.55	1178.4	63	135.50	34.41	1354.6
24	136.38	36.35	2592.4	64	129.18	34.12	2139.2
25	137.12	36.42	2296.1	65	130.23	33.35	1604.3
26	138.12	36.40	938.3	66	130.18	33.16	1836.4
27	138.15	37.06	2880.4	67	131.37	33.14	1637.5
28	139.52	36.33	1382.3	68	129.52	32.44	1945.3
29	136.14	36.03	2368.3	69	130.43	32.49	1967.7
30	137.15	36.09	1756.8	70	130.33	31.33	2236.8
31	137.58	36.15	1010.5	71	131.25	31.55	2434.6
32	138.33	36.20	1211.3	72	128.50	32.42	2372.0
33	139.04	36.24	1130.2	73	132.47	33.50	1286.0
34	139.23	36.09	1167.5	74	134.03	34.19	1147.2
35	140.28	36.23	1307.8	75	133.33	33.34	2582.4
36	136.04	35.39	2418.9	76	134.35	34.04	1614.6
37	136.46	35.24	1933.8	77	133.01	32.43	2487.7
38	136.58	35.10	1535.0	78	134.11	33.15	2435.5
39	137.50	35.31	1591.6	79	129.30	28.23	2870.7
40	138.33	35.40	1055.0	80	127.41	26.12	2036.8