Modified Median Polish Kriging and its Application to the Wolfcamp-Aquifer Data

Olaf Berke

Institute of Biometry, Epidemiology and Information Processing,
School of Veterinary Medicine Hannover, D-30559 Hannover, Germany
email: olaf.berke@tiho-hannover.de

SUMMARY. In geostatistics, spatial data will be analysed that often come from irregularly distributed sampling locations. Interest is in modelling the data, i.e. estimating distributional parameters, and then to predict the phenomenon under study at unobserved sites within the corresponding sampling domain. The method of universal kriging for spatial prediction was introduced to cover the problem of spatial trend effects. This is done by incorporating linear trend models, e.g. polynomial functions of the spatial co-ordinates. However, universal kriging is sensitive to additive outliers. An outlier resistant method for spatial prediction is median polish kriging. Both methods have certain advantages but also some drawbacks. Here, universal kriging and median polish kriging will be combined to the robust spatial prediction method called modified median polish kriging. An example illustrates the method of modified median polish kriging along with piezometric-head data from the Wolfcamp-Aquifer.

KEY WORDS: Geostatistics, Robust Methods, Spatial Prediction, Trend Estimation.

1. INTRODUCTION

Geostatistics was developed to analyse geological and meteorological data in the sixties (Cressie, 1990a). Now the field of applications for geostatistics has wide spread to such diverse areas like Hydrogeology (Kitanidis, 1997), Biology (Ecker and Heltshe, 1994), Pedometry (Burrough *et al.*, 1994), and Epidemiology (Webster *et al.*, 1994) to name a few.

Especially in environmetrics the geostatistical methods for the analysis of spatially continuous phenomena have become standard analytical tools. Many applications have been published in the literature, recent examples include the mapping of PM_{10} exposure in the Pittsburgh region (Cressie, 2000), spatial prediction of dynamic factor loadings from spatial time series of groundwater levels in the Transdanubian Mountains of Hungary (Márkus *et al.*, 1999), modelling and prediction of acid deposition in Lower Saxony (Berke, 1999), and modelling the nitrogen pollution occurrence by nitrogen runoff into surface water in the Midwest and Northern Plains region (Opsomer *et al.*, 1999).

Kriging is the optimal geostatistical prediction method to interpolate and extrapolate observed spatial data $z(s)$ from regularly or irregularly located sample sites $s_1, ..., s_n$ within a certain survey area $\mathcal{D}$. These data are modelled to be a partial realisation from a spatial process or random field $\mathcal{Z} = \{Z(s) : s \in \mathcal{D}\}$. Kriging is based on accurately estimating the spatial dependence structure. This is generally achieved using certain types of spatial structure models (i.e. the exponential or spherical among other models) that will be fit to an estimated structure function (e.g. the semivariogram or covariogram). Then, the spatial dependence is used to optimally weight the observations for predicting the phenomenon under study at an unobserved site, say s. For mapping purposes, i.e. isopleth maps or three-dimensional surface plots, the prediction location s will be moved over a collection of knots forming a fine grid over the survey area $\mathcal{D}$.

However, estimation of the spatial dependence structure presumes that the phenomenon under study is ergodic (Cressie, 1993, p. 53). Amongst others, the mean structure must be constant over $\mathcal{D}$. This is rarely the case in practical applications. So, generally some kind of trend elimination is used before estimation of the spatial dependence structure. An exception is intrinsic random function modelling (Matheron, 1973; Cressie, 1993, p. 299). This method is based on filtering out the spatial trend and is the spatial equivalent to integrated time series modelling (Box and Jenkins, 1970). There are plenty non-parametric methods for smoothing or estimation and extraction of the spatial trend out of the data, e.g. splines (Wahba, 1990) or locally weighted regression (Cleveland and Devlin, 1988). The mainly used geostatistical methods for detrending are the parametric method of trend surface estimation and the non-parametric median polishing approach. These result in the prediction methods of universal kriging (Matheron, 1969; Cressie, 1993, p. 151) and median polish kriging (Cressie, 1986), respectively.

The advantages of kriging in predicting spatial data over competing methods, such as spline smoothing, are given by theory and practice. Statistical theory states that (under the condition that an adequate model is fit to the data) the universal kriging predictions are optimal with respect to the squared loss function. On the other hand, several comparative studies showed that universal kriging outperforms other interpolators and never performs worse (Laslett, 1994; Brus et al., 1996).

Furthermore, unlike most of its competitors, kriging allows to quantify the accuracy of the predictions by means of the mean squared prediction error. But, the estimate of the mean squared prediction error relies on the estimate of the structure function and, thus, on an accurate choice of the trend model.

The paper is outlined as follows. In Section 2, the spatial prediction methods of universal kriging and median polish kriging are summarised and compared to certain properties. Then, Section 3 outlines the method of modified median polish kriging. The piezometric-head data set of the Wolfcamp-Aquifer is analysed by use of different kriging methods in Section 4. The main ideas and results are brought together and discussed in Section 5.

2. SPATIAL PREDICTION: KRIGING

There are many different kriging methods depending on certain model assumptions. For completeness, the methods of universal kriging (UK) and median polish kriging (MPK) will be reviewed in this Section.

Behind the spatial analysis is an assumption that data $\mathbf{Z} = (Z(s_1), ..., Z(s_n))'$

observed at known spatial locations $\{s_1, ..., s_n\} \in \mathcal{D}$ are a realisation from a spatial process $\mathcal{Z} = \{Z(s) : s \in \mathcal{D}\}$ that could be decomposed

$$Z(s) = \mu(s) + \delta(s), \qquad s \in \mathcal{D}. \tag{1}$$

Here, $\mu(\cdot)$ denotes the large scale variation or deterministic spatial mean component of the process, and $\delta(\cdot)$ is the small scale variation or stochastic residual component that accounts for all the spatial dependence and random fluctuations, i.e.

$$\begin{aligned} E(Z(s)) &= \mu(s), \\ E(\delta(s)) &= 0, \\ Cov(\delta(s), \delta(u)) &= \sigma(s, u), \qquad s, u \in \mathcal{D}. \end{aligned}$$

The decomposition of data into large and small scale variation is not unique and depends on how much variation is attributed to the trend or the residual component. For example, someone may have knowledge about the phenomenon under study and wants to explain as much of the variation through the trend component $\mu(\cdot)$ as possible. Someone else, on the other hand, follows the principle of parsimony and prefers smaller or less complex models. The latter approach results in larger residuals, but is allied with smaller statistical errors from unknown model parameter estimations. In the end, the choice of the model to be fit to the data is subjective and may be more important than the choice of the fitting algorithm.

For simple kriging (Cressie, 1993, p. 359), the large scale variation $\mu(s)$ is assumed to be known at any location $s \in \mathcal{D}$. This is, however, unlikely in practice, so for UK the large scale variation is assumed to be an unknown linear combination of known functions of exploratory variables $\mu(s) = \sum_j^{p+1} x_{j-1}(s)\beta_{j-1}$. On the other hand, MPK allows for an unknown and variable spatial mean as well. The mean is not explained externally through exploratory variables but rather than by variation of the data itself. This is done via a main effect and several row and column effects, which will be introduced later in Section 2.2. Unlike UK, MPK is based on an outlier resistant estimation of the spatial trend. In addition, UK is statistically optimal but operationally more difficult than MPK.

Both methods give predictions that do not differ drastically, but the mean squared prediction errors do. Indeed, an important distinction between UK and MPK is that in practise the MSPE for UK is estimated, whereas for MPK it is approximated. Lastly, the frequently used polynomial trend models in UK are not flexible enough to model local characteristics of the mean function. On the other hand, median polishing may result in an over fitting of the spatial mean function, making the median polish surface rough, and, its artificial appearence can be visually extracted from the prediction map, compare the Figures 2 and 3 latter in this article.

2.1 Universal Kriging

Matheron (1969) introduced UK to overcome the problem associated with spatially predicting a process influenced by a trend. This is achieved by modelling the trend via a linear combination of functions of explanatory variables associated with the sampling locations $s_1, ..., s_n$ and any prediction location $s \in \mathcal{D}$ under study. Generally, the trend is given by low order polynomials in the spatial co-ordinates. However, explanatory variables, other than the co-ordinates are often useful. See Cressie and Majure (1997) for an example. For this purpose, the spatial process $\mathcal{Z}$ is modelled by the two component model (1) with linear trend component

$$\mu(s) = \sum_{j=1}^{p+1} x_{j-1}(s)\beta_{j-1}.$$

The covariogram $\sigma(h, \theta)$, where h measures the spatial translation between two points in $\mathcal{D}$, models the spatial dependence structure exhibited by the residual process and is generally estimated by use of the semivariogram. To use the linear model theory to receive the best linear unbiased predictor (BLUP), i.e. the UK predictor, the spatial process $\mathcal{Z} = \{Z(s) : s \in \mathcal{D}\}$ is put into the form of a spatial linear model. The spatial linear model for the sample variable vector is of the form

$$\mathbf{Z} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\delta}, \qquad \boldsymbol{\delta} \sim \mathcal{N}(\mathbf{0}, \, \boldsymbol{\Sigma}(\boldsymbol{\theta})), \tag{2}$$

where $\mathbf{X}$ is an $n \times (p+1)$ matrix whose $(i,\, j)$th element is $x_{j-1}(s_i)$, $\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta_p)'$ is the unknown vector of trend parameters and $\boldsymbol{\delta} = (\delta(s_1), ..., \delta(s_n))'$ is the vector of the residual process at the sampling locations. Furthermore, the spatial linear model is assumed to hold for all sites $s \in \mathcal{D}$

$$Z(s) = \mathbf{x}'\boldsymbol{\beta} + \delta, \qquad \delta \sim \mathcal{N}(0,\, \sigma^2), \tag{3}$$

where $\mathbf{x}$ is a p-vector of functions of the explanatory variables connected with site s and $\delta = \delta(s)$ is the corresponding residual with variance σ^2 determined from the covariogram or the sill value of the semivariogram, respectively: $\sigma^2 = \sigma(\mathbf{0}) = \gamma(\infty)$.

Within the setup of the spatial linear model (2) and (3), the trend surface estimator is given by

$$\widehat{\mu}(s) = \mathbf{x}'\widehat{\boldsymbol{\beta}}, \qquad s \in \mathcal{D},$$

based on the general least squares estimator of the unknown trend parameter

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Z},$$

with $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\theta)$ denoting the covariance matrix of the sample vector based on knowledge of spatial structure parameter $\boldsymbol{\theta}$. Then, the UK predictor is expressed by

$$\widehat{Z}_{uk}(s) = \mathbf{x}'\widehat{\boldsymbol{\beta}} + \boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}(\mathbf{Z} - \mathbf{X}\widehat{\boldsymbol{\beta}}), \qquad s \in \mathcal{D},$$

where $\boldsymbol{\sigma} = (\sigma(s - s_1), ..., \sigma(s - s_n))'$. The mean squared prediction error of the UK predictor $\widehat{Z}_{uk}(s)$, also called kriging variance which depends only on the sampling network configuration and the spatial dependence structure, is given for any site $s \in \mathcal{D}$ by

$$\begin{aligned} MSPE(Z(s), \widehat{Z}_{uk}(s)) &= \sigma^2_{uk}(s) \\ &= \sigma^2 - \boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma} + (\mathbf{x} - \mathbf{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma})'(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}(\mathbf{x} - \mathbf{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}). \end{aligned}$$

A special case of UK is called ordinary kriging (OK) and is concerned with spatial prediction in the case of an unknown but constant spatial mean structure, i.e.

$$\mu(s) \equiv \mu, \qquad s \in \mathcal{D}.$$

OK is used in connection with the MPK approach that is reviewed in the following subsection.

2.2 Median Polish Kriging

Cressie (1986) introduced the MPK method which is based on the median polish approach for two-way tables (Tukey, 1977; Emerson and Hoaglin, 1983) to extract a mean surface from spatial data and then to use OK for spatially predicting the residual process. MPK is described in more detail in Cressie (1993), along with applications to such different types of data like regular or irregular distributed geostatistical data as well as regional count data. There are two advantages of MPK over UK. Firstly, the mean component of model (1) is estimated by the outlier resistant method of median polishing, which gives less biased residuals for estimating the structure function (Cressie and Glonek, 1984). And secondly, the spatial structure function, i.e. the variogram, is not assumed to be known a priori.

In this work, MPK with irregularly distributed geostatistical data is achieved as follows. With respect to the network of sample sites define a $p \times q$ rectangular grid to be laid over the investigation area. Furthermore, the sample data $Z(s_i)$, $i = 1, ..., n$, are allied to the closest grid knots $s_{kl}, k = 1, ..., p$, $l = 1, ..., q$, and the coordinates of the sample sites s_i are converted to that of the corresponding gridlines s_{kl}. For multiple data allied with a single grid knot, the data is replaced by an appropriate summary, i.e. the median. Now, the edited sample data, say $Y(s_{kl})$, have the structure of a two-way table and are suitable for median polishing.

Median polishing proceeds by repeated extraction of the row and column medians until convergence, with respect to a stopping criterion to be chosen by the investigator. A mathematical formulation of the median polish algorithm is given in Cressie *et al.* (1990). For data $\{y(s_{kl}) : k = 1, ..., p, \, l = 1, .., q, \, s_{kl} \in \mathcal{D}\}$, the median polish algorithm gives as an estimate of the mean component $\mu_y(\cdot)$,

$$\widehat{\mu}_y(s_{kl}) = \widehat{a} + \widehat{r}_k + \widehat{c}_l,$$

where $\widehat{a}$ is an over all effect, $\widehat{r}_k$ is the kth row effect and $\widehat{c}_l$ is the lth column effect.

After having obtained the median polish estimates of the spatial trend component, then some kind of spatial inter- or extrapolation is needed to calculate mean component estimates at the sampling locations as well as at desired prediction locations, i.e. grid knots of a fine meshed regular lattice. Cressie (1986) proposes spline smoothing for this purpose. Suppose an estimate is needed at location $s = (u, v)$ anywhere inside the region bounded by the grid lines, say $s = (u, v)$ with $u_k \le u < u_{k+1}$ and $v_l \le v < v_{l+1}$, where $k \in \{1, ..., p-1\}$ and $l \in \{1, ..., q-1\}$. Then, the interpolation of the spatial trend is given by splines

$$\widehat{\mu}_{mp}(s) \;=\; \widehat{a} + \widehat{r}_k + \left(\frac{u - u_k}{u_{k+1} - u_k}\right)(\widehat{r}_{k+1} - \widehat{r}_k) + \widehat{c}_l + \left(\frac{v - v_l}{v_{l+1} - v_l}\right)(\widehat{c}_{l+1} - \widehat{c}_l).$$

For sites outside the region bounded by the grid lines, extrapolation of the spatial trend estimates are needed. This is done by use of the following extrapolating splines. For $s = (u, v)$ with $u < u_1$ and $v_1 \le v_l \le v < v_{l+1} \le v_q$, calculate

$$\widehat{\mu}_{mp}(s) \;=\; \widehat{a} + \widehat{r}_k + \left(\frac{u - u_k}{u_{k+1} - u_k}\right)(\widehat{r}_{k+1} - \widehat{r}_k) + \widehat{c}_1 + \left(\frac{v - v_1}{v_2 - v_1}\right)(\widehat{c}_2 - \widehat{c}_1).$$

For $s = (u, v)$ with $u_1 \le u_k \le u < u_{k+1} \le u_p$ and $v < v_1$, calculate

$$\widehat{\mu}_{mp}(s) \;=\; \widehat{a} + \widehat{r}_1 + \left(\frac{u - u_1}{u_2 - u_1}\right)(\widehat{r}_2 - \widehat{r}_1) + \widehat{c}_l + \left(\frac{v - v_l}{v_{l+1} - v_l}\right)(\widehat{c}_{l+1} - \widehat{c}_l).$$

For $s = (u, v)$ with $u < u_1$ and $v < v_1$, calculate

$$\widehat{\mu}_{mp}(s) \;=\; \widehat{a} + \widehat{r}_1 + \left(\frac{u - u_1}{u_2 - u_1}\right)(\widehat{r}_2 - \widehat{r}_1) + \widehat{c}_1 + \left(\frac{v - v_1}{v_2 - v_1}\right)(\widehat{c}_2 - \widehat{c}_1).$$

Similar extrapolation formulas apply to the cases where $u > u_p$ and/or $v > v_q$.

Lastly, the MPK prediction is the sum of the median polish surface plus the surface obtained from OK of the residual process $\delta(s)$, $s \in \mathcal{D}$:

$$\widehat{Z}_{mpk}(s) = \widehat{\mu}_{mp}(s) + \widehat{\delta}_{ok}(s),$$

where

$$\widehat{\delta}_{ok}(s) = \widehat{\beta}_\delta + \boldsymbol{\sigma}_\delta' \boldsymbol{\Sigma}_\delta^{-1}(\boldsymbol{\delta} - \mathbf{1}\widehat{\beta}_\delta), \qquad s \in \mathcal{D}$$

$$\boldsymbol{\delta} = \mathbf{Z} - \widehat{\mu}_{mp}$$

are the OK predictions of the residual process and the estimated trend residuals, respectively. Furthermore,

$$\widehat{\beta}_\delta \quad = \quad (\mathbf{1}'\boldsymbol{\Sigma}_\delta^{-1}\mathbf{1})^{-1}\mathbf{1}'\boldsymbol{\Sigma}_\delta^{-1}\boldsymbol{\delta}$$

denotes the GLSE of the constant mean for the residual process $\delta(\cdot)$ based on the spatially structured covariance matrix of the residual vector $\boldsymbol{\Sigma}_\delta$.

Lastly, the MSPE for the MPK preditctor is approximated by the MSPE of the OK predictor for the residual process, i.e.:

$$\begin{aligned} &MSPE(Z(s), \widehat{Z}_{mpk}(s)) \approx MSPE(\delta(s), \widehat{\delta}_{mpk}(s)) \\ &\quad = \sigma^2 - \boldsymbol{\sigma}_\delta'\boldsymbol{\Sigma}_\delta^{-1}\boldsymbol{\sigma}_\delta + (1 - \boldsymbol{\sigma}_\delta'\boldsymbol{\Sigma}_\delta^{-1}\mathbf{1})(\mathbf{1}'\boldsymbol{\Sigma}_\delta^{-1}\mathbf{1})^{-1}(1 - \boldsymbol{\sigma}_\delta'\boldsymbol{\Sigma}_\delta^{-1}\mathbf{1})', \qquad s \in \mathcal{D}. \end{aligned}$$

3. MODIFIED MEDIAN POLISH KRIGING

The method of modified median polish kriging (MMPK) is a modification of the MPK method introduced by Cressie (1986) and outlined above. The modification is concerned with the estimation of the mean function, i.e. the mathematical spline interpolation is replaced by the geostatistical approach of UK.

In MMPK, the spatial mean function $\mu(\cdot)$ is again estimated in a first step by the median polish procedure

$$\widehat{\mu}_y(s_{kl}) \quad = \quad \widehat{a} + \widehat{c}_k + \widehat{r}_l,$$

where the estimation locations s_{kl} are the grid knots of the two-way table layout connected with the network of sample sites, and the index Y is used to indicate the case where multiple data are connected to the same grid knot and hence, are replaced by the corresponding median.

These mean estimates $\widehat{\mu}_y(s_{kl})$ give a low resolution map of the mean function. To predict the phenomenon under study to unobserved sites, a high resolution map based on a much finer grid is needed. Therefore, some kind of interpolation technique is needed in a second step to produce a high resolution map from the low resolution map. In MPK, this is done by use of (extra- and) interpolating splines of degree 1. The modification proposed here is to just use UK instead. This gives a smooth and robust estimate of the mean function $\widehat{\mu}_{mmp}(s)$ at any location s in the investigation area $\mathcal{D}$. This is due to the fact that spline technology is built for surfaces that are "deterministic" or "deterministic plus white noise" (Cressie, 1990b), but from model (1), it is clear that the residual component is spatially correlated. Therefore, kriging will be superior to splining in this case.

Doing this improves the UK and MPK methods as follows:

- The mean function has a parametric interpretation alike to UK but unlike MPK, which is a useful feature for example in spatiotemporal modelling approaches (Berke, 1998).

- The mean function is smoother than with the original MPK approach, because it avoids the artificial angular shape of the estimated mean surface and the prediction surface but retains its local flexibility which outperforms low order polynomial trend surfaces used with UK.

- The mean function is an outlier resistant algorithm alike to MPK, since it is based on the median polish estimates of $\mu(\cdot)$ at the low resolution grid knots. If the structure function is modelled without nugget effect, the trend surface estimates from MPK are equal to the MMPK estimates at the low resolution grid knots, since UK is an exact interpolation method in this case.

- In contrast to UK, the MMPK approach avoids problems arising in connection with poor network design, i.e. sampling sites that are bad leverage points (Martin, 1992).

After the short explanation of the modification of MPK to MMPK and the resulting benefits, the method of MMPK is now outlined step by step.

Firstly, median polish the data. Therefore, define a grid $\mathcal{G}$ with possibly unequal spacings to which the sample variables $Z = (Z(s_1), ..., Z(s_n))'$ are connected. For regular networks, the grid is given in a natural way. For irregular networks, the resolution of the grid and the grid orientation are choices to be made. Cressie (1993, p. 193) recommends to maintain the same co-ordinate directions as the original map and to use a grid with approximately one observation per grid knot if possible. However, one should avoid empty cells in the two-way table and take care of anisotropical behaviour of the phenomenon. If there is more than one observation per grid knot, calculate the median of this data and use the medians along with the median polish algorithm. This means, the sample $Z(s_i)$, $i = 1, ..., n$, will be replaced by $Y(s_{kl})$, $k = 1, ..., p$, $l = 1, ..., q$. Next, choose a stopping criterion for median polishing. Perform median polish to receive the overall, row and column effects. Check whether the residuals contain cross-product trend, i.e. row and column interaction. Calculate the mean effects at the grid knots, i.e.

$$\widehat{\mu}_y(s_{kl}) = \widehat{a} + \widehat{r}_k + \widehat{c}_l, \qquad s_{kl} \in \mathcal{G}$$

Secondly, predict the median polish estimates by UK onto a fine grid. For this, take median polish estimates as new observations and calculate the corresponding spatial structure model (for example a semivariogram $\gamma_y(h;\theta)$). Now, choose a fine meshed grid $\mathcal{M}$, fine enough for the mapping purposes. Identify an appropriate linear trend model $E(Y(s)) = X\beta_y$ and perform UK with the identified trend model $X\beta_y$ and structure model $\gamma_y(h;\widehat{\theta}_y)$ for each grid knot of the fine grid $\mathcal{M}$. The UK predictions, $\{\widehat{y}(s) : s \in \mathcal{M}\} = \{\widehat{\mu}_{mmp}(s) : s \in \mathcal{M}\}$, now define the smoothed spatial trend of the originally sampled process $\mathcal{Z}$ from the MMPK approach.

Thirdly, calculate the trend residuals for further analysis. Predict the trend surface at the sampling locations, i.e. calculate $(\widehat{\mu}_{mmp}(s_1), ..., \widehat{\mu}_{mmp}(s_n))'$, and then take the differences $(z(s_1) - \widehat{\mu}_{mmp}(s_1), ..., z(s_n) - \widehat{\mu}_{mmp}(s_n))'$ to receive the residual vector $(\widehat{\delta}(s_1), ..., \widehat{\delta}(s_n))'$.

Fourthly, investigate the spatial dependence structure in the residuals. Fit an appropriate structure function (e.g. fit a semivariogram model $\gamma_\delta(h;\widehat{\theta}_\delta)$) and use this for OK of the residual process along with the fine grid $\mathcal{M}$ of prediction locations: $\{\widehat{\delta}_{ok}(s) : s \in \mathcal{M}\}$.

Fifthly, sum up the predictions of the trend and residual processes to receive a prediction of the original phenomenon $\widehat{Z}_{mmpk}(s) = \widehat{\mu}_{mmp}(s) + \widehat{\delta}_{ok}(s)$, where $s \in \mathcal{M}$. These predictions may then be used for mapping.

Finally, approximate the mean squared prediction error by

$$EMSPE(Z(s);\, \widehat{Z}_{mmpk}(s)) \;\; = \;\; EMSPE(\delta(s);\, \widehat{\delta}_{ok}(s)),$$

where $EMSPE(\delta(s);\, \widehat{\delta}_{ok}(s))$ is the estimated or empirical mean squared prediction error connected with OK of the residual process, similar to the original MPK approach.

4. APPLICATION: THE WOLFCAMP-AQUIFER DATA

When the Wolfcamp-Aquifer data were collected, there was an interest in a potential high-level nuclear-waste site in Texas. For this reason, the directions of groundwater flow were investigated. The data were reported and geostatistically analysed in Harper and Furr (1986) and Cressie (1989, 1993). The data show a clear downwards trend from south-west to north-east. Harper and Furr (1986) modelled this trend by use of a linear trend surface, whereas Cressie (1993) used the median polish approach. Another geostatistical approach is reported in Cressie (1989), where the power model for the structure function is used to filter out the trend using intrinsic random function kriging. In Figure 1, the spatial sampling network is presented. The dashed lines correspond to the low resolution grid to be used for median polishing in this work and the dotted lines indicate the border of the cells which belong to the grid knots. Thus, one can easily identify the rows and columns used along with median polishing. An open circle around one of the sampling locations at the western border of the investigation area indicates site 78 which is of interest in the following.

By use of the Wolfcamp-Aquifer data, the three spatial prediction methods UK, MPK and MMPK will be compared. At first, some specifics to the modelling approaches are given followed by visual display of the trend surfaces, the prediction surfaces, the error maps and normal probability plots for standardised cross-validation residuals.

A first order polynomial trend model is chosen for UK. For MPK, an equally spaced 7×6 low resolution grid is used. In MMPK, the same grid was used in the first step followed by fitting a first order trend polynomial to the median polish estimates. In all cases, isotropic spherical structure models are estimated by use of the robust semivariogram estimator (Cressie and Hawkins, 1980).

The Figures 2a, 2b and 2c show the resulting trend surfaces for UK, MPK and MMPK, respectively. With respect to the problem under consideration, i.e. groundwater level modelling and prediction, one might expect that the mean function of the spatial distribution somehow follows the elevation of the earth surface. Thus, the planar or angular surfaces in Figure 2a and 2b look artificial whilst Figure 2c shows a more realistic fit obtained by the modified median polish approach. Furthermore, unlike median polishing or low order polynomial trend surface modelling, the modified median polish approach in Figure 2c shows the property of smooth interpolation combined with high local flexibility.

The spatial prediction maps resulting from UK, MPK and MMPK are given in Figures 3a, 3b and 3c. The UK and MMPK prediction maps are extremely similar. The map resulting from OK lets the map reader identify the underlying median polish fit, i.e. the artificial angular behaviour of the median polish surface is still visible in the prediction map.

The error maps corresponding to the three spatial prediction methods are given by Figures 4a, 4b and 4c. Again, the results from UK and MMPK are visually equivalent. Although the errors from OK are in contrast much smaller, the MSPE from OK are known to be underestimated.

For comparison of the three fitted models, the following Figures 5a, 5b and 5c show normal probability plots of internally standardised cross-validation residuals resulting from each of the three spatial prediction methods UK, MPK and MMPK, respectively. The term "internally" standardised is used because the cross-validation residuals,

$$\widehat{\varepsilon}_{-i}^{int} = \frac{Z(s_i) - \widehat{Z}^{int}(s_{-i})}{\widehat{\sigma}(s_i)},$$

are standardised by the fitted model $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\theta}}$ based on the whole sample $\mathbf{Z}$ including $Z(s_i)$, whereas $\widehat{Z}^{int}(s_{-i})$ denotes the cross-validation predictor $p(Z(s_i)|\mathbf{Z_{-i}}, \widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\theta}})$ of $Z(s_i)$ based on the reduced sample information $\mathbf{Z}_{-i} = (..., Z(s_{i-1}), Z(s_{i+1}), ...)'$. Furthermore, $\widehat{\sigma}(s_i)$ denotes the root mean squared prediction error estimated from the whole sample information $\mathbf{Z}$ and based on $\widehat{\boldsymbol{\theta}}$. What is seen in Figure 5 is that all three spatial models give an excellent fit to the data. Furthermore, the data are Gaussian and are not contaminated by additive outliers. Therefore, it is preferable to use the BLUP method, i.e. UK. The robust methods MPK and MMPK are of no value in this situation.

To gain more insight, an error of 300 is added to the observation at site No. 78 to introduce an outlier:

$$\widetilde{Z}(s_{78}) = Z(s_{78}) + 300 = 3571 + 300 = 3871.$$

Sample site No. 78 is indicated in Figure 1 by an open-circle around the site locating bullet. The normal probabilty plots of internally standardised cross-validation residuals based on the outlier contaminated sample $\widetilde{\mathbf{Z}} = (Z(s_1), ..., \widetilde{Z}(s_{78}), ..., Z(s_n))'$ for all three methods UK, MPK and MMPK are shown in Figure 6. It follows that the single outlier has no effect on the predictions from MPK and MMPK but on the UK predictions.

Lastly, the model fitting procedures will be compared using normal probability plots of the externally standardised cross-validation residuals

$$\widehat{\varepsilon}_{-i}^{ext} = \frac{Z(s_i) - \widehat{Z}^{ext}(s_{-i})}{\widehat{\sigma}(s_i)}.$$

Here, the term "externally" standardised is used to indicate that the mean vector at the prediction site and the rest of the sample sites is now externally specified from $\widetilde{\mathbf{Z}}_{-i}$, i.e. the sample information without $Z(s_i)$: $\widehat{\boldsymbol{\mu}}_{-i}$; and $\widehat{Z}^{ext}(s_{-i})$ denotes the cross-validation predictor $p(Z(s_i)|\widetilde{\mathbf{Z}}_{-i}, \widehat{\boldsymbol{\mu}}_{-i}, \widehat{\boldsymbol{\theta}})$ of $Z(s_i)$.

The outlier can be easily identified from Figure 6a for the modelling approach used along with UK. The same holds for the robust methods of MPK and MMPK, but now the outlier is smaller. Nevertheless, an additional negative outlier could be detected from Figure 6b and 6c. This is due to the fact that the first column effect in median polishing is externally estimated based on just one of two observations, including the outlier sample $\widetilde{Z}(s_{78})$. Thus, the column effect is both overestimated and underestimated one time each. If the outlier would be introduced at an other site, i.e. where the sampling in the rows and columns is more dense, the residuals will be considerably downweighted by median polishing.

5. DISCUSSION

The aim of this work is to present a new method for geostatistical prediction called modified median polish prediction (MMPK) and to demonstrate its performance in comparison to median polish kriging (MPK). The comparison with universal kriging (UK) is presented mainly to show the results from the (perhaps) best or optimal prediction method for the Wolfcamp-Aquifer data, in the case of no outliers. And, unlike MPK, UK and MMPK result in visually equal prediction and error maps. However, when outliers occur in the given data set, MPK is recommended over UK but MMPK may be an even better choice for the purpose of spatial prediction.

So, the MPK and the MMPK methods are compared with respect to the results from UK. This is to distinguish from a comparison of all three prediction algorithms used along with certain trend models. Here, UK is based on the linear polynomial trend model, whereas MPK and MMPK start with a linear trend model given by

$$trend = all\ effect + row\ effects + column\ effects.$$

It is possible to use this trend model along with UK as well. For this define

$$\mu(s) = \beta_0 + \sum_{k=1}^{p} r_k(s)\beta_k + \sum_{l=1}^{q} c_l(s)\beta_l, \qquad s \in \mathcal{D}$$

where β_0 is the all effect, β_k, $k = 1, ..., p$, are the row effects and β_l, $l = 1, ..., q$, are the column effects. Furthermore, $r_k(s)$, $k = 1, ..., p$, and $c_l(s)$, $l = 1, ..., q$, denote indicator variables with $r_k(s) = 1$ if $s \in \mathcal{D}$ is a site in the k-th row and 0 otherwise, and $c_l(s)$ for the columns respectively. This gives surely a more flexible trend model for UK (see e.g. Cressie and Majure, 1997, for an example) and is similar to the mean polish approach followed by OK described in Cressie (1993, p. 187). Nevertheless, the polynomial trend model is more parsimonious and hence preferred, to show what may be the unknown truth in this application.

The proposed modification of MPK is simply the replacement of the spline interpolation step with an UK step. This is done to smooth the resulting trend surface and prediction map. Another possibility would be to use higher order splines instead of splines of degree 1 in MPK. This would considerably smooth the median polish surface. However, here, the statistical optimal smoothing method, i.e. UK, is preferred. This is due to the fact that UK is the appropriate technique to interpolate spatially correlated data. Nonetheless, there are similarities between spline interpolation and kriging, which is largely debated in the literature (Cressie, 1990b; Laslett, 1994; Kent and Mardia, 1994).

Another possibility to interpolate the median polish estimates is to use the traditional geostatistical approach of trend surface estimation. However, the use of UK is adopted here for the following reasons. Firstly, the aim is to interpolate the estimated mean values from median polishing and not to re-estimate the spatial mean function from previously obtained estimates. Secondly, trend surfaces are too smooth to allow for local flexibility in the mean surface. And lastly, UK may result in a direct interpolation of the median polish estimates if the corresponding structure function does not contain a nugget effect, i.e. $\widehat{\mu}_y(s_{kl}) = \widehat{\mu}_{mmp}(s)$, $s = s_{kl} \in \mathcal{G}$.

The main advantage of MMPK is the robust fit of the smooth and flexible trend surface. But, this is achieved at the expense of an additional variogram modelling step. Furthermore, in the present example, the error map from MMPK is close to the one from UK, whereas the error map from MPK is biased downwards. The reasons for this have not yet been investigated.

ACKNOWLEDGEMENTS

I thank Noel Cressie from the Ohio State University for the stimulating discussions and for encouraging me to work out the details presented so far. This work was partly supported by the Deutsche Forschungsgemeinschaft (DFG) through the Sonderforschungsbereich: "Reduction of complexity for multivariate data structures" (SFB 475) at the Department of Statistics, University of Dortmund.

REFERENCES

Berke, O. (1998). On spatiotemporal prediction for online monitoring data. *Communications in Statistics - Theory and Methods* **27**, 2343-2369.

Berke, O. (1999). Estimation and prediction in the spatial linear model. *Water, Air, and Soil Pollution* **110**, 215-237.

Box, G.E.P. and Jenkins, G.M. (1970). *Time Series Analysis: Forecasting and Control.* Holden-Day, San Francisco.

Brus, D.J., De Gruijter, J.J., Marsman, B.A. Visschers, R., Bregt, A., Breeuwsma, A. and Bouma, J. (1996). The performance of spatial interpolation methods and choropleth maps to estimate properties at points: a soil survey case study. *Environmetrics* **7**, 1-16.

Burrough, P.A., Bouma, J. and Yates, S.R. (1994). The state of art in Pedometrics. *Geoderma* **62**, 311-326.

Cleveland, W.S. and Devlin, S.J. (1988). Robust locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association* **83**, 596-610.

Cressie, N. (1986). Kriging nonstationary data. *Journal of the American Statistical Association* **81**, 625-634.

Cressie, N. (1989). Geostatistics. *The American Statistician* **43**, 197-202.

Cressie, N. (1990a). The origins of kriging. *Mathematical Geology* **22**, 239-252.

Cressie, N. (1990b). Reply to Whaba's letter. *The American Statistician* **44**, 256-258.

Cressie, N. (1993). *Statistics for Spatial Data*, revised edition. Wiley, New York.

Cressie, N. (2000). Geostatistical methods for mapping environmental exposures. In: Elliot, P., Wakefield, J., Best, N. and Briggs, D. (eds.) *Disease and Exposure Mapping.* Oxford University Press, Oxford.

Cressie, N. and Glonek, G. (1984). Median based covariogram estimators reduce bias. *Statistics and Probability Letters* **2**, 299-304.

Cressie, N., Gotway, C.A. and Grondona, M.O. (1990). Spatial prediction from networks. *Chemometrics and Intelligent Laboratory Systems* **7**, 251-271.

Cressie, N. and Hawkins, D.M. (1980). Robust estimation of the variogram: I. *Journal of the International Association for Mathematical Geology* **12**, 115-125.

Cressie, N. and Majure, J.J. (1997). Spatio-Temporal modeling of livestock waste in streams, *Journal of Agricultural, Biological, and Environmental Statistics* **2**, 24-47.

Ecker, M.D. and Heltshe, J.F. (1994). Geostatistical estimates of scallop abundance. In: Lange, N., Ryan, N., Billard, L., Brillinger, D., Conquest, L., and Greenhouse, J. (eds.) *Case Studies in Biometry.* Wiley, New York. 107-124.

Emerson, J.D. and Hoaglin, D.C. (1983). Analysis of two-way tables by medians. In: Hoaglin, D.C., Mosteller, F. and Tukey, J.W. (eds.) *Understanding Robust and Exploratory Data Analysis.* Wiley, New York, p. 166-210.

Harper, W.V. and Furr, J.M. (1986). Geostatistical analysis of potentiometric data in the Wolfcamp Aquifer of the Palo Duro Basin, Texas. *Technical Report* ONWI-587, Battle Memorial Institute, Columbus, OH.

Kent, J.T. and Mardia, K.V. (1994). The link between kriging and thin-plate splines. In: Kelly, F.P. (ed.) *Probability, Statistics and Optimisation.* Wiley, New York.

Kitanidis, P.K. (1997). *Introduction to Geostatistics: Applications to Hydrogeology.* Cambridge University Press, Cambridge.

Laslett, G.M. (1994). Kriging and splines: An empirical comparison of their predictive performance in some applications. *Journal of the American Statistical Association* **89**, 391-400.

Márkus, L., Berke, O., Kovács, J. and Urfer, W. (1999). Spatial prediction of the intensity of latent effects govering hydrogeological phenomena. *Environmetrics* **10**, 633-654.

Martin, R. J. (1992). Leverage, influence and residuals in regression models when observations are correlated. *Communications in Statistics - Theory and Methods* **21**, 1183-1212.

Matheron, G. (1969). Le Krigeage Universel. *Cahiers du Centre de Morphologie Mathematique*, No. 1, Fontainebleau, France.

Matheron, G. (1973). The intrinsic random functions and their applications. *Advances in Applied Probability* **5**, 439-468.

Opsomer, J.D., Ruppert, D., Wand, M.P., Holst, U. and Hössjer, O. (1999). Kriging with nonparametric variance function estimation. *Biometrics* **55**, 704-710.

Tukey, J.W. (1977). *Exploratory Data Analysis.* Addison-Wesley, Reading MA.

Wahba, G. (1990). *Spline Models for Observational Data.* SIAM, Philadelphia.

Webster, R., Oliver, M.A., Muir, K.R. and Mann, J.R. (1994). Kriging the local risk of a rare disease from a register of diagnoses. *Geographical Analysis* **26**, 168-185.

FIGURES

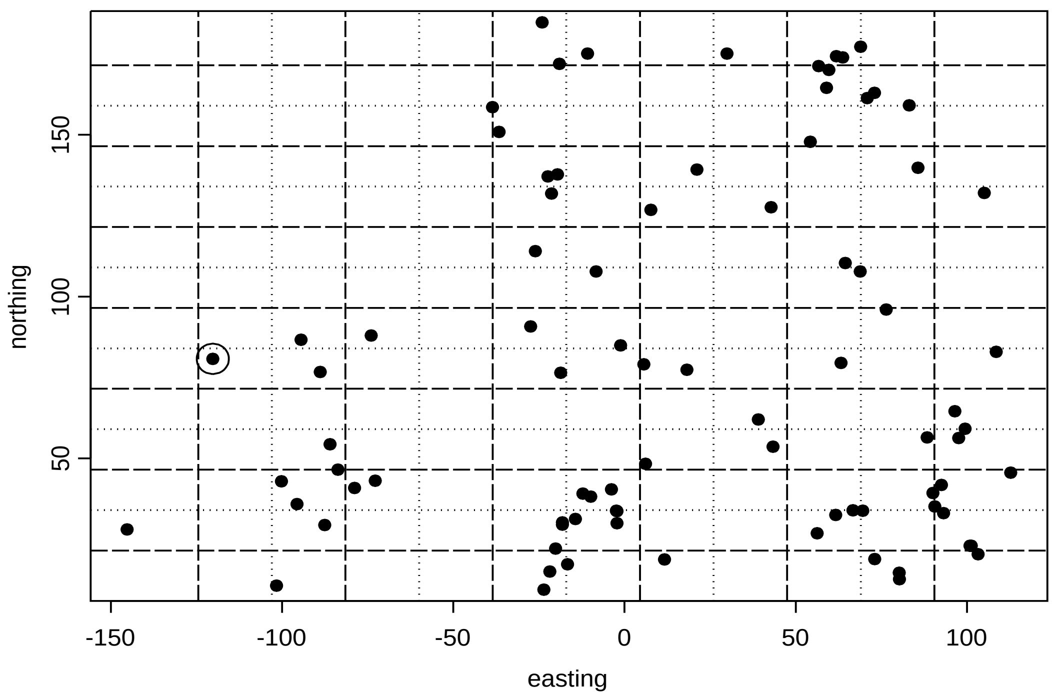


Figure 1. Map of the Wolfcamp-Aquifer sampling network.

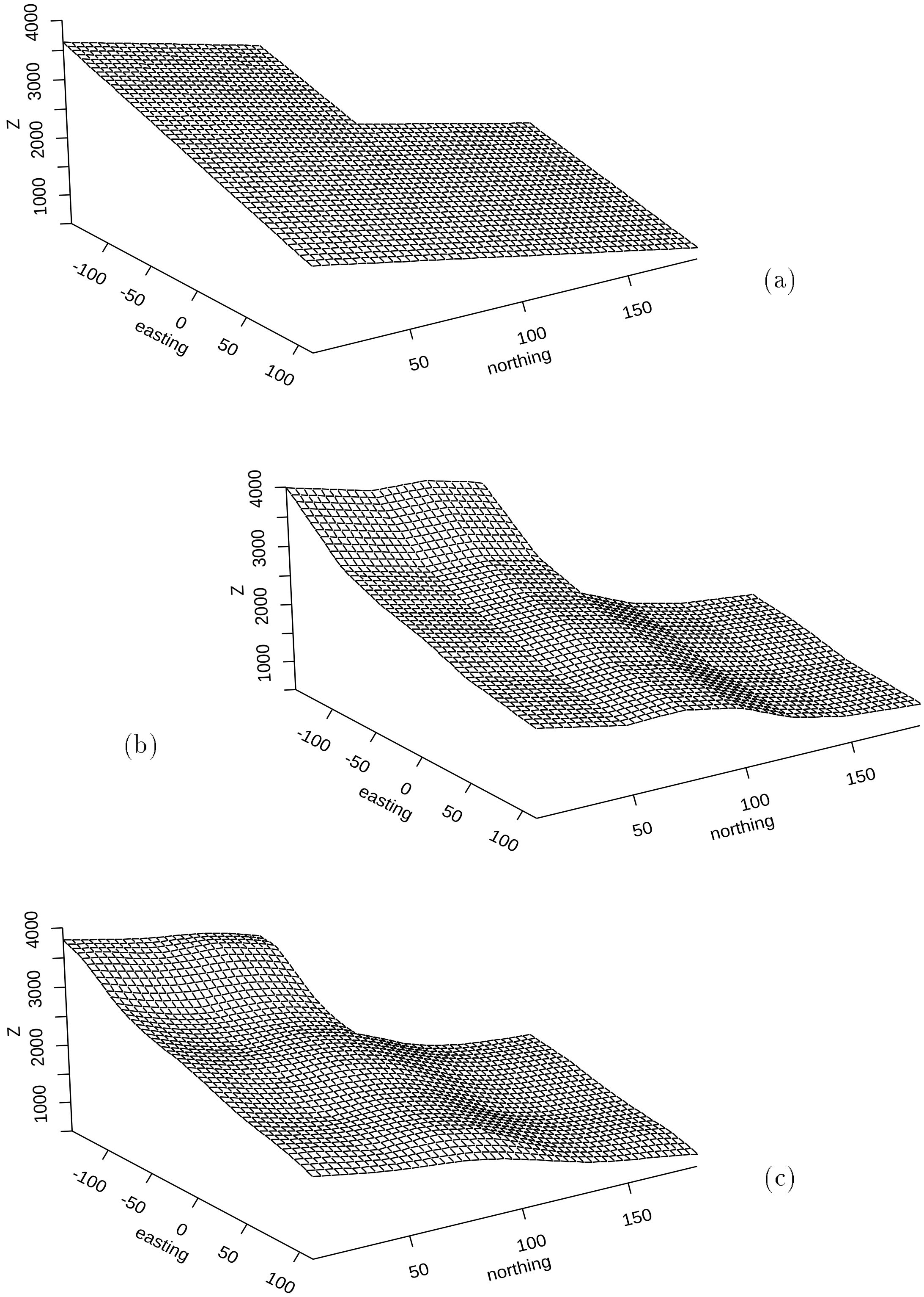


Figure 2. Trend surfaces from UK, MPK and MMPK.

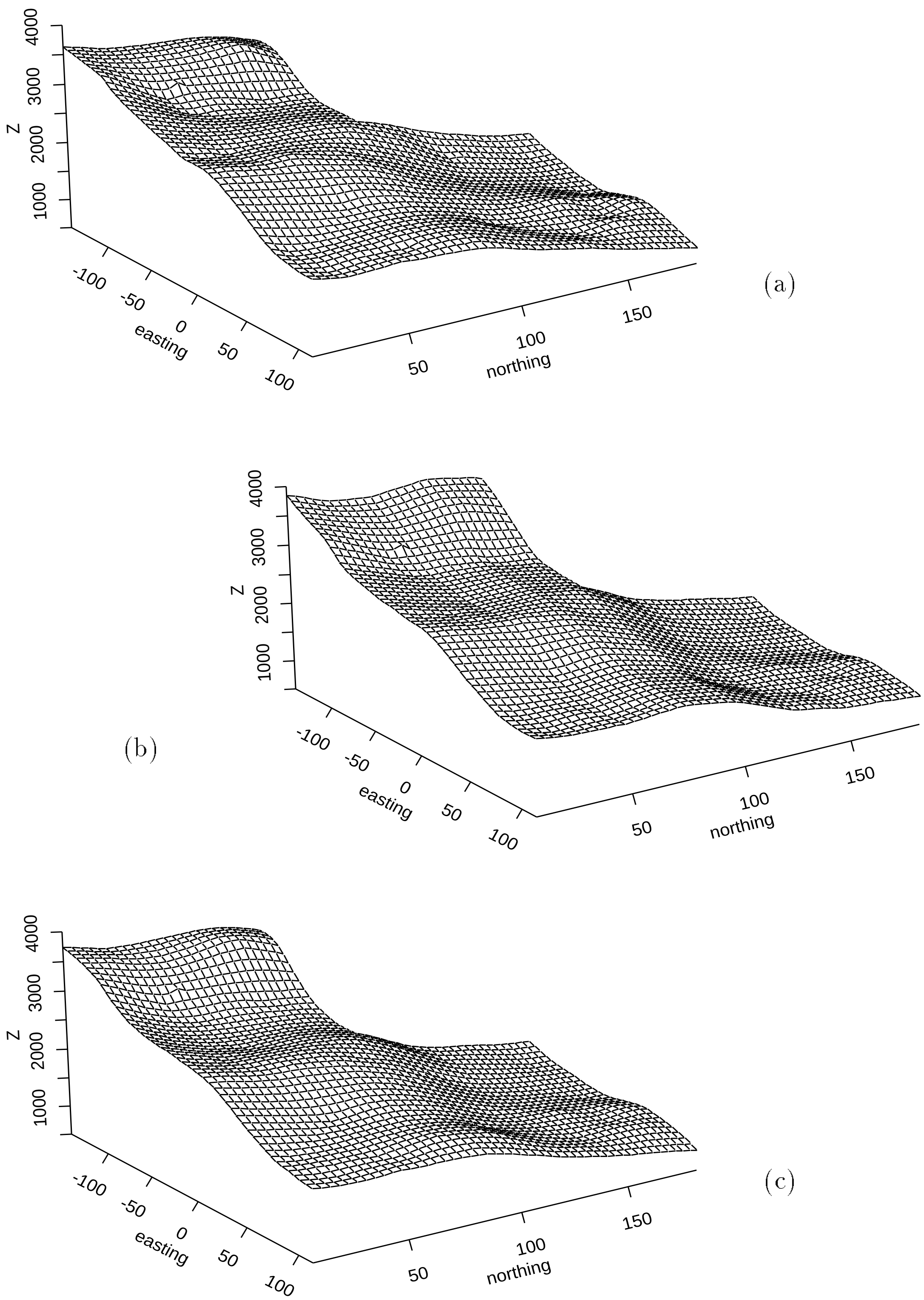
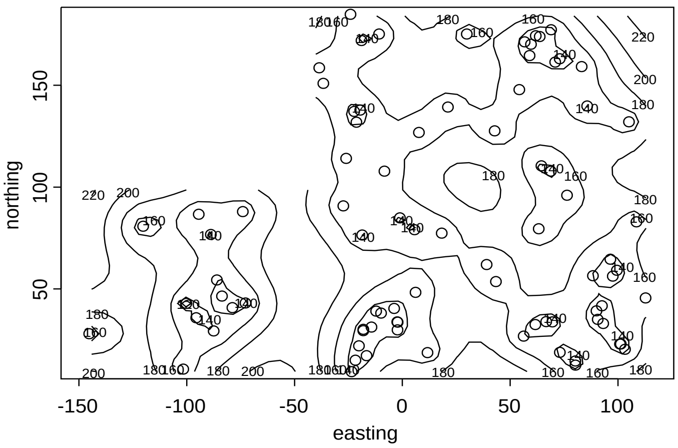
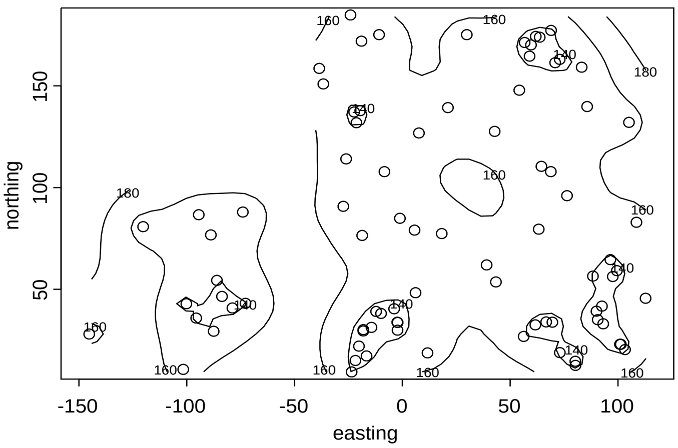


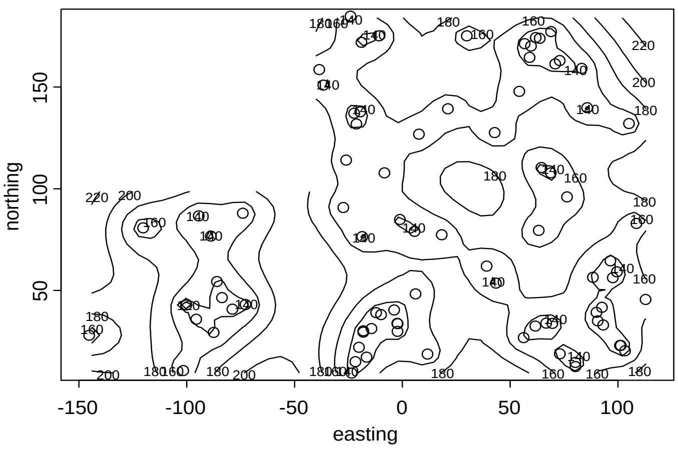
Figure 3. Prediction surfaces from UK, MPK and MMPK.



(a)

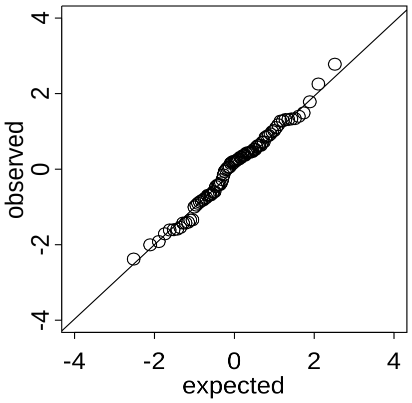


(b)

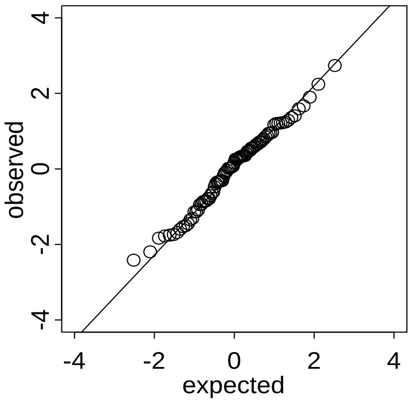


(c)

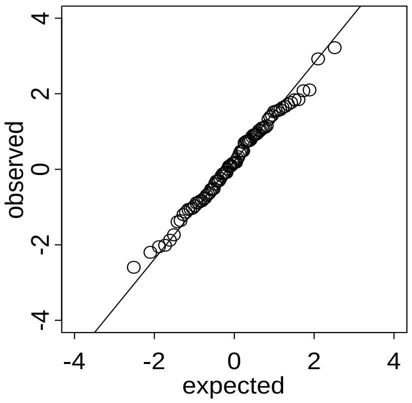
Figure 4. Error maps from UK, MPK and MMPK.



(a)

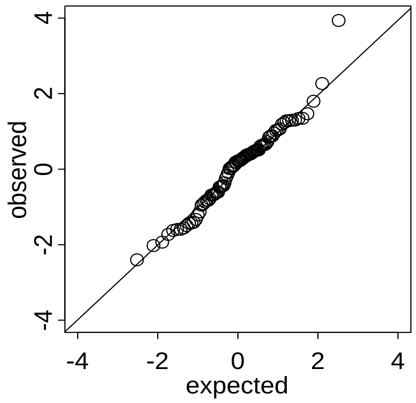


(b)

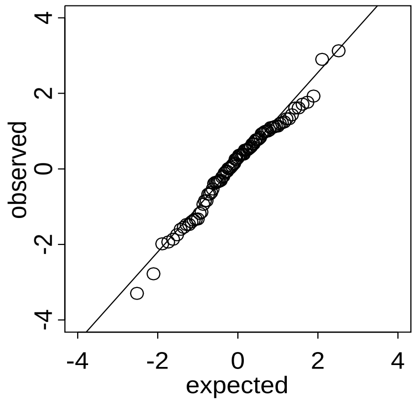


(c)

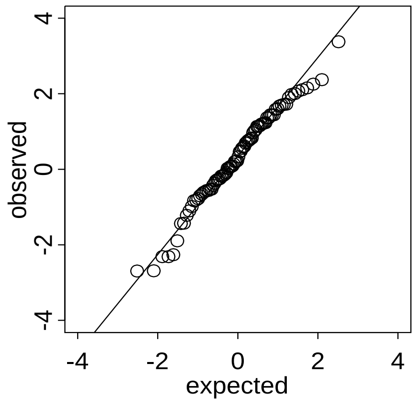
Figure 5. Normal probability plots for internally standardised cross-validation residuals from UK, MPK and MMPK.



(a)

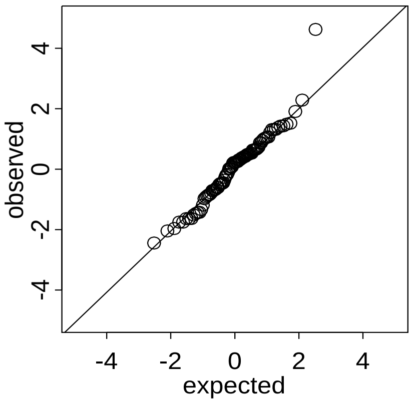


(b)

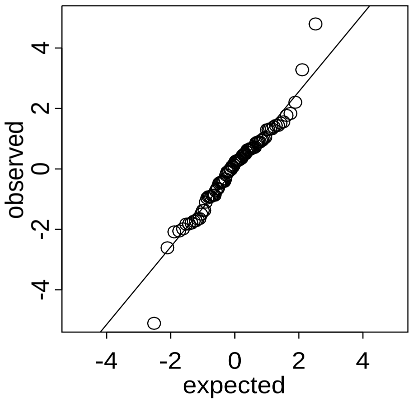


(c)

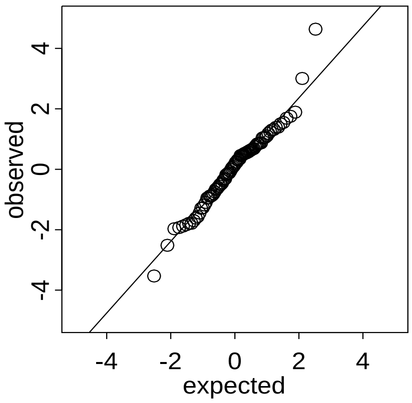
Figure 6. Normal probability plots for internally standardised cross-validation residuals from UK, MPK and MMPK based on the mean shift outlier data.



(a)



(b)



(c)

Figure 7. Normal probability plots for externally standardised cross-validation residuals from UK, MPK and MMPK based on the mean shift outlier data.