Robust estimation of the external drift and the variogram of spatial data

Künsch, Hans R.
ETH Zurich, Seminar for Statistics
Rämistrasse 101
8092 Zürich, Switzerland
E-mail: hans-ruedi.kuensch@stat.math.ethz.ch

Papritz, Andreas
ETH Zurich, Institute of Terrestrial Ecosystems
Universitätstrasse 16
8092 Zürich, Switzerland
E-mail: papritz@env.ethz.ch

Schwierz, Cornelia
ETH Zurich, Seminar for Statistics & Institute of Terrestrial Ecosystems
Rämistrasse 101
8092 Zürich, Switzerland
E-mail: cornelia.schwierz@stat.math.ethz.ch

Stahel, Werner A.
ETH Zurich, Seminar for Statistics
Rämistrasse 101
8092 Zürich, Switzerland
E-mail: stahel@stat.math.ethz.ch

Introduction

Most of the geostatistical software tools, available either commercially in GIS and in statistical software packages or as open source software (e.g. R, R Development Core Team, 2011) rely on non-robust algorithms. This is unfortunate, because outlying observations are rather the rule than the exception, in particular in environmental data sets. Outlying observations may results from errors (e.g. in data transcription) or from local perturbations in the processes that are responsible for a given pattern of spatial variation. As an example, the spatial distribution of some trace metal in the soils of a region may be distorted by emissions of local anthropogenic sources. Outliers affect the modelling of the large-scale spatial variation, the so-called external drift or trend, the estimation of the spatial dependence of the residual variation and the predictions by kriging. Identifying outliers manually is cumbersome and requires expertise. A better approach is to use robust algorithms that prevent automatically that outlying observations have undue influence.

Former studies on robust geostatistics focused on robust estimation of the sample variogram (cf. Lark, 2000, for a review), kriging without external drift (Hawkins and Cressie, 1984) and on robust estimation of parameters of a linear model for the external drift (Militino and Ugarte, 1997). Furthermore, Richardson and Welsh (1995) proposed a robustified version of (restricted) maximum likelihood ([RE]ML) estimation for the variance components of a linear mixed model, which was later used by Marchant and Lark (2007) for robust REML estimation of the variogram. We propose here a novel method for robust REML estimation of the variogram of a Gaussian random field that is possibly contaminated by independent errors from a long-tailed distribution. Besides robust estimates of the parameters of the external drift and of the variogram, the method also provides robustified kriging predictions.

Theory

We use the model $Y(\boldsymbol{s}) = \boldsymbol{x}(\boldsymbol{s})^{\mathrm{T}}\boldsymbol{\beta} + Z(\boldsymbol{s}) + \varepsilon(\boldsymbol{s})$ where $\boldsymbol{x}(\boldsymbol{s})^{\mathrm{T}}\boldsymbol{\beta}$ is the external drift, $Z(\boldsymbol{s})$ is a stationary Gaussian field with mean zero and covariance $R(\boldsymbol{h}; \sigma_0^2, \alpha)$, and $\varepsilon(\boldsymbol{s})$ is an independently distributed error with a probability density function proportional to $1/\sigma \exp(-\rho(\varepsilon/\sigma))$ so that $\rho(x) = x^2/2$ gives the case where $\varepsilon(\boldsymbol{s})$ has a Gaussian distribution. σ^2 is the squared scale parameter of ε and is commonly called the nugget, and $\boldsymbol{\theta}^{\mathrm{T}} = (\sigma_0^2,\, \alpha)$ are the sill and range parameters of the covariance function (or variogram). For some variogram models $\boldsymbol{\theta}$ contains further parameters. Using the observations, $\boldsymbol{y}^{\mathrm{T}} = (y(\boldsymbol{s}_1), y(\boldsymbol{s}_2), \ldots, y(\boldsymbol{s}_n))$ we want to estimate $\boldsymbol{\beta}$, σ^2 and $\boldsymbol{\theta}$ and to predict $Z(\boldsymbol{s})$ for some $\boldsymbol{s}$ where $\boldsymbol{s}$ can be equal to one of the sampled locations $\boldsymbol{s}_i$ or different.

The starting point for developing our new procedure is the Gaussian log-likelihood

$$(1) \qquad l(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}|\boldsymbol{y}) = -\frac{1}{2}\log\big(\det(\sigma^2\boldsymbol{I} + \boldsymbol{V_\theta})\big) - \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X\beta})^{\mathrm{T}}(\sigma^2\boldsymbol{I} + \boldsymbol{V_\theta})^{-1}(\boldsymbol{y} - \boldsymbol{X\beta})$$

that arises when $\varepsilon(\boldsymbol{s})$ is Gaussian. $\boldsymbol{V_\theta} = \sigma_0^2\boldsymbol{V}_\alpha$ is the covariance and $\boldsymbol{V}_\alpha$ the correlation matrix of $\boldsymbol{z}^{\mathrm{T}} = \big(z(\boldsymbol{s}_1), z(\boldsymbol{s}_2), \ldots, z(\boldsymbol{s}_n)\big)$. We define the pseudo log-likelihood that depends in addition to the parameters on the latent variable $\boldsymbol{z}$

$$(2) \qquad l^*(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}, \boldsymbol{z}|\boldsymbol{y}) = -\frac{1}{2}\log\big(\det(\sigma^2\boldsymbol{I} + \boldsymbol{V_\theta})\big) - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{y(\boldsymbol{s}_i) - \boldsymbol{x}(\boldsymbol{s}_i)^{\mathrm{T}}\boldsymbol{\beta} - z(\boldsymbol{s}_i)}{\sigma}\right)^2 - \frac{1}{2}\boldsymbol{z}^{\mathrm{T}}\boldsymbol{V_\theta}^{-1}\boldsymbol{z}.$$

The Gaussian log-likelihood may be considered as a profile log-likelihood of $l^*(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}, \boldsymbol{z}|\boldsymbol{y})$ that has been maximized with respect to $\boldsymbol{z}$. Hence, $l(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}|\boldsymbol{y}) = l^*(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}, \breve{\boldsymbol{z}}|\boldsymbol{y})$ where

$$\breve{\boldsymbol{z}} = \underset{z}{\operatorname{argmax}}\big(l^*(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}, \boldsymbol{z}|\boldsymbol{y})\big) = \boldsymbol{V_\theta}(\sigma^2\boldsymbol{I} + \boldsymbol{V_\theta})^{-1}(\boldsymbol{y} - \boldsymbol{X\beta}).$$

Maximizing (1) with respect to $\boldsymbol{\beta}$, σ^2 and $\boldsymbol{\theta}$ is thus equivalent to solving the expanded system of ML estimating equations

$$(3) \qquad \frac{\partial l^*}{\partial \boldsymbol{z}} = \frac{1}{\hat\sigma}\frac{\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{z}}}{\hat\sigma} - \boldsymbol{V}_{\hat{\boldsymbol{\theta}}}^{-1}\hat{\boldsymbol{z}} = \boldsymbol{0},$$

$$(4) \qquad \frac{\partial l^*}{\partial \boldsymbol{\beta}} = \boldsymbol{X}^{\mathrm{T}}\,\frac{\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{z}}}{\hat\sigma} = \boldsymbol{0},$$

$$(5) \qquad \frac{\partial l^*}{\partial \sigma^2} = \operatorname{trace}\big((\boldsymbol{I} + \frac{1}{\hat\sigma^2}\boldsymbol{V}_{\hat{\boldsymbol{\theta}}})^{-1}\big) - \sum_{i=1}^{n}\left(\frac{y(\boldsymbol{s}_i) - \boldsymbol{x}(\boldsymbol{s}_i)^{\mathrm{T}}\hat{\boldsymbol{\beta}} - \hat z(\boldsymbol{s}_i)}{\hat\sigma}\right)^2 = 0,$$

$$(6) \qquad \frac{\partial l^*}{\partial \sigma_0^2} = n - \operatorname{trace}\big((\boldsymbol{I} + \frac{1}{\hat\sigma^2}\boldsymbol{V}_{\hat{\boldsymbol{\theta}}})^{-1}\big) - \hat{\boldsymbol{z}}^{\mathrm{T}}\boldsymbol{V}_{\hat{\boldsymbol{\theta}}}^{-1}\hat{\boldsymbol{z}} = 0,$$

$$(7) \qquad \frac{\partial l^*}{\partial \alpha} = \operatorname{trace}\big((\frac{\hat\sigma^2}{\hat\sigma_0^2}\boldsymbol{I} + \boldsymbol{V}_{\hat\alpha})^{-1}\frac{\partial \boldsymbol{V}_\alpha}{\partial \alpha}\Big|_{\alpha=\hat\alpha}\big) - \frac{1}{\hat\sigma_0^2}\hat{\boldsymbol{z}}^{\mathrm{T}}\boldsymbol{V}_{\hat\alpha}^{-1}\frac{\partial \boldsymbol{V}_\alpha}{\partial \alpha}\Big|_{\alpha=\hat\alpha}\boldsymbol{V}_{\hat\alpha}^{-1}\hat{\boldsymbol{z}} = 0.$$

Closed-form expressions exist in general only for the generalized least squares (GLS) estimates of $\boldsymbol{\beta}$

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^{\mathrm{T}}(\hat\sigma^2\boldsymbol{I} + \boldsymbol{V}_{\hat{\boldsymbol{\theta}}})^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}(\hat\sigma^2\boldsymbol{I} + \boldsymbol{V}_{\hat{\boldsymbol{\theta}}})^{-1}\boldsymbol{y},$$

and the plug-in mean square predictions of $\boldsymbol{z}$

$$\hat{\boldsymbol{z}} = \boldsymbol{V}_{\hat{\boldsymbol{\theta}}}(\hat\sigma^2\boldsymbol{I} + \boldsymbol{V}_{\hat{\boldsymbol{\theta}}})^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}),$$

which are called universal (UK) or external drift kriging predictions in geostatistics.

Equations (3)–(7) show that the Gaussian ML estimates and the UK predictions depend on the standardized residuals $\hat{\boldsymbol{r}}/\hat{\sigma} = (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{z}})/\hat{\sigma}$, and it is obvious that both estimates and predictions are sensitive to outliers in $\boldsymbol{y}$. Welsh and Richardson (1997) proposed several approaches to make ML estimates in linear mixed models robust to outlying observations. A first approach consists of robustifying the Gaussian log-likelihood. One has then to replace $(y(\boldsymbol{s}_i) - \boldsymbol{x}(\boldsymbol{s}_i)^{\mathrm{T}}\boldsymbol{\beta} - z(\boldsymbol{s}_i))^2/\sigma^2$ in (2) by a function $\rho_c(\cdot)$ of the standardized errors that grows more slowly than the quadratic (c denotes here a tuning parameter that controls the amount of robustness of the estimates). Initially, we tried such a strategy and approximated the restricted likelihood function (Harville, 1974) of a model with long-tailed $\varepsilon(\boldsymbol{s})$ by a Laplace approximation and maximized the approximation to get robust REML estimates of the drift and variogram parameters (Schwierz *et al.*, 2010). However, to bound the influence of outliers, we had to use a bounded $\rho_c(\cdot)$ function, which made the numerical maximization of the restricted log-likelihood more difficult because the related system of estimating equations has then multiple roots. Furthermore, we had difficulties to find appropriate bias corrections to obtain Fisher consistent estimates for Gaussian $\varepsilon(\boldsymbol{s})$.

Therefore, we eventually pursued another suggestion by Welsh and Richardson (1997) and robustified (3)–(7) by replacing $\hat{\boldsymbol{r}}/\hat{\sigma}$ by a bounded and odd function $\psi_c(\cdot)$ of the standardized residuals, yielding thereby a system of robustified estimating equations

$$\frac{1}{\tilde{\sigma}}\,\psi_c\Big(\frac{\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}} - \tilde{\boldsymbol{z}}}{\tilde{\sigma}}\Big) - \boldsymbol{V}_{\tilde{\boldsymbol{\theta}}}^{-1}\tilde{\boldsymbol{z}} + a_1 = \boldsymbol{0}, \tag{8}$$

$$\boldsymbol{X}^{\mathrm{T}}\,\psi_c\Big(\frac{\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}} - \tilde{\boldsymbol{z}}}{\tilde{\sigma}}\Big) + a_2 = \boldsymbol{0}, \tag{9}$$

$$\mathrm{trace}\Big(\big(\boldsymbol{I} + \frac{1}{\tilde{\sigma}^2}\boldsymbol{V}_{\tilde{\boldsymbol{\theta}}}\big)^{-1}\Big) - \sum_{i=1}^{n}\psi_c^2\Big(\frac{y(\boldsymbol{s}_i) - \boldsymbol{x}(\boldsymbol{s}_i)^{\mathrm{T}}\tilde{\boldsymbol{\beta}} - \tilde{z}(\boldsymbol{s}_i)}{\tilde{\sigma}}\Big) + a_3 = 0, \tag{10}$$

$$n - \mathrm{trace}\Big(\big(\boldsymbol{I} + \frac{1}{\tilde{\sigma}^2}\boldsymbol{V}_{\tilde{\boldsymbol{\theta}}}\big)^{-1}\Big) - \tilde{\boldsymbol{z}}^{\mathrm{T}}\boldsymbol{V}_{\tilde{\boldsymbol{\theta}}}^{-1}\tilde{\boldsymbol{z}} + a_4 = 0, \tag{11}$$

$$\mathrm{trace}\Big(\big(\frac{\tilde{\sigma}^2}{\tilde{\sigma}_0^2}\boldsymbol{I} + \boldsymbol{V}_{\tilde{\alpha}}\big)^{-1}\frac{\partial \boldsymbol{V}_\alpha}{\partial \alpha}\Big|_{\alpha=\tilde{\alpha}}\Big) - \frac{1}{\tilde{\sigma}_0^2}\tilde{\boldsymbol{z}}^{\mathrm{T}}\boldsymbol{V}_{\tilde{\alpha}}^{-1}\frac{\partial \boldsymbol{V}_\alpha}{\partial \alpha}\Big|_{\alpha=\tilde{\alpha}}\boldsymbol{V}_{\tilde{\alpha}}^{-1}\tilde{\boldsymbol{z}} + a_5 = 0. \tag{12}$$

In the above equations, a_1 to a_5 are adjustments for Fisher consistency at the Gaussian model. They are determined by the condition that the expectations of (8)–(12) must vanish for Gaussian $\varepsilon(\boldsymbol{s})$ (e.g. Maronna *et al.*, 2006, p. 67). Hence, $a_1 = a_2 = 0$ since $\psi_c(\cdot)$ is odd, and

$$a_3 = a_3^* - \mathrm{trace}\Big(\big(\boldsymbol{I} + \frac{1}{\tilde{\sigma}^2}\boldsymbol{V}_{\tilde{\boldsymbol{\theta}}}\big)^{-1}\Big) \tag{13}$$

$$a_4 = a_4^* - \Big(n - \mathrm{trace}\Big(\big(\boldsymbol{I} + \frac{1}{\tilde{\sigma}^2}\boldsymbol{V}_{\tilde{\boldsymbol{\theta}}}\big)^{-1}\Big)\Big), \tag{14}$$

$$a_5 = a_5^* - \mathrm{trace}\Big(\big(\frac{\tilde{\sigma}^2}{\tilde{\sigma}_0^2}\boldsymbol{I} + \boldsymbol{V}_{\tilde{\alpha}}\big)^{-1}\frac{\partial \boldsymbol{V}_\alpha}{\partial \alpha}\Big|_{\alpha=\tilde{\alpha}}\Big), \tag{15}$$

where (using (8) to derive (16))

$$a_3^* = \sum_{i=1}^{n}\mathrm{E}\Big[\psi_c^2\Big(\frac{y(\boldsymbol{s}_i) - \boldsymbol{x}(\boldsymbol{s}_i)^{\mathrm{T}}\tilde{\boldsymbol{\beta}} - \tilde{z}(\boldsymbol{s}_i)}{\tilde{\sigma}}\Big)\Big] = \tilde{\sigma}^2\,\mathrm{trace}\big(\boldsymbol{V}_{\tilde{\boldsymbol{\theta}}}^{-2}\mathrm{E}\,[\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{z}}^{\mathrm{T}}]\big), \tag{16}$$

$$a_4^* = \mathrm{E}\,[\tilde{\boldsymbol{z}}^{\mathrm{T}}\boldsymbol{V}_{\tilde{\boldsymbol{\theta}}}^{-1}\tilde{\boldsymbol{z}}] = \mathrm{trace}\big(\boldsymbol{V}_{\tilde{\boldsymbol{\theta}}}^{-1}\mathrm{E}\,[\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{z}}^{\mathrm{T}}]\big), \tag{17}$$

$$a_5^* = \mathrm{E}\Big[\frac{1}{\tilde{\sigma}_0^2}\tilde{\boldsymbol{z}}^{\mathrm{T}}\boldsymbol{V}_{\tilde{\alpha}}^{-1}\frac{\partial \boldsymbol{V}_\alpha}{\partial \alpha}\Big|_{\alpha=\tilde{\alpha}}\boldsymbol{V}_{\tilde{\alpha}}^{-1}\tilde{\boldsymbol{z}}\Big] = \frac{1}{\tilde{\sigma}_0^2}\mathrm{trace}\Big(\boldsymbol{V}_{\tilde{\alpha}}^{-1}\frac{\partial \boldsymbol{V}_\alpha}{\partial \alpha}\Big|_{\alpha=\tilde{\alpha}}\boldsymbol{V}_{\tilde{\alpha}}^{-1}\mathrm{E}\,[\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{z}}^{\mathrm{T}}]\Big). \tag{18}$$

To correct the biases we thus need the covariance matrix of $\tilde{\boldsymbol{z}}$, which we approximated in the following way: The starting point was a first-order Taylor series approximation for $\psi_c(\tilde{\boldsymbol{r}}/\sigma)$ at $\boldsymbol{\varepsilon}/\sigma$

$$\psi_c(\tilde{\boldsymbol{r}}/\sigma) \approx \psi_c(\boldsymbol{\varepsilon}/\sigma) - \frac{1}{\sigma}\mathrm{diag}\big(\psi_c'(\boldsymbol{\varepsilon}/\sigma)\big)\big(\tilde{\boldsymbol{z}} - \boldsymbol{z} + \boldsymbol{X}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})\big)$$

where $\tilde{\boldsymbol{r}}^{\mathrm{T}} = \boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}} - \tilde{\boldsymbol{z}}$ and $\boldsymbol{\varepsilon}^{\mathrm{T}} = (\varepsilon(\boldsymbol{s}_1), \varepsilon(\boldsymbol{s}_2), \ldots, \varepsilon(\boldsymbol{s}_n))$. Then we replaced $\psi_c'(\varepsilon(\boldsymbol{s}_i)/\sigma)$ by its expectation, say $b = \mathrm{E}\left[\psi_c'(\varepsilon(\boldsymbol{s}_i)/\sigma)\right]$ (for Gaussian $\varepsilon(\boldsymbol{s})$), and substituted the modified Taylor approximation in (8)–(9) for $\psi_c(\tilde{\boldsymbol{r}}/\sigma$, which resulted in

$$\begin{bmatrix} \tilde{\boldsymbol{z}} \\ \tilde{\boldsymbol{\beta}} - \boldsymbol{\beta} \end{bmatrix} \approx \boldsymbol{M}^{-1} \begin{bmatrix} \boldsymbol{I} & \boldsymbol{I} \\ \boldsymbol{X}^{\mathrm{T}} & \boldsymbol{X}^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} b\,\boldsymbol{z} \\ \sigma\psi(\boldsymbol{\varepsilon}/\sigma) \end{bmatrix}$$

where

$$\boldsymbol{M} = \begin{bmatrix} b\,\boldsymbol{I} + \sigma^2 \boldsymbol{V}_{\boldsymbol{\theta}}^{-1} & b\,\boldsymbol{X} \\ b\,\boldsymbol{X}^{\mathrm{T}} & b\,\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} \end{bmatrix}.$$

Since $\boldsymbol{z}$ and $\psi_c(\boldsymbol{\varepsilon}/\sigma)$ are independent, and letting $a = \mathrm{E}\left[\psi_c^2(\varepsilon(\boldsymbol{s})/\sigma)\right]$ (for Gaussian $\varepsilon(\boldsymbol{s})$), we finally obtained the approximation

$$\begin{bmatrix} \mathrm{Cov}\,[\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{z}}^{\mathrm{T}}] & \mathrm{Cov}\,[\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{\beta}}^{\mathrm{T}}] \\ \mathrm{Cov}\,[\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{z}}^{\mathrm{T}}] & \mathrm{Cov}\,[\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\beta}}^{\mathrm{T}}] \end{bmatrix} \approx \boldsymbol{M}^{-1} \begin{bmatrix} \boldsymbol{\Lambda} & \boldsymbol{\Lambda X} \\ \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\Lambda} & \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\Lambda X} \end{bmatrix} \boldsymbol{M}^{-1}, \tag{19}$$

where $\boldsymbol{\Lambda} = b^2 \boldsymbol{V}_{\boldsymbol{\theta}} + a\,\sigma^2 \boldsymbol{I}$.

Solving (8)–(12) subject to (13)–(15) and (19) for the non-robust case yields the customary REML estimates of σ^2 and $\boldsymbol{\theta}$ and the related plug-in GLS estimate of $\boldsymbol{\beta}$ and the UK prediction of $\boldsymbol{z}$. Hence, apart from the Bayesian interpretation of the REML estimates (Harville, 1974) and the interpretation as ML estimates obtained from error contrasts (Patterson and Thompson, 1971), the REML estimates can also be considered as ML estimates, forced to be Fisher consistent for finite sample size.

Estimation procedure

We wrote an R function that solves the robustified estimating equations for $\tilde{\boldsymbol{z}}$, $\tilde{\boldsymbol{\beta}}$, $\tilde{\sigma}^2$ and $\tilde{\boldsymbol{\theta}}$. Instead of the popular Huber ψ_c-function (Maronna *et al.*, 2006, p. 26), we used the continuously differentiable function

$$\psi_c(x) = \frac{2c}{1 + \exp(-2x/c)} - c, \tag{20}$$

which corresponds to a shifted and scaled logistic cumulative distribution function. Note that $\lim_{x\to\pm\infty} \psi_c(x) = \pm c$. We computed the roots of the estimating equations by a combination of iteratively re-weighted least squares (IRWLS, e.g. Maronna *et al.*, 2006, p. 105) and a Broyden's scheme. For given $\tilde{\sigma}^2$ and $\tilde{\boldsymbol{\theta}}$, $\tilde{\boldsymbol{z}}$ and $\tilde{\boldsymbol{\beta}}$ were determined by IRWLS. The estimates were then inserted into (10)–(12), and the roots of these equations were computed by Broyden's method as implemented in the R package nleqslv.

The estimation procedure needs initial values for $\boldsymbol{z}$ and all the parameters. An initial guess of $\tilde{\boldsymbol{\beta}}$ was obtained by MM-estimation (Maronna *et al.*, 2006, pp. 124). The robust regression residuals were then spatially smoothed by loess to get an initial $\tilde{\boldsymbol{z}}$. The MAD of the loess residuals was taken as initial $\tilde{\sigma}$. An initial guess of $\tilde{\boldsymbol{\theta}}$ was obtained by fitting the sill and range parameters of a parametric variogram model (with nugget fixed at the initial $\tilde{\sigma}^2$) by weighted non-linear least squares (Cressie, 1993, p. 96) to the sample variogram of the regression residuals that had been computed by the MAD estimator (Dowd, 1984).

Table 1: Statistics of relative errors (%) of the approximations of a_3^*, a_4^* and a_5^* in simulations of $n =$ 50–500 $Y(\boldsymbol{s})$ ($\boldsymbol{s}$ uniformly distributed on the unit square) for tuning parameter $c = 1$. The data were generated with a linear external drift in the coordinates and an exponential variogram with ranges varying from 0.01 to 0.2 and nugget:sill ratios from 0.25 to 4.

	minimum	1st quartile	median	3rd quartile	maximum
a_3^*	−11.5	−7.6	−4.5	−2.7	−0.9
a_4^*	−10.7	−6.6	−3.8	−2.3	−0.9
a_5^*	−17.7	−9.1	−4.8	−2.5	5.2

Simulation study

We explored the properties of our robust REML estimation method by simulations. First, we checked approximation (19) for $\mathrm{Cov}\,[\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{z}}^{\mathrm{T}}]$. We simulated for several external drifts and exponential or cubic variograms (with various combinations of nugget, sill and range) data sets consisting of 50–500 Gaussian $Y(\boldsymbol{s})$ on grids or for uniformly distributed $\boldsymbol{s}$ in one- and two-dimensional domains.

For each scenario, we simulated 5 000 realizations and computed for each realization a_3^*, a_4^* and a_5^* using (19) and the true σ^2 and $\boldsymbol{\theta}$. Then we compared the mean of these 5 000 a_i^* values with the "true" value that was computed with the empirical covariances of the 5 000 estimates $\tilde{\boldsymbol{z}}$, obtained by

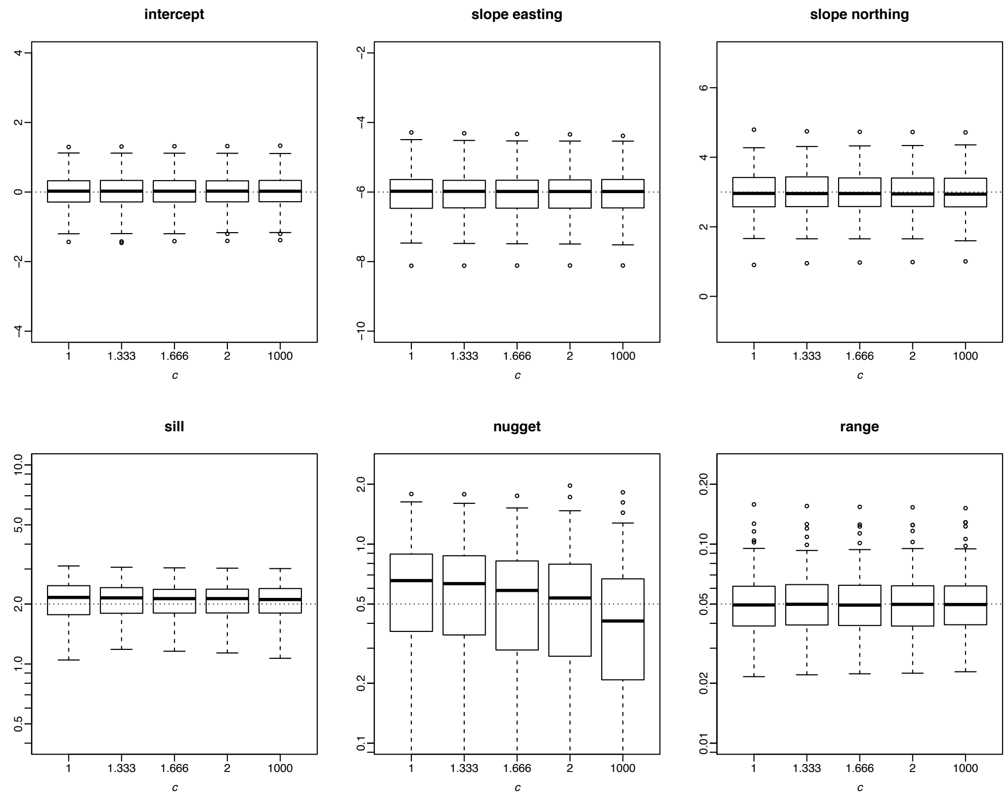

Figure 1: Boxplots of estimates $\tilde{\boldsymbol{\beta}}$, $\tilde{\sigma}^2$ and $\tilde{\boldsymbol{\theta}}$ computed from 200 Gaussian data sets that were simulated with a linear external drift in the coordinates and an exponential variogram with $\sigma^2 = 0.5$, $\sigma_0^2 = 2$, and $\alpha = 0.05$ for $n = 200$ uniformly distributed $\boldsymbol{s}$ in the unit square (dotted lines mark true parameter values, for $c = 1000$ estimates correspond to customary REML estimates).

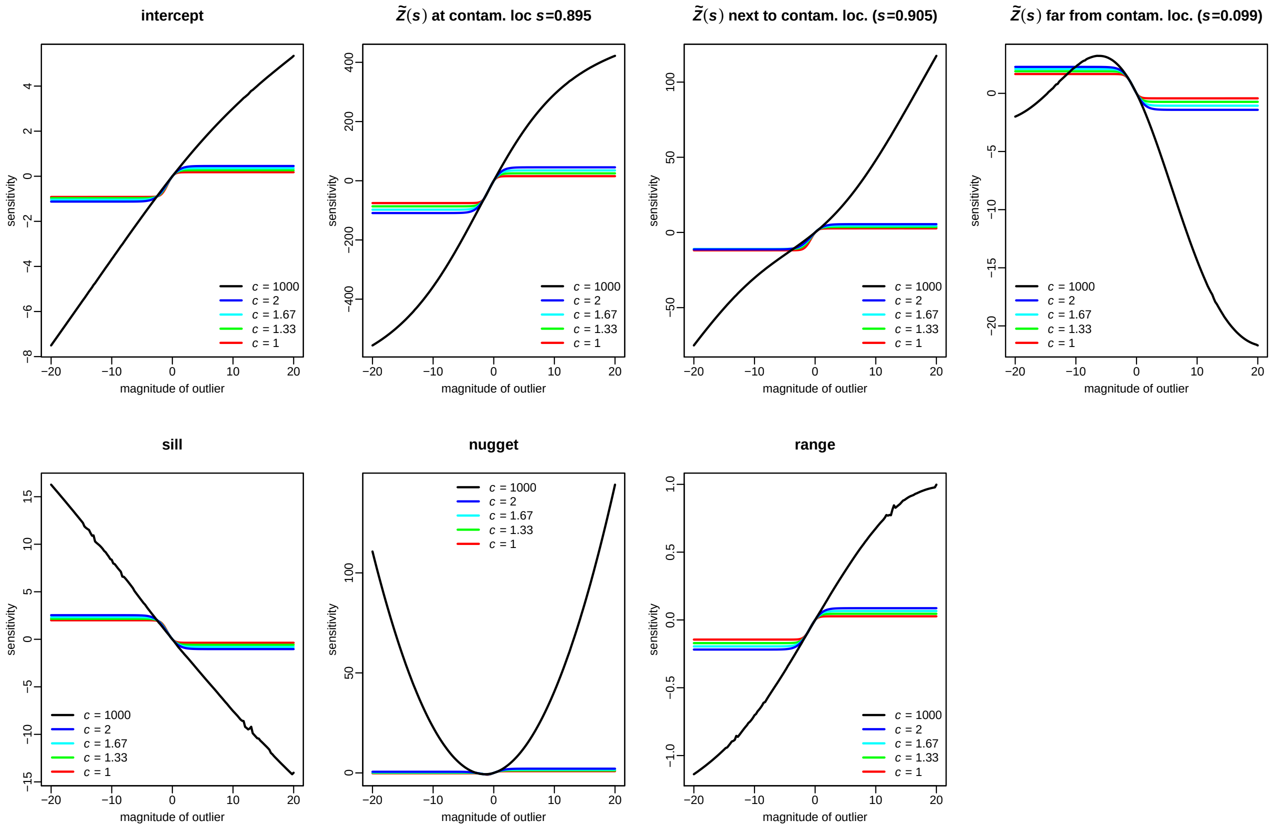

Figure 2: Standardized sensitivity curves for one realization of the simulation scenario 500 Gaussian $Y(\boldsymbol{s})$ on centred regular grid in $[0, 1]$; no trend ($\beta = -0.5$); exponential variogram with $\sigma^2 = 0.5$, $\sigma_0^2 = 2$, $\alpha = 0.05$. The observation at $\boldsymbol{s} = 0.895$ was contaminated by adding $x \in (-20, 19.75, -19.5, \ldots, 19.75, 20)$. Then we estimated for each x the model parameters and $\boldsymbol{z}$, and plotted $n(\tilde{\phi}(x) - \tilde{\phi}(0))$ vs x, where ϕ denotes a parameter or an element of $\boldsymbol{z}$. Estimates and predictions correspond for $c = 1000$ to customary REML estimates and UK predictions.

solving (8)–(9) with the true covariance parameters. Table 1 lists as example the relative errors of the approximations for the simulations with the spatially uniformly distributed $\boldsymbol{s}$ on the unit square. On average, the approximation underestimated the a_i^*, but, except for the very short ranges, the modulus of the relative errors rarely exceeded 10 %. Errors in the approximations of a_3^* to a_5^* lead to some bias in the parameter estimates. Figure 1 shows boxplots of the estimates $\tilde{\boldsymbol{\beta}}$, $\tilde{\sigma}^2$ and $\tilde{\boldsymbol{\theta}}$ for a simulation scenario where the approximation resulted for $c = 1$ in relatively large errors (a_3^*: -10.4 %, a_4^*: -8.3 %, a_5^* -13.0 %). The medians of the variance estimates deviate with decreasing c increasingly from the true values, but compared to the biases that we faced when maximizing the robustified log-likelihood these systematic errors seem tolerable.

Second, we computed for a single realization of one simulation scenario sensitivity curves (Maronna *et al.*, 2006, p. 55). Figure 2 shows that the sensitivity curves remained bounded for $c \leq 2$, both for the parameter estimates and the predictions. As expected from theory, the sensitivity decreased with decreasing c. In contrast, the non-robust estimates ($c = 1000$) got increasingly corrupted by increasing the severity of the outlier.

Third, we simulated 200 realizations of a contaminated Gaussian field where 5 % of the $\varepsilon(\boldsymbol{s})$ had been replaced by normal variates with ten times as large standard deviation. The boxplots of the estimates are shown in Figure 3. The contamination strongly inflated the non-robust ($c = 1000$)

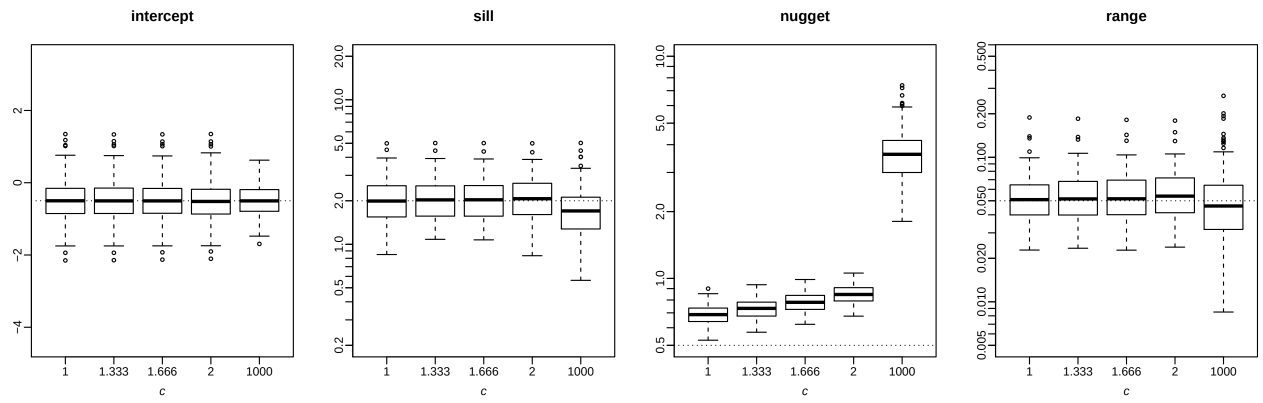

Figure 3: Boxplots of estimates $\tilde{\boldsymbol{\beta}}$, $\tilde{\sigma}^2$ and $\tilde{\boldsymbol{\theta}}$ computed from 200 contaminated Gaussian data sets (500 $Y(\boldsymbol{s})$ on centred regular grid in $[0,1]$; no trend ($\beta = -0.5$); exponential variogram with $\sigma^2 = 0.5$, $\sigma_0^2 = 2$, $\alpha = 0.05$; 5 % of $\varepsilon(\boldsymbol{s})$ randomly replaced by normal variates with mean zero and standard deviation 7.9) (dotted lines mark true parameter values, for $c = 1000$ estimates correspond to customary REML estimates).

REML estimate of the nugget, biased (negatively) the REML estimate of $\boldsymbol{\theta}$ and increased the variances of the estimates of all three covariance parameters. The robust estimates of $\boldsymbol{\theta}$ showed no systematic errors and $\tilde{\sigma}^2$ was much less biased and more efficient than the REML estimate. The systematic error in $\tilde{\sigma}^2$ decreased with decreasing c.

Summary and conclusions

We developed a novel robust REML method to estimate the external drift and the variogram parameters of Gaussian spatial data that are possibly contaminated by outliers. The methods also provides robustified universal kriging predictions. Simulations showed that our procedure bounds the influence of outliers on parameter estimates and predictions, and yields approximately Fisher consistent estimates at the Gaussian model. For large c, the estimates and predictions coincide with the customary REML estimates and the plug-in UK predictions.

So far, we did not yet check whether (19) provides an accurate approximation of Cov $[\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\beta}}^{\mathrm{T}}]$ and Cov $[(\tilde{\boldsymbol{z}} - \boldsymbol{z}), (\tilde{\boldsymbol{z}} - \boldsymbol{z})^{\mathrm{T}}]$. These quantities are required for computing the mean square prediction error at non-sampled $\boldsymbol{s}$. Furthermore, we have to develop robust procedures for testing hypotheses about $\boldsymbol{\beta}$. Some of these further developments will be presented at the conference, along with an analysis of data on heavy metals in the soils around a metal smelter in Switzerland where removal and displacement of contaminated soil has led to unpredictable irregularities in the spatial distribution of the metals.

Acknowledgements

We gratefully acknowledge the support of the Swiss Federal Office for the Environment (FOEN), which funded a part of this research.

REFERENCES

Cressie, N. A. C. (1993). *Statistics for Spatial Data*. John Wiley & Sons, New York, revised edition.

Dowd, P. A. (1984). The variogram and kriging: Robust and resistant estimators. In G. Verly, M. David,

A. Journel, and A. Maréchal, editors, *Geostatistics for Natural Resources Characterization*, volume 1, pages 91–106, Dordrecht. D. Reidel Publishing Company.

Harville, D. A. (1974). Bayesian inference for variance components using only error contrasts. *Biometrika*, **61**, 383–385.

Hawkins, D. M. and Cressie, N. (1984). Robust kriging—a proposal. *Mathematical Geology*, **16**, 3–18.

Lark, R. M. (2000). A comparison of some robust estimators of the variogram for use in soil survey. *European Journal of Soil Science*, **51**, 137–157.

Marchant, B. P. and Lark, R. M. (2007). Robust estimation of the variogram by residual maximum likelihood. *Geoderma*, **140**, 62–72.

Maronna, R. A., Martin, R. D., and Yohai, V. J. (2006). *Robust Statistics Theory and Methods*. John Wiley & Sons, Chichester.

Militino, A. F. and Ugarte, M. D. (1997). A GM estimation of the location parameters in a spatial linear model. *Communications in Statistics, Theory and Methods*, **26**(7), 1701–1725.

Patterson, H. D. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, **58**(3), 545–554.

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Richardson, A. M. and Welsh, A. H. (1995). Robust restricted maximum likelihood in mixed linear models. *Biometrics*, **51**, 1429–1439.

Schwierz, C., Papritz, A., Stahel, W. A., and Künsch, H. R. (2010). Robust REML estimation of the variogram and robust kriging. In *Proceedings of the International Conference on Robust Statistics Prague 2010*, Book of Abstracts, pages 91–92, Prague. http://icors2010.karlin.mff.cuni.cz/Book-of-abstracts-ICORS-2010.pdf.

Welsh, A. H. and Richardson, A. M. (1997). Approaches to the robust estimation of mixed models. In *Robust Inference*, volume 15 of *Handbook of Statistics*, pages 343–384. Elsevier.