ETH zürich

ETH Library

Predicting Threshold Exceedance by Local Block Means in Soil Pollution Surveys

Journal Article

Author(s):
Hofer, Christoph; Papritz, Andreas Jürg

Publication date:
2010-08

Permanent link:
https://doi.org/10.3929/ethz-b-000156400

Rights / license:
In Copyright - Non-Commercial Use Permitted

Originally published in:
Mathematical Geosciences 42(6), https://doi.org/10.1007/s11004-010-9287-4

This page was generated automatically upon download from the ETH Zurich Research Collection.
For more information, please consult the Terms of use.

Math Geosci (2010) 42: 631–656
DOI 10.1007/s11004-010-9287-4

Predicting Threshold Exceedance by Local Block Means in Soil Pollution Surveys

Christoph Hofer · Andreas Papritz

Received: 1 October 2009 / Accepted: 24 May 2010 / Published online: 18 June 2010
© International Association for Mathematical Geosciences 2010

Abstract Soil contamination by heavy metals and organic pollutants around industrial premises is a problem in many countries around the world. Delineating zones where pollutants exceed tolerable levels is a necessity for successfully mitigating related health risks. Predictions of pollutants are usually required for blocks because remediation or regulatory decisions are imposed for entire parcels. Parcel areas typically exceed the observation support, but are smaller than the survey domain. Mapping soil pollution therefore involves a local change of support. The goal of this work is to find a simple, robust, and precise method for predicting block means (linear predictions) and threshold exceedance by block means (nonlinear predictions) from data observed at points that show a spatial trend. By simulations, we compared the performance of universal block kriging (UK), Gaussian conditional simulations (CS), constrained (CK), and covariance-matching constrained kriging (CMCK), for linear and nonlinear local change of support prediction problems. We considered Gaussian and positively skewed spatial processes with a nonstationary mean function and various scenarios for the autocorrelated error. The linear predictions were assessed by bias and mean square prediction error and the nonlinear predictions by bias and Peirce skill scores.

For Gaussian data and blocks with locally dense sampling, all four methods performed well, both for linear and nonlinear predictions. When sampling was sparse CK and CMCK gave less precise linear predictions, but outperformed UK for nonlinear predictions, irrespective of the data distribution. CK and CMCK were only outperformed by CS in the Gaussian case when threshold exceedance was predicted by the conditional quantiles. However, CS was strongly biased for the skewed data whereas CK and CMCK still provided unbiased and quite precise nonlinear predictions. CMCK did not show any advantages over CK. CK is as simple to compute as UK. We recommend therefore this method to predict block means and nonlinear

C. Hofer (✉) · A. Papritz
Institute of Terrestrial Ecosystems, ETH Zurich, Universitätstrasse 16, 8092 Zürich, Switzerland
e-mail: christoph.hofer@alumni.ethz.ch

transforms thereof because it offers a good compromise between robustness, simplicity, and precision.

Keywords Constrained kriging · Covariance-matching constrained kriging · Local change of support · Nonlinear predictions

1 Introduction

Heavy metal contamination of soils by emissions of metal processing industries is a world wide problem. Several such cases have been documented in recent years (Altfelder et al. 2002; Buxton et al. 1997; Frangi and Richard 1997; McGrath et al. 2004; Van Meirvenne et al. 1993; Papritz et al. 2005; Rawlins et al. 2006; Saito and Goovaerts 2001). The metals, emitted as dust or fumes, are dispersed in the atmosphere and brought back to the ground by dry and wet deposition. Hence, the metal content of the soil usually decreases with increasing distance from the source. Regional topography and wind may further influence the contamination pattern (Frangi and Richard 1997). With respect to emissions by stationary point sources, heavy metals are not the only contaminants of concern: Emissions of persistent organic pollutants such as polycyclic aromatic hydrocarbons (Van Brummelen et al. 1996), polychlorinated biphenyls (Xing et al. 2005), or dioxins (Goovaerts et al. 2008) by chemical industries create similar contamination patterns. The contamination of the soils near the industrial premises is sometimes so severe that the health of humans and other organisms is at risk. To avert harm from the subjects of protection, one must therefore delineate the zone where the pollutants exceed tolerable levels. Many countries (e.g., Germany, the Netherlands, Switzerland, United Kingdom) enacted soil protection laws to establish allowable maximum concentrations and defined regulations on how to proceed if they are exceeded. In general, protective measures are imposed for entire parcels of land (Papritz et al. 2005), and the mean pollutant content of some soil layer (usually topsoil) on the parcel is decisive for the measures to be taken. To adopt the usual geostatistical terminology, we use in the sequel block and block mean for the parcel and the mean of the target variable on the parcel.

In regional surveys of soil contamination around a known source, soil is usually sampled at different distance and orientation from the source (Starks et al. 1987). Compared to the area of the blocks, the spatial support of the samples, i.e., the area over which the material of a bulked soil sample is collected, is usually much smaller than the blocks. At the same time, the blocks are several orders of magnitude smaller than the survey domain. Thus, for planning protective measures, one faces a nonlinear local change of support problem (Chilès and Delfiner 1999, pp. 435–437; Gotway and Young 2002). Based on usually sparse quasi-point-support observations, one has to predict for blocks—which are small compared to the area of the survey domain—whether their means exceed a threshold. The popular approach to such nonlinear and nonstationary prediction problems is conditional Gaussian simulation (CS) (Chilès and Delfiner 1999, pp. 449–592). However, Aldworth and Cressie (2003) rightly observe that CS is highly parametric. CS predictions of nonlinear functionals of a spatial variable may be badly biased if the probabilistic model is misspecified. Universal

block kriging (UK) (Cressie 1993, p. 155) is less sensitive to model misspecification. But UK predictions of nonlinear functionals of a spatial variable are often severely biased because UK underestimates the variances of the target quantities (Aldworth and Cressie 2003). The constrained kriging (CK) predictor, proposed by Cressie in 1993, is less biased than UK in this instance, and it is exactly unbiased for nonlinear functionals of a Gaussian variable. Like UK, the CK predictor is linear in the data, but it satisfies the unbiasedness constraint of UK, as well as a second constraint that matches the variances of the predictions to the variances of the block means. Aldworth and Cressie (2003) extended CK to covariance-matching constrained kriging (CMCK), which matches for a set of blocks both the variances and covariances of predictions and block means. Like CK, CMCK is less biased than UK for nonlinear predictions and exactly unbiased if the spatial variable is Gaussian.

Aldworth and Cressie (2003) compared the performance of CS, CK, CMCK, and ordinary kriging (OK) by simulation for various stationary random process models. Among other target quantities, the following predictands were considered: (i) the fraction of the simulation domain where the spatial variable was below a threshold and (ii) its mean over the area where the threshold was exceeded. Although not best, CMCK and CK consistently performed well in the simulations. The performance of CS and OK was more variable. For the non-Gaussian models, CS was best (with respect to the mean square error) for predictand (i) and, excepting the lognormal process, worst for predictand (ii). Rather surprisingly, OK showed the reverse performance.

Clearly, the predictands considered by Aldworth and Cressie (2003) were nonlinear functionals of the variable of interest, but they were global quantities (Chilès and Delfiner 1999, pp. 430–434), because they depended on the spatial distribution of the target variable in the whole domain of interest. In soil pollution surveying, as set out before, one is mostly interested in predicting local nonlinear functionals (Chilès and Delfiner 1999, pp. 435–437) of a spatial variable. Aldworth and Cressie's study is inconclusive in this respect. Moreover, the authors did not investigate the performance of the methods for models with nonstationary means. These two issues are not addressed by the other studies (Cressie and Johannesson 2001; Cressie et al. 2006; Tercan 2004) published to date on CK and CMCK. Thus, we lack information about their performance in local change of support problems. The goal of this work is therefore to explore their merits and disadvantages in comparison to CS and UK for predicting both linear (block means) and nonlinear functionals (threshold exceedance by block means) of a spatial variable. To this end, we used simulations, and we studied Gaussian and positively skewed random processes with a common nonstationary mean function and various scenarios for the autocorrelated component of the models.

The remainder of the article is organised as follows: Sect. 2 reviews spatial interpolation methods for (nonlinear) block predictions. This section contains a summary of CK and CMCK. Section 3 describes the simulation experiment (models, target quantities, validation criteria). Section 4 presents the results of the simulations and, by discussing the performance of CK relative to UK, sheds some further light on the properties of this method. Finally, Sect. 5 concludes with some advice on the choice of methods in nonlinear predictions problems with local change of support.

2 Review of Spatial Interpolation Methods for Block Prediction

In this work we consider a spatial random process $\{Z(\mathbf{s}):\mathbf{s} \in D\}$ in a survey domain $D \subset \mathbb{R}^2$ around a stationary point source, where $\mathbf{s} = (x, y)'$ indicates a location in D ($'$ denotes transpose). Let $\mathbf{Z} = (Z(\mathbf{s}_1), Z(\mathbf{s}_2), \ldots, Z(\mathbf{s}_n))'$ denote the vector of random variables that model the observations $z(\mathbf{s}_i)$, $i = 1, 2, \ldots, n$. For $Z(\mathbf{s})$ we assume the model

$$Z(\mathbf{s}) = Y(\mathbf{s}) + \epsilon(\mathbf{s}) = \mu(\mathbf{s}) + \delta(\mathbf{s}) + \epsilon(\mathbf{s}), \tag{1}$$

where $Y(\mathbf{s}) = \mu(\mathbf{s}) + \delta(\mathbf{s})$ is the variable of interest or signal, whose expectation $\mathrm{E}[Y(\mathbf{s})] = \mu(\mathbf{s}) = \mathbf{x}(\mathbf{s})'\boldsymbol{\beta}$ is modelled by a linear regression with $\mathbf{x}(\mathbf{s})$ denoting the vector with the p covariates for location $\mathbf{s}$ and $\boldsymbol{\beta}$ the vector with the p regression coefficients; $\delta(\mathbf{s})$ is a zero mean weakly stationary variable with isotropic covariance function $\mathrm{Cov}[\delta(\mathbf{s}), \delta(\mathbf{s}+\mathbf{h})] = \mathrm{Cov}[Y(\mathbf{s}), Y(\mathbf{s}+\mathbf{h})] = C(h)$, $h = \|\mathbf{h}\|$; and $\epsilon(\mathbf{s})$ is a zero mean white noise variable with variance σ_ϵ^2.

Let

$$Y(B_i) = \frac{1}{|B_i|}\int_{B_i} Y(\mathbf{s})\,\mathrm{d}\mathbf{s}$$

denote the mean of the signal over some block B_i of land with area $|B_i| = \int_{B_i} \mathrm{d}\mathbf{s} \ll |D|$. We consider m such blocks and denote the vector of mean values by $\mathbf{Y} = (Y(B_1), Y(B_2), \ldots, Y(B_m))'$. Suppose we wish to predict $g(\mathbf{Y})$ where $g(\cdot)$ is a (possibly nonlinear) scalar function of $\mathbf{Y}$. We consider predictors of the form $g(\widehat{\mathbf{Y}})$, where $\widehat{\mathbf{Y}} = \boldsymbol{\Lambda}'\mathbf{Z}$ is a linear predictor of $\mathbf{Y}$, and $\boldsymbol{\Lambda} = (\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_m)$ is a $n \times m$ matrix of weights. Note that we confine ourselves to compute point predictions of $g(\mathbf{Y})$ and its variance rather than estimating the conditional distribution of $g(\mathbf{Y})$ given the data.

For squared error loss, known $C(\mathbf{h})$ and σ_ϵ^2, and linear $g(\cdot)$, the universal kriging predictor (UK), $g(\widehat{\mathbf{Y}}_{\mathrm{UK}})$, is best (Harville 1977, p. 322), where $\widehat{\mathbf{Y}}_{\mathrm{UK}}$ is of course given by

$$\widehat{\mathbf{Y}}_{\mathrm{UK}} = \boldsymbol{\Lambda}_{\mathrm{UK}}'\mathbf{Z} = \mathbf{X}_m\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}} + \mathbf{C}'\boldsymbol{\Sigma}^{-1}(\mathbf{Z} - \mathbf{X}\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}}), \tag{2}$$

and $\boldsymbol{\Lambda}_{\mathrm{UK}} = \boldsymbol{\Sigma}^{-1}\{(\mathbf{I} - \mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1})\mathbf{C} + \mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}_m'\}$ is the $n \times m$ matrix of the UK weights; $\mathbf{X}_m = (\mathbf{x}(B_1), \ldots, \mathbf{x}(B_m))'$ and $\mathbf{X} = (\mathbf{x}(\mathbf{s}_1), \ldots, \mathbf{x}(\mathbf{s}_n))'$ are the $m \times p$ and $n \times p$ design matrices of the target blocks and the data, respectively; $\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Z}$ is the p vector with the generalised least square estimate of $\boldsymbol{\beta}$; $\mathbf{C} = (\mathbf{c}(\mathbf{s}_{1\ldots n}, B_1), \ldots, \mathbf{c}(\mathbf{s}_{1\ldots n}, B_m))$ is a $n \times m$ matrix where $\mathbf{c}(\mathbf{s}_{1\ldots n}, B_i) = (C(\mathbf{s}_1, B_i), \ldots, C(\mathbf{s}_n, B_i))'$ are the n covariances between $\mathbf{Z}$ and $Y(B_i)$; $\boldsymbol{\Sigma} = \mathrm{Cov}[\mathbf{Z}, \mathbf{Z}']$ is the $n \times n$ covariance matrix of the data and $\mathbf{I}$ is the $n \times n$ identity matrix. $\widehat{\mathbf{Y}}_{\mathrm{UK}}$ typically underestimates the true variation of $\mathbf{Y}$. Consequently, $g(\widehat{\mathbf{Y}}_{\mathrm{UK}})$ is a biased predictor of $g(\mathbf{Y})$ for nonlinear $g(\cdot)$ (Appendix A).

For nonlinear $g(\cdot)$, UK is usually given up in favour of Gaussian conditional simulations (CS). Conditional realisations of $\mathbf{Y}$ given $\mathbf{Z}$, say $\mathbf{Y}_\omega|\mathbf{Z}$, can be efficiently simulated by (Chilès and Delfiner 1999, pp. 465–472)

$$\mathbf{Y}_\omega|\mathbf{Z} = \widehat{\mathbf{Y}}_{\mathrm{UK}} + \left(\mathbf{Y}_\omega - \mathbf{C}'\boldsymbol{\Sigma}^{-1}\mathbf{Z}_\omega\right), \tag{3}$$

where $(\mathbf{Y}'_\omega, \mathbf{Z}'_\omega)$ is a $m+n$ vector with an unconditionally simulated realisation of $(\mathbf{Y}', \mathbf{Z}')$ (ω denotes a realisation). The marginal distributions of $\mathbf{Y}_\omega|\mathbf{Z}$ and $\mathbf{Y}$ are the same as long as the latter is also Gaussian (Chilès and Delfiner 1999, p. 466). Hence, any realisation of $g(\mathbf{Y}_\omega|\mathbf{Z})$ is an unbiased predictor of $g(\mathbf{Y})$ for nonlinear $g(\cdot)$. Of course, a single $g(\mathbf{Y}_\omega|\mathbf{Z})$ is highly variable, and one uses in practise the mean $\frac{1}{N}\sum_{\omega=1}^{N} g(\mathbf{Y}_\omega|\mathbf{Z})$ of N realisations as an approximation of $\mathrm{E}[g(\mathbf{Y})|\mathbf{Z}]$.

Aldworth and Cressie (2003) pursued another approach and proposed to predict $g(\mathbf{Y})$ by $g(\widehat{\mathbf{Y}}_{\mathrm{CMCK}})$, where $\widehat{\mathbf{Y}}_{\mathrm{CMCK}} = \mathbf{A}'\mathbf{Z}$ is the covariance-matching constrained kriging (CMCK) predictor and $\mathbf{A} = (\mathbf{a}_1, \ldots, \mathbf{a}_m)$ is a $n \times m$ matrix of the weights that satisfy in addition to the usual unbiasedness constraints

$$\mathrm{E}[\mathbf{A}'\mathbf{Z}] = \mathrm{E}[\mathbf{Y}], \tag{4}$$

of UK the covariance constraints

$$\mathrm{Cov}\big[\mathbf{A}'\mathbf{Z}, (\mathbf{A}'\mathbf{Z})'\big] = \mathrm{Cov}[\mathbf{Y}, \mathbf{Y}']. \tag{5}$$

Consequently, $g(\widehat{\mathbf{Y}}_{\mathrm{CMCK}})$ is approximately unbiased for any smooth, nonlinear $g(\cdot)$ and exactly unbiased if $\mathbf{Y}$ is Gaussian (Appendix A). The CMCK predictor of $\mathbf{Y}$ is given by Aldworth and Cressie (2003)

$$\widehat{\mathbf{Y}}_{\mathrm{CMCK}} = \mathbf{A}'\mathbf{Z} = \mathbf{X}_m\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}} + \mathbf{K}'\mathbf{C}'\boldsymbol{\Sigma}^{-1}(\mathbf{Z} - \mathbf{X}\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}}), \tag{6}$$

where $\mathbf{A} = \boldsymbol{\Sigma}^{-1}\{(\mathbf{I} - \mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1})\mathbf{C}\mathbf{K} + \mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'_m\}$ is the $n \times m$ matrix of the CMCK weights, $\mathbf{K} = \mathbf{Q}_1^{-1}\mathbf{P}_1$ is a $m \times m$ matrix and $\mathbf{Q}_1$ and $\mathbf{P}_1$ are defined by

$$\mathbf{Q}_1\mathbf{Q}_1 = \mathbf{Q} = \mathrm{Cov}[\widehat{\mathbf{Y}}_{\mathrm{UK}}, \widehat{\mathbf{Y}}'_{\mathrm{UK}}] - \mathrm{Cov}\big[\mathbf{X}_m\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}}, (\mathbf{X}_m\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}})'\big], \tag{7}$$

$$\mathbf{P}_1\mathbf{P}_1 = \mathbf{P} = \mathrm{Cov}[\mathbf{Y}, \mathbf{Y}'] - \mathrm{Cov}\big[\mathbf{X}_m\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}}, (\mathbf{X}_m\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}})'\big]. \tag{8}$$

The symmetric $m \times m$ matrices $\mathbf{Q}_1$ and $\mathbf{P}_1$ exist and are positive definite if $\mathbf{Q}$ and $\mathbf{P}$ are themselves positive definite, and they can be uniquely determined by the square root decomposition described by Harville (1997, pp. 543–545). In practise, the CMCK predictor thus exists if $\mathbf{Q}$ and $\mathbf{P}$ are positive definite (Aldworth and Cressie 2003, p. 14). Unlike $\mathbf{Q}$, which is always nonnegative definite (Aldworth and Cressie 2003, p. 15) the matrix $\mathbf{P}$ may become negative definite, and the CMCK predictor no longer exists. This is likely to happen if m is large, if the areas of the blocks are large compared to the support of the samples and if one extrapolates the trend, i.e., if the $\mathbf{x}(B_i)$ are far from the covariates at the sample locations.

For $m = 1$, $\mathbf{Y}$, $\widehat{\mathbf{Y}}_{\mathrm{UK}}$, $\mathbf{P}$ and $\mathbf{Q}$ reduce to the scalars $Y(B_1)$, $\widehat{Y}_{\mathrm{UK}}(B_1)$, P and Q, respectively, $\mathbf{K}$ is thus also a scalar

$$K = \left(\frac{P}{Q}\right)^{\frac{1}{2}} = \left(\frac{\mathrm{Var}[Y(B_1)] - \mathrm{Var}[\mathbf{x}(B_1)'\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}}]}{\mathrm{Var}[\widehat{Y}_{\mathrm{UK}}(B_1)] - \mathrm{Var}[\mathbf{x}(B_1)'\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}}]}\right)^{\frac{1}{2}}, \tag{9}$$

and the CMCK predictor simplifies to the constrained kriging (CK) predictor

$$\widehat{Y}_{\mathrm{CK}}(B_1) = \mathbf{x}(B_1)'\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}} + K\mathbf{c}(\mathbf{s}_{1\ldots n}, B_1)'\boldsymbol{\Sigma}^{-1}(\mathbf{Z} - \mathbf{X}\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}}). \tag{10}$$

K exists if the numerator and denominator of (9) are positive.

Of course, $Q \geq 0$ (this follows from the fact that $\mathbf{Q}$ is nonnegative definite) but P may become negative if $|B_1|$ is large ($\mathrm{Var}[Y(B_1)]$ small) and/or if the trend is extrapolated ($\mathrm{Var}[\mathbf{x}(B_1)'\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}}]$ large).

There is a price to pay for the covariance constraint in (5); the CMCK and the CK predictors have larger mean square prediction errors (MSPE) than the UK predictor. It is not difficult to show that

$$\mathrm{MSPE}[\widehat{\mathbf{Y}}_{\mathrm{CMCK}}] = \mathrm{MSPE}[\widehat{\mathbf{Y}}_{\mathrm{UK}}] + (\mathbf{P}_1 - \mathbf{Q}_1)(\mathbf{P}_1 - \mathbf{Q}_1), \tag{11}$$

where

$$\begin{aligned}\mathrm{MSPE}[\widehat{\mathbf{Y}}_{\mathrm{UK}}] &= \mathrm{Cov}\big[\widehat{\mathbf{Y}}_{\mathrm{UK}} - \mathbf{Y}, (\widehat{\mathbf{Y}}_{\mathrm{UK}} - \mathbf{Y})'\big] \\ &= \mathrm{Cov}[\mathbf{Y}, \mathbf{Y}'] - \mathbf{C}'\boldsymbol{\Sigma}^{-1}\mathbf{C} \\ &\quad + \big(\mathbf{X}'_m - \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{C}\big)'\big(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\big)^{-1}\big(\mathbf{X}'_m - \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{C}\big)\end{aligned} \tag{12}$$

is the covariance matrix of the UK predictions errors. The second term of the right-hand side of (11) is a positive definite matrix and therefore any linear combination, say $\boldsymbol{\nu}'\widehat{\mathbf{Y}}_{\mathrm{CMCK}}$, has a larger mean square prediction error than $\boldsymbol{\nu}'\widehat{\mathbf{Y}}_{\mathrm{UK}}$.

For $m = 1$ (11) simplifies to

$$\begin{aligned}\mathrm{MSPE}\big[\widehat{Y}_{\mathrm{CK}}(B_1)\big] &= \mathrm{MSPE}\big[\widehat{Y}_{\mathrm{UK}}(B_1)\big] + \big(\sqrt{P} - \sqrt{Q}\big)^2 \\ &= \mathrm{MSPE}\big[\widehat{Y}_{\mathrm{UK}}(B_1)\big] + Q(K-1)^2.\end{aligned} \tag{13}$$

The CMCK predictor has a larger MSPE than the CK predictor. From (11) and the fact that $\mathbf{P}_1$ and $\mathbf{Q}_1$ are symmetric immediately follows

$$\begin{aligned}&\mathrm{MSPE}\big[\widehat{Y}_{\mathrm{CMCK}}(B_k)\big] \\ &= \mathrm{MSPE}\big[\widehat{Y}_{\mathrm{UK}}(B_k)\big] + \sum_{i=1}^{m}\big([\mathbf{P}_1]_{ik} - [\mathbf{Q}_1]_{ik}\big)^2 \\ &= \mathrm{MSPE}\big[\widehat{Y}_{\mathrm{UK}}(B_k)\big] + \big([\mathbf{P}_1]_{kk} - [\mathbf{Q}_1]_{kk}\big)^2 + \sum_{i=1, i\neq k}^{m}\big([\mathbf{P}_1]_{ik} - [\mathbf{Q}_1]_{ik}\big)^2 \\ &= \mathrm{MSPE}\big[\widehat{Y}_{\mathrm{CK}}(B_k)\big] + \underbrace{\sum_{i=1, i\neq k}^{m}\big([\mathbf{P}_1]_{ik} - [\mathbf{Q}_1]_{ik}\big)^2}_{\geq 0},\end{aligned}$$

where $[\mathbf{A}]_{ik}$ is the element in the ith row and kth column of $\mathbf{A}$.

3 Simulation Experiment

We compared the performance of CS, UK, CK, and CMCK by simulations. We were guided by the example of the metal smelter in Dornach, NW Switzerland (Papritz et al. 2005), when we designed the simulation experiment.

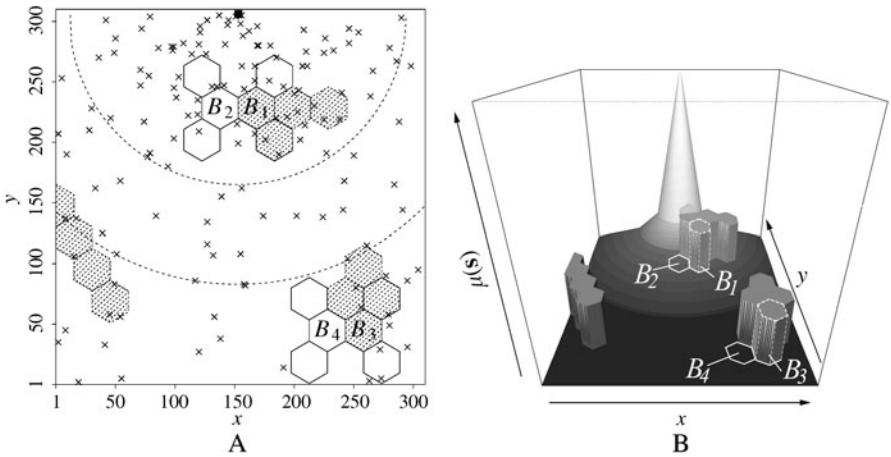

Fig. 1 (**A**) Simulation domain D. The crosses denote the 146 sample points, the hexagons B_1–B_4 are the target blocks. (**B**) Pseudo-three-dimensional display of the trend function, $\mu(\mathbf{s})$, used in the simulations. See text for details

3.1 Simulation Domain

We used a two-dimensional domain D, discretized by a square grid with 305×305 nodes, for the simulation experiments (Fig. 1A). The assumed source of the pollutants (symbol ■ in Fig. 1A) was located outside of D at $\mathbf{s}_{\mathrm{PS}} = (x = 153, y = 306)'$. We had four target blocks, $B_1, \ldots, B_4$, all with hexagonal shape and area $|B_i| = 800$. Compared to real soil pollution cases, the blocks were quite large relative to the area of the simulation domain, but we made this choice on purpose because we wanted to see if the CK and CMCK predictors were likely to fail to exist. For covariance matching constrained kriging, each target block had three neighbours (hexagons with solid lines in Fig. 1A), each sharing an edge with the target block. B_1 and B_2 as well as B_3 and B_4 were neighbours. We chose 146 sample points (marked by crosses in Fig. 1A) where the values of the simulated fields were assumed to be observed. We selected more sample points close to the source than at farther distance from it. To this aim, we divided D into three zones by two concentric circles (dashed circles in Fig. 1A) with radii 141 and 223 centred on $\mathbf{s}_{\mathrm{PS}}$. Then we selected randomly 88, 32, and 26 grid nodes within the three zones. The support of the measurements was equal to one grid cell, i.e., $|\mathbf{s}_i| = 1, i = 1, \ldots, 146$.

3.2 Spatial Models

We used two variants of (1) as spatial models for the simulations. Both models shared the same trend, $\mu(\mathbf{s})$, but they differed in the way they modelled the stochastic model component $\delta(\mathbf{s})$.

3.2.1 Modelling the Trend

The trend function $\mu(\mathbf{s})$ mimicked the topsoil heavy metal contamination around the Dornach metal smelter. We modelled a piecewise linear, isotropic decrease of

$\mu(\mathbf{s})$ with increasing distance $d(\mathbf{s}) = |\mathbf{s} - \mathbf{s}_{\mathrm{PS}}|$ from the source. In Dornach, we had observed that the metal concentrations were larger in the soils of forests compared to open land. We therefore assumed that the pollution was more severe in three distinct zones of D (groups of hatched polygons in Fig. 1A). In the remainder, we denote these zones by forests. Note that the target blocks B_1 and B_3 belonged to the forests.

In more detail, we modelled the trend by a segmented or broken stick regression (Faraway 2004, pp. 121–122) with 3 knot distances, c_i, and combined this with an indicator term for forests

$$\mu(\mathbf{s}) = \beta_0 + \sum_{j=1}^{i-1} \beta_j (c_j - c_{j-1}) + \beta_i \big(d(\mathbf{s}) - c_{i-1}\big) + \beta_5 I(\mathbf{s} \in \text{forest}), \tag{14}$$

where $i = 1$ if $c_0 < d(\mathbf{s}) \le c_1$, $i = 2$ if $c_1 < d(\mathbf{s}) \le c_2$, $i = 3$ if $c_2 < d(\mathbf{s}) \le c_3$, $i = 4$ if $d(\mathbf{s}) > c_3$ and

$$I(x) = \begin{cases} 1, & \text{if } x \text{ is true,} \\ 0, & \text{otherwise.} \end{cases} \tag{15}$$

In the simulation, we set $\boldsymbol{\beta} = (104, -1.4, -0.2, -0.05, 0, 15)'$. The knots for the distances were equal to $c_0 = 0, c_1 = 40, c_2 = 75, c_3 = 200$ and $c_4 = \infty$. Figure 1B illustrates this trend function.

3.2.2 Modelling the Stochastic Component

As aforementioned, we considered two different models for the signal process: a Gaussian, $\{\delta_{\mathrm{G}}(\mathbf{s})\}$, and a positively skewed spatial process, $\{\delta_{\mathrm{S}}(\mathbf{s})\}$. To simulate $\{\delta_{\mathrm{G}}(\mathbf{s})\}$, we chose an isotropic exponential covariance function

$$C_{\mathrm{G}}(h) = \sigma^2_{\mathrm{G},0} I(h = 0) + \sigma^2_{\mathrm{G},1} \exp(-h/\alpha), \tag{16}$$

with nugget $\sigma^2_{\mathrm{G},0} = 7$, partial sill $\sigma^2_{\mathrm{G},1} = 63$ and range parameter $\alpha = 25$. At the sample points, we added independent, identically (iid) normally distributed zero mean measurements errors, $\epsilon(\mathbf{s}_i)$, $\mathrm{Var}[\epsilon(\mathbf{s}_i)] = 3.5$, to the signal.

As we aimed for a positively skewed process with an additive white noise component, we simulated $\{\delta_{\mathrm{S}}(\mathbf{s})\}$ by adding two standardised, lognormal spatial random fields, of which one, $\{\exp(\delta_{\mathrm{G},0}(\mathbf{s}))\}$, was lognormal white noise and the other, $\{\exp(\delta_{\mathrm{G},1}(\mathbf{s}))\}$, spatially autocorrelated. Thus, $\{\delta_{\mathrm{S}}(\mathbf{s})\}$ was simulated from

$$\delta_{\mathrm{S}}(\mathbf{s}) = \sum_{i=0}^{1} \delta_{\mathrm{S,i}}(\mathbf{s}) = \sum_{i=0}^{1} \sigma_{\mathrm{S},i} \frac{\exp(\delta_{\mathrm{G},i}(\mathbf{s})) - \mathrm{E}[\exp(\delta_{\mathrm{G},i}(\mathbf{s}))]}{\sqrt{\mathrm{Var}[\exp(\delta_{G,i}(\mathbf{s}))]}}, \tag{17}$$

where

$$\mathrm{E}\big[\exp\big(\delta_{\mathrm{G},i}(\mathbf{s})\big)\big] = \exp\left(\mathrm{E}\big[\delta_{G,i}(\mathbf{s})\big] + \frac{\sigma^2_{\mathrm{G},i}}{2}\right)$$

and

$$\mathrm{Var}\big[\exp\big(\delta_{\mathrm{G},i}(\mathbf{s})\big)\big] = \mathrm{E}\big[\exp\big(\delta_{\mathrm{G},i}(\mathbf{s})\big)\big]^2 \big(\exp\big(\sigma^2_{\mathrm{G},i}\big) - 1\big).$$

Table 1 Parameter values used to simulate the positively skewed spatial process $\{\delta_S(\mathbf{s})\}$

i	$\mathrm{E}[\delta_{\mathrm{G},i}(\mathbf{s})]$	$\sigma^2_{\mathrm{G},i}$	$\sigma^2_{\mathrm{S},i}$	α
0	3.359	0.085	80	0
1	3.107	0.588	720	25

The covariance function of $\{\delta_{\mathrm{S}}(\mathbf{s})\}$ is then equal to Diggle and Ribeiro (2007, p. 61)

$$C_{\mathrm{S}}(h) = \sigma^2_{\mathrm{S},0} l(h=0) + \sigma^2_{\mathrm{S},1} \frac{\exp(\sigma^2_{\mathrm{G},1}\exp(-h/\alpha)) - 1}{\exp(\sigma^2_{\mathrm{G},1}) - 1}. \tag{18}$$

The parameters $\sigma_{\mathrm{S},0}$, $\sigma_{\mathrm{S},1}$, α, and the means and variances of the Gaussian process required for the simulation of $\{\delta_{\mathrm{S}}(\mathbf{s})\}$ are listed in Table 1. The parameters were selected such that $\delta_{\mathrm{G}}(\mathbf{s})$ and $\delta_{\mathrm{S}}(\mathbf{s})$ had the same correlation length (Chilès and Delfiner 1999, p. 74) and that their nugget-to-total-sill ratios matched. Normal iid zero mean measurement errors were again added to the signal, but this time we chose $\mathrm{Var}[\epsilon(\mathbf{s})] = 40$. In addition to these models, we also considered Gaussian and positively skewed processes with larger nugget-to-sill ratios including models with larger measurement error variances. However, since the results did not differ much we do not describe the models here and we do not report the results.

3.3 Target Quantities

For each B_i, we studied the prediction of two quantities:

1. The block mean, $Y(B_i)$, which is a linear functional of $\{Y(\mathbf{s})\}$.
2. The binary indicator, $I(Y(B_i) > T)$, that indicates if the block mean exceeds a given threshold T. Clearly, $I(Y(B_i) > T)$ depends nonlinearly on $\{Y(\mathbf{s})\}$.

3.4 Implementation of the Simulations

We used the programming environment R (R Development Core Team 2009) for all the computations. For both signal processes, we simulated 2000 realisations on the grid with 305×305 nodes and added μ to obtain the realisations of $\{Y(\mathbf{s})\}$. Realisations of the various $\{\delta_{\mathrm{G},i}(\mathbf{s})\}$ were simulated by the circulant embedding algorithm of Chan and Wood (1997), implemented in the R package RandomFields (Schlather 2001). Realisations of the block means were computed by arithmetically averaging the values simulated at the grid nodes within the respective blocks. To obtain a realisation of the observations $\mathbf{Z}$, we added simulated measurement errors to the $Y(\mathbf{s}_i)$ at the sample points.

We compared the kriging methods UK, CK, CMCK, and conditional simulations. In addition, we computed the GLS trend surface predictions, $\mathbf{x}(B_i)'\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}}$, of the block means. In CS, we simulated 500 conditional realisations, say $Y_\omega(B_i)|\mathbf{Z}$, of the block means for each of the 2000 realisations by the kriging method (3). The block means were then predicted by the estimated conditional means,

$$\widehat{Y}_{\mathrm{CS}}(B_i) = 1/500 \sum_{\omega=1}^{500} Y_\omega(B_i)|\mathbf{Z},$$

and the exceedance of the thresholds by their medians. The latter choice should minimise the total number of misclassifications, provided the model is not misspecified. Note that up to some numerical inaccuracies, $\widehat{Y}_{\mathrm{CS}}(B_i)$ and $\widehat{Y}_{\mathrm{UK}}(B_i)$ were the same, see (3). The conditional simulations and the kriging predictions were computed with the true covariance parameters (and not with estimates inferred from the simulated values at the sample points). The (co-)variances of the block means, required for UK and (CM)CK predictions, were computed by the R package spatialCovariance (Clifford 2009). Finally, to compute the predictions, we did not use local search windows but used all the 146 observations.

3.5 Validation

3.5.1 Block Means

We assessed the quality of the predictions of the block means by calculating the empirical bias (eBIAS) and the empirical mean square prediction error (eMSPE)

$$\mathrm{eBIAS}\big[\widehat{Y}_k(B_i)\big] = \frac{1}{2000}\sum_{\omega=1}^{2000}\big(\widehat{Y}_{\omega k}(B_i) - Y_\omega(B_i)\big), \tag{19}$$

$$\mathrm{eMSPE}\big[\widehat{Y}_k(B_i)\big] = \frac{1}{2000}\sum_{\omega=1}^{2000}\big(\widehat{Y}_{\omega k}(B_i) - Y_\omega(B_i)\big)^2, \tag{20}$$

where $\widehat{Y}_{\omega k}(B_i)$ denotes the prediction of the ωth realisation, $Y_\omega(B_i)$, of the block mean by method k.

3.5.2 Threshold Exceedance

We validated predictions of threshold exceedance for multiple thresholds. In fact, for a given block B_i, we used the ordered simulated block means $Y_\omega(B_i)$, denoted as $T_{[l]}$, as thresholds. Table 2 shows the 2×2 contingency table used for validating the predictions of threshold exceedance. The quantities $\mathrm{TN}_{kl}(B_i), \ldots, \mathrm{TP}_{kl}(B_i)$ are counts, estimated from the simulation results by

$$\mathrm{TN}_{kl}(B_i) = \sum_{\omega=1}^{2000} I\big(\widehat{Y}_{\omega k}(B_i) \le T_{[l]}\big)\cdot I\big(Y_\omega(B_i) \le T_{[l]}\big),$$

$$\mathrm{FN}_{kl}(B_i) = \sum_{\omega=1}^{2000} I\big(\widehat{Y}_{\omega k}(B_i) \le T_{[l]}\big)\cdot I\big(Y_\omega(B_i) > T_{[l]}\big),$$

Table 2 Contingency table for validating whether the prediction, $\widehat{Y}_k(B_i)$, of the block mean, $Y(B_i)$, by method k exceeds the threshold $T_{[l]}$. See text for details

	$Y(B_i) \le T_{[l]}$	$Y(B_i) > T_{[l]}$
$\widehat{Y}_k(B_i) \le T_{[l]}$	$\mathrm{TN}_{kl}(B_i)$ (true negative)	$\mathrm{FN}_{kl}(B_i)$ (false negative)
$\widehat{Y}_k(B_i) > T_{[l]}$	$\mathrm{FP}_{kl}(B_i)$ (false positive)	$\mathrm{TP}_{kl}(B_i)$ (true positive)

$$\mathrm{FP}_{kl}(B_i) = \sum_{\omega=1}^{2000} I\big(\widehat{Y}_{\omega k}(B_i) > T_{[l]}\big) \cdot I\big(Y_\omega(B_i) \le T_{[l]}\big),$$

$$\mathrm{TP}_{kl}(B_i) = \sum_{\omega=1}^{2000} I\big(\widehat{Y}_{\omega k}(B_i) > T_{[l]}\big) \cdot I\big(Y_\omega(B_i) > T_{[l]}\big).$$

Various accuracy and goodness of prediction measures can be calculated from such contingency tables (Wilks 2006, pp. 255–335). We used the bias score

$$\mathrm{BS}_{kl}(B_i) = \frac{\mathrm{FP}_{kl}(B_i) + \mathrm{TP}_{kl}(B_i)}{\mathrm{FN}_{kl}(B_i) + \mathrm{TP}_{kl}(B_i)} \tag{21}$$

to see whether the kth method systematically over- or underestimated the exceedance frequency (the nominator [denominator] of (21) is the total number of predictions [simulated block means] exceeding the threshold). Thus, $\mathrm{BS}_{kl}(B_i) = 1$ indicates that method k predicts the correct number of threshold exceedances for block B_i, whereas $\mathrm{BS}_{kl}(B_i) > 1$ ($\mathrm{BS}_{kl}(B_i) < 1$) signals systematic overestimation (underestimation) of exceedance.

The bias score does not characterise the conditional bias of the predictions (BS ≈ 1 as long as FP $\approx$ FN). We therefore computed the Peirce skill score, $\mathrm{PSS}_{kl}(B_i)$, which characterises conditional bias. $\mathrm{PSS}_{kl}(B_i)$ was first proposed by Peirce (1884) and is also known as the true skill statistic or Hanssen–Kuipers discriminant (Wilks 2006, p. 266). It is defined by

$$\mathrm{PSS}_{kl}(B_i) = \frac{\mathrm{TN}_{kl}(B_i) \cdot \mathrm{TP}_{kl}(B_i) - \mathrm{FN}_{kl}(B_i) \cdot \mathrm{FP}_{kl}(B_i)}{(\mathrm{FN}_{kl}(B_i) + \mathrm{TP}_{kl}(B_i)) \cdot (\mathrm{TN}_{kl}(B_i) + \mathrm{FP}_{kl}(B_i))}. \tag{22}$$

PSS equals the difference between the hit and the false alarm rate

$$\mathrm{PSS}_{kl}(B_i) = \mathrm{H}_{kl}(B_i) - \mathrm{F}_{kl}(B_i),$$

where the hit rate,

$$\mathrm{H}_{kl}(B_i) = \frac{\mathrm{TP}_{kl}(B_i)}{\mathrm{FN}_{kl}(B_i) + \mathrm{TP}_{kl}(B_i)}, \tag{23}$$

is the conditional probability that exceedance is correctly predicted given that the block mean indeed exceeds the threshold, and the false alarm rate,

$$\mathrm{F}_{kl}(B_i) = \frac{\mathrm{FP}_{kl}(B_i)}{\mathrm{TN}_{kl}(B_i) + \mathrm{FP}_{kl}(B_i)}, \tag{24}$$

is the conditional probability of wrongly predicting exceedance when there is none.

$\mathrm{PSS}_{kl}(B_i)$ varies between -1 (perfect misclassification: $\mathrm{TN}_{kl}(B_i) = \mathrm{TP}_{kl}(B_i) = 0$) and 1 (perfect classification: $\mathrm{FN}_{kl}(B_i) = \mathrm{FP}_{kl}(B_i) = 0$). Random predictions by the marginal row and column probabilities of the contingency table result in $\mathrm{PSS}_{kl}(B_i) = 0$ (Wilks 2006, p. 266). Equation (22) shows that the Peirce skill scores for exceedance and nonexceedance are the same because PSS is invariant to swapping positives and negatives. We computed the bias and Peirce skill scores for

the ordered thresholds, $T_{[l]}, l = 1, 2, \ldots, 2000$, and plotted them against the relative ranks $\tilde{l} = \frac{l}{2000}$ of the thresholds.

4 Results and Discussion

In the sequel, we first compare how precisely the methods predicted the block means, and we discuss what factors influenced the precision of CK and CMCK and by that further explore their nature. Then we explore how successfully the methods predicted threshold exceedance. The results for the Gaussian and skewed signal processes with larger nugget-to-sill ratios are not presented because they were very similar to the ones shown below.

4.1 Block Means

As one expects from theory, the predictions of the block means were unbiased, no matter what block was predicted, what method was used, and regardless which signal process was considered (Table 3). The contribution of eBIAS^2 to the empirical mean square prediction errors (Figs. 2 and 3) was at most 2.6×10^{-3}, which confirms that the considered methods were unbiased for all practical purposes. The empirical mean square prediction errors differed more strongly than the biases between the blocks and methods. Figure 2 shows the eMSPE for the Gaussian signal process. The predictions were most precise for block B_2, followed by B_1, B_3, and B_4. CS and UK were always best, and their eMSPEs were, up to numerical inaccuracies, the same. CK and CMCK had larger eMSPE than UK or CS, with CK being consistently more precise than CMCK. This findings agree with theory, as shown in (11) and (13), but are in contrast with the simulation results of Aldworth and Cressie (2003). In their study, CK showed for linear and nonlinear predictions almost always a higher eMSPE than CMCK, which is not consistent with theory.

For B_1 to B_3, there were only small differences between CS/UK, CK, and CMCK. Furthermore, CMCK was not much worse than CK. For these blocks, kriging and CS resulted in noticeable gains in precision compared to trend surface prediction by GLS regression. By contrast, GLS regression outperformed CK and CMCK for block B_4, and it was not much worse than CS/UK for this block. For the positively skewed

Table 3 Empirical bias (eBIAS) for predicting the block means $Y(B_i)$ for the Gaussian, $\{\delta_G(\mathbf{s})\}$, and the positively skewed signal process, $\{\delta_S(\mathbf{s})\}$

	$\delta_G(\mathbf{s})$				$\delta_S(\mathbf{s})$			
	B_1	B_2	B_3	B_4	B_1	B_2	B_3	B_4
CS	−0.038	0.002	0.153	0.148	−0.038	0.028	−0.471	0.060
UK	−0.007	0.031	0.149	0.142	−0.339	0.050	−0.472	0.104
CK	−0.017	−0.003	0.102	0.073	−0.305	0.181	−0.454	0.106
CMCK	0.008	0.000	0.067	0.137	−0.226	0.065	−0.517	0.012
GLS	0.094	0.206	0.293	0.188	−0.575	−0.466	−0.517	0.103

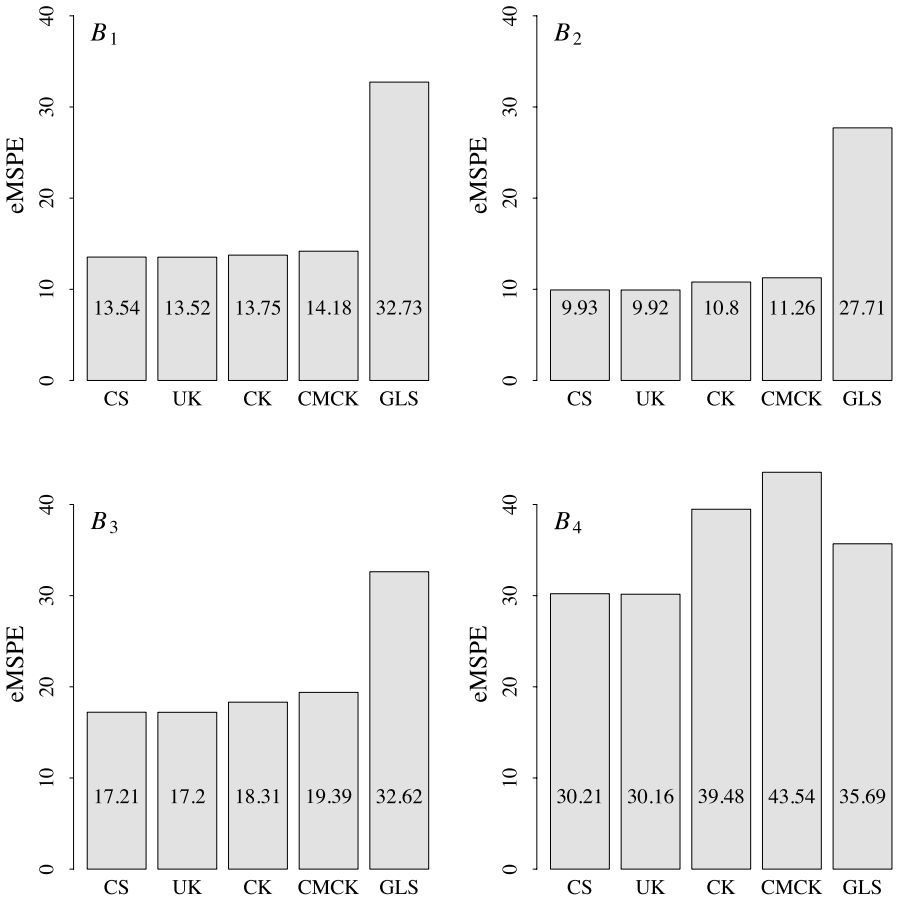

Fig. 2 Empirical mean square prediction errors for predicting the block means $Y(B_i)$ for the Gaussian signal process $\{\delta_G(\mathbf{s})\}$. B_1–B_4 are the four target blocks

signal, the empirical mean square errors showed largely the same patterns (Fig. 3). The only remarkable difference was the larger, absolute magnitude of the eMSPEs. No matter what model was used for the signal process, all the methods performed best for block B_2 followed (in order of increasing eMSPE) by B_1, B_3, and B_4. The precision of the fitted trend surface and the local density of the support points controlled how much larger the mean square errors of CMCK and CK were, compared to UK; see (11) and (13). For convenience, we discuss only CK point predictions in the sequel, but similar findings also hold for CMCK and for block predictions.

Figures 4A and 4B show the spatial distributions of $\sqrt{P}$ and $\sqrt{Q}$ for the Gaussian signal. $\sqrt{P}$, being the square root of the difference between the signal variance and the variance of the fitted trend, depends on the prediction location by the Mahalanobis distance $(\underline{\mathbf{x}}(\mathbf{s}_0) - \underline{\tilde{\bar{\mathbf{x}}}})'\mathbf{V}^{-1}(\underline{\mathbf{x}}(\mathbf{s}_0) - \underline{\tilde{\bar{\mathbf{x}}}})$ where $\underline{\mathbf{x}}(\mathbf{s}_0)$ is defined by $\mathbf{x}(\mathbf{s}_0)' = (1, \underline{\mathbf{x}}(\mathbf{s}_0)')$, $\underline{\tilde{\bar{\mathbf{x}}}}$ is the centroid of the orthogonalised covariates at the sample locations and $\mathbf{V}$ is the cross-product matrix of the centred, orthogonalised covariates (Appendix B).

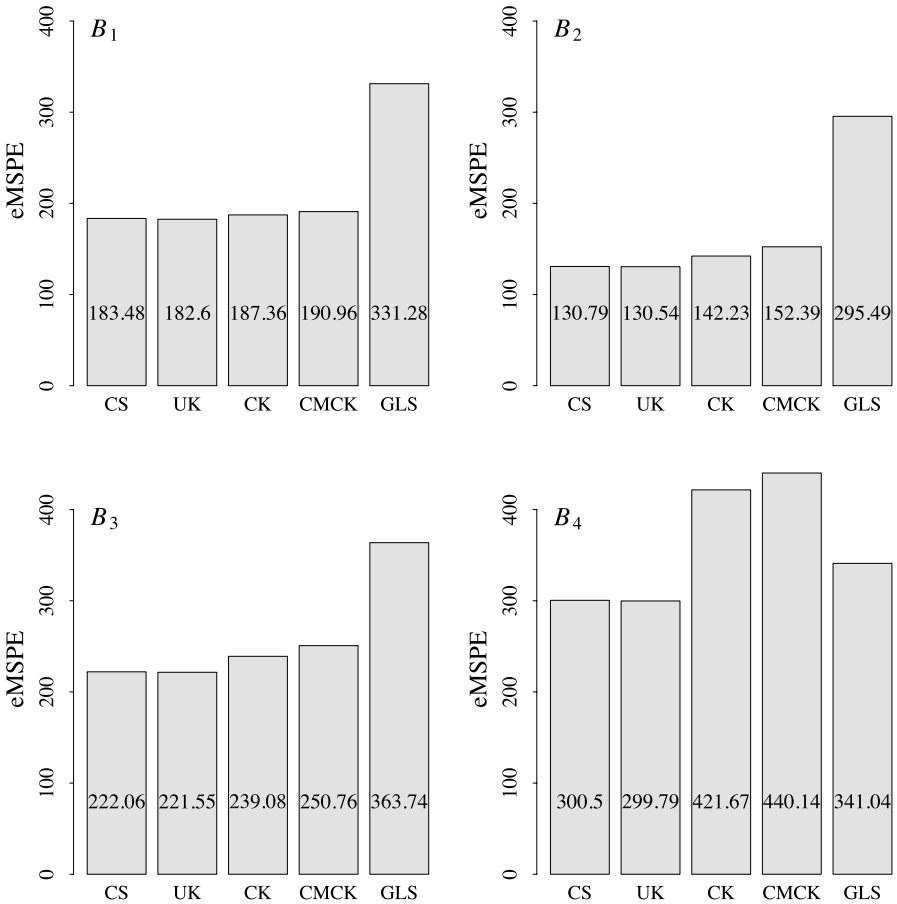

Fig. 3 Empirical mean square prediction errors for predicting the block means $Y(B_i)$ for the positively skewed signal process $\{\delta_S(\mathbf{s})\}$. B_1–B_4 are the four target blocks

$\sqrt{P}$ was largest for locations for which $\underline{\mathbf{x}}(\mathbf{s}_0)$ was close to the centroid. For the broken-stick regression, this resulted in two concentric rings (Fig. 4A). $\sqrt{Q}$, on the other hand, depends on $\mathbf{s}_0$ only through the covariance $\mathbf{c}(\mathbf{s}_{1\ldots n}, \mathbf{s}_0)$ (Appendix C). If the distance between $\mathbf{s}_0$ and the nearest sample point increases then the universal kriging predictions converges to the fitted trend and, therefore, $\sqrt{Q} \to 0$. If $\mathbf{s}_0$ approaches a sample point, then $\sqrt{Q}$ grows, and depending on the magnitude of the noise variance, σ_ε^2, converges to a value $\leq \sqrt{P}$ (equality is reached for $\sigma_\varepsilon^2 = 0$). This behaviour explains the spatial distribution of $\sqrt{Q}$ shown in Fig. 4B.

Compared to UK, CK lost precision where the difference $\sqrt{P} - \sqrt{Q}$ was large. As $\sqrt{P}$ varied little, this happened in zones where $\sqrt{Q}$ was small, i.e., where sampling was sparse. Of course, also K (Fig. 4C) was related to the difference between the mean square errors of CK and UK, see (13); large values of K coincided with zones where CK lost precision. As long as K was close to one, CK performed nearly equally well as UK. Table 4 shows that the above relations also hold if we predict

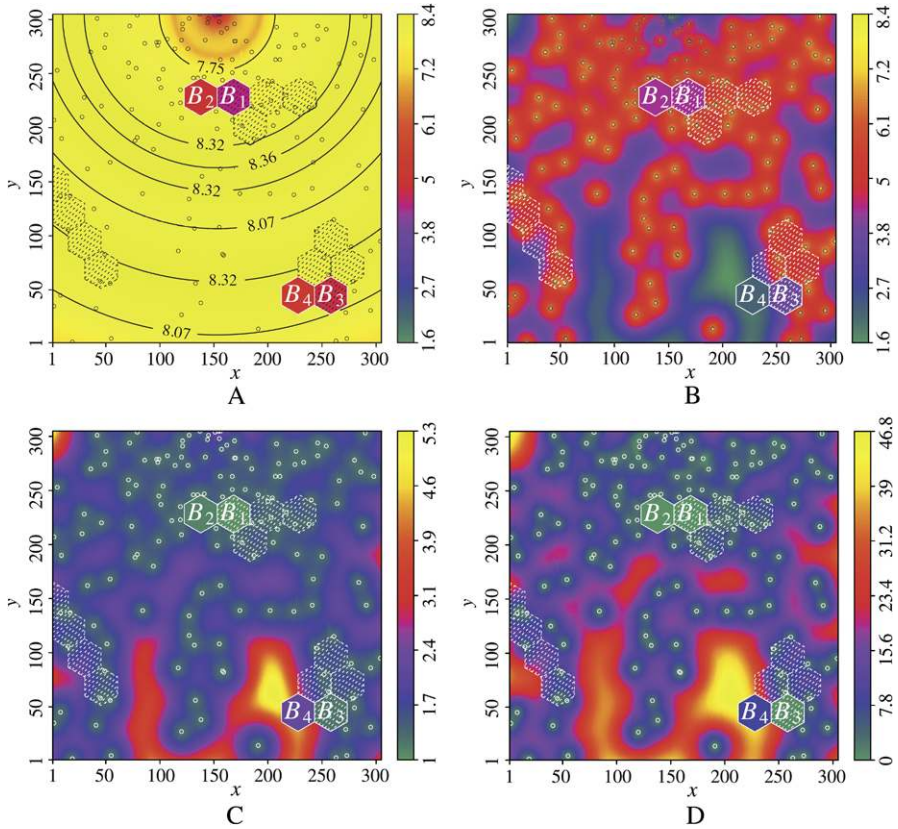

Fig. 4 Spatial distribution of terms relevant for point predictions by CK (Gaussian signal process $\{\delta_G(\mathbf{s})\}$). (**A**) $\sqrt{P}$, (**B**) $\sqrt{Q}$, (**C**) K and (**D**) MSPE[$\widehat{Y}_{CK}(\mathbf{s})$]–MSPE[$\widehat{Y}_{UK}(\mathbf{s})$]. Note that at the target blocks B_1–B_4 the block values of the corresponding terms are shown, which are also listed in Table 4

Table 4 Relevant terms for computing the CK predictions of the block means for the Gaussian signal process $\{\delta_G(\mathbf{s})\}$. The spatial point distributions of those terms within the simulation domain D are shown in Fig. 4

	B_1	B_2	B_3	B_4
$\sqrt{P}$	4.444	4.846	4.732	5.094
$\sqrt{Q}$	4.302	4.341	3.868	2.207
K	1.033	1.116	1.223	2.308
MSPE[$\widehat{Y}_{CK}$]–MSPE[$\widehat{Y}_{UK}$]	0.0202	0.255	0.745	8.336

block means. The sample points were densest around B_1 and B_2 and sparsest in the neighbourhood of B_4. Consequently, unlike for B_1 and B_2, $\sqrt{Q}$ was small for B_4 and the difference in the precision of CK and UK was largest for this block as the smoothing effect of UK increases in sparsely sampled areas (Goovaerts 1997, p. 370).

We can thus conclude that the use of CK (and CMCK) does not result in a substantial loss of precision as long as the sampling is not too sparse. Densely sampled data convey enough information to tie the UK predictions locally to the data. The UK predictions differ quite strongly from the fitted trend, and their variance is not much

smaller than the variance of the target data. In such circumstances, CK (CMCK) need not amplify the GLS residuals strongly to match the (co-)variances of the data, therefore $K \approx 1$ ($\mathbf{K} \approx \mathbf{I}$), and, consequently, the mean square errors of CK (CMCK) and UK do not differ much.

4.2 Threshold Exceedance

4.2.1 Marginal Bias

The bias score (BS) of UK reflected the well-known smoothing bias of that method for nonlinear predictions. No matter what block was considered and irrespective of the signal, we found BS > 1 if $\tilde{l} < c$ and BS < 1 if $\tilde{l} > c$ with $c \approx 0.5$ for $\{\delta_{\mathrm{G}}(\mathbf{s})\}$ (Figs. 5A–5C) and $c \approx 0.65$–0.8 for $\{\delta_{\mathrm{S}}(\mathbf{s})\}$ (Figs. 5D–5F). Exceedance of small and large thresholds was thus overestimated and underestimated by UK, respectively. Much the same results were obtained from the conditional simulations when threshold exceedance was predicted by their conditional medians. The CS results are therefore not shown in Fig. 5. The magnitude of the bias of UK differed between the blocks and the signal processes. For $\{\delta_{\mathrm{G}}(\mathbf{s})\}$, the bias was generally smaller and for block B_1 the predictor $I(\widehat{Y}_{\mathrm{UK}}(B_1) > T_{[l]})$ was nearly unbiased for $\tilde{l} < 0.9$ because $\mathrm{Var}[\widehat{Y}_{\mathrm{UK}}(B_1)] \approx \mathrm{Var}[Y(B_1)]$ (Fig. 6A). However, for block B_4, the bias of UK was substantial, in particular for $\{\delta_{\mathrm{S}}(\mathbf{s})\}$, because $\mathrm{Var}[\widehat{Y}_{\mathrm{UK}}(B_4)]$ was distinctly smaller than $\mathrm{Var}[Y(B_4)]$ (Figs. 6C and 6F), and this resulted in strongly biased predictions for the small and large thresholds.

Unlike UK, the bias scores of CMCK and CK were close to one for $\{\delta_{\mathrm{G}}(\mathbf{s})\}$ as long as $\tilde{l} < 0.9$ (Figs. 5A–5C). There were some deviations beyond that point, but one must bear in mind that BS becomes unstable for $\tilde{l} \to 1$ as the counts in the numerator and denominator of (21) decrease. For the positively skewed signal, also CMCK and CK exhibited some bias (Figs. 5D–5F): BS was negative for the small ($\tilde{l} < 0.3$–0.4) and large thresholds ($\tilde{l} > 0.9$) and positive in between. This pattern resulted from matching the means and the (co-)variances of the predictions and the block means. As the distributions of the latter were skewed for $\{\delta_{\mathrm{S}}(\mathbf{s})\}$, CMCK and CK overestimated the extent of the lower tails in these instances (Figs. 6D–6F). UK modelled the lower tails more accurately, but it strongly underestimated the extent of the upper tails of the distributions of the block means, in particular for block B_4. On the whole, CMCK and CK reproduced the skewed distributions of the block means more successfully than UK, and this is the main reason for their lesser bias. There were no distinct differences between CK and CMCK. Based on the BS criterion neither method can be preferred.

4.2.2 Conditional Bias

The BS criterion signals if a method systematically over or underpredicts threshold exceedances, but it does not provide information on how well a method detects (non)exceedance of a threshold given that there is one (none). Only conditional statistics such as the hit (H) and false alarm rate (F) or their difference, the PSS criterion, convey this information. Figure 7 shows the hit and false alarm rates of UK, CK, and

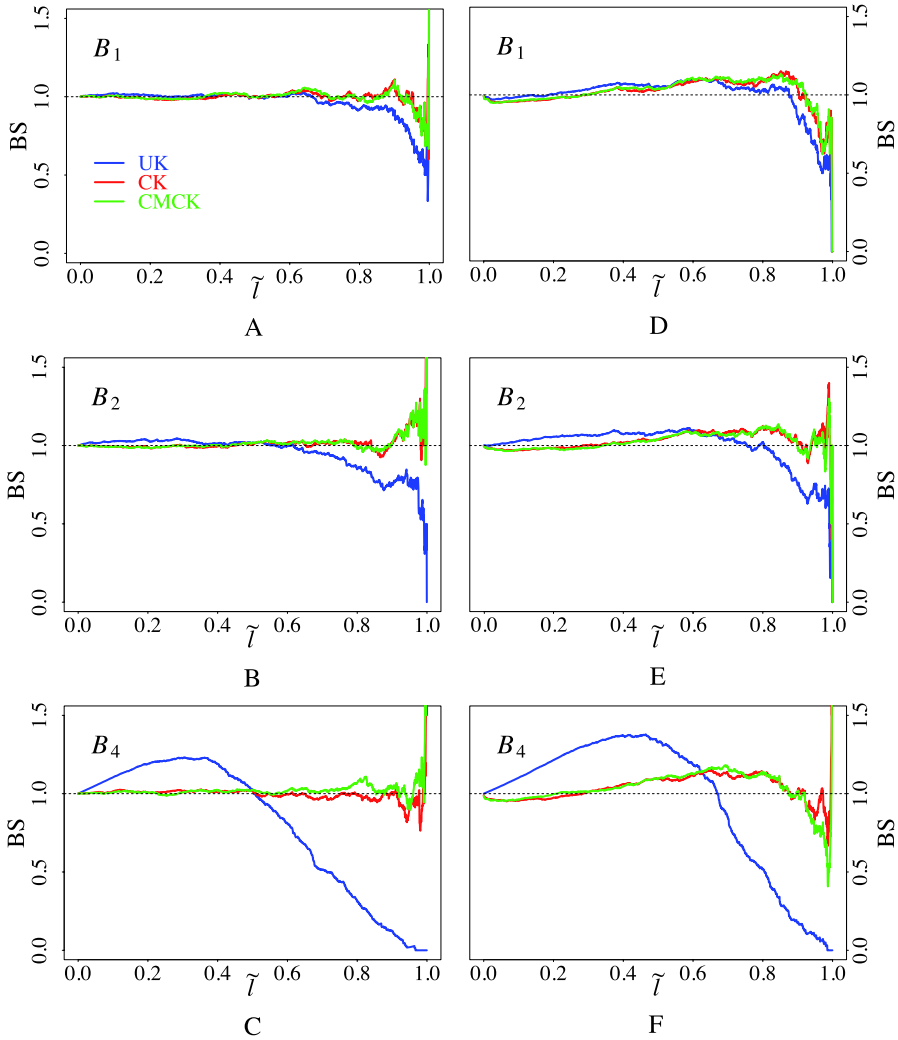

Fig. 5 Bias scores (BS) for predicting threshold exceedance vs. the relative ranks $\tilde{l}$ of the thresholds $T_{[l]}$: (**A–C**) $\{\delta_G(\mathbf{s})\}$; (**D–F**) $\{\delta_S(\mathbf{s})\}$

CMCK for block B_4 and $\{\delta_G(\mathbf{s})\}$, and Fig. 8 displays PSS of these methods for blocks B_1, B_2, and B_4 (both signal processes). Predicting threshold exceedance by the conditional medians of CS gave again very similar results to UK, and they are therefore not shown in Figs. 7 and 8. Although UK had a larger hit rate than CMCK and CK for the small thresholds (Fig. 7A), it did not outperform these methods with respect to PSS (Fig. 8C) because the advantage in H was more than compensated by a poor F (Fig. 7B). For large thresholds, the advantage in F was similarly overcompensated by a bad hit rate so that CMCK and CK again outperformed UK with respect to PSS. UK matched the performance of CMCK and CK only for the intermediate thresholds ($\tilde{l} \approx 0.4$–0.6, Fig. 8C), below and above this range, the PSS of UK deteriorated. In

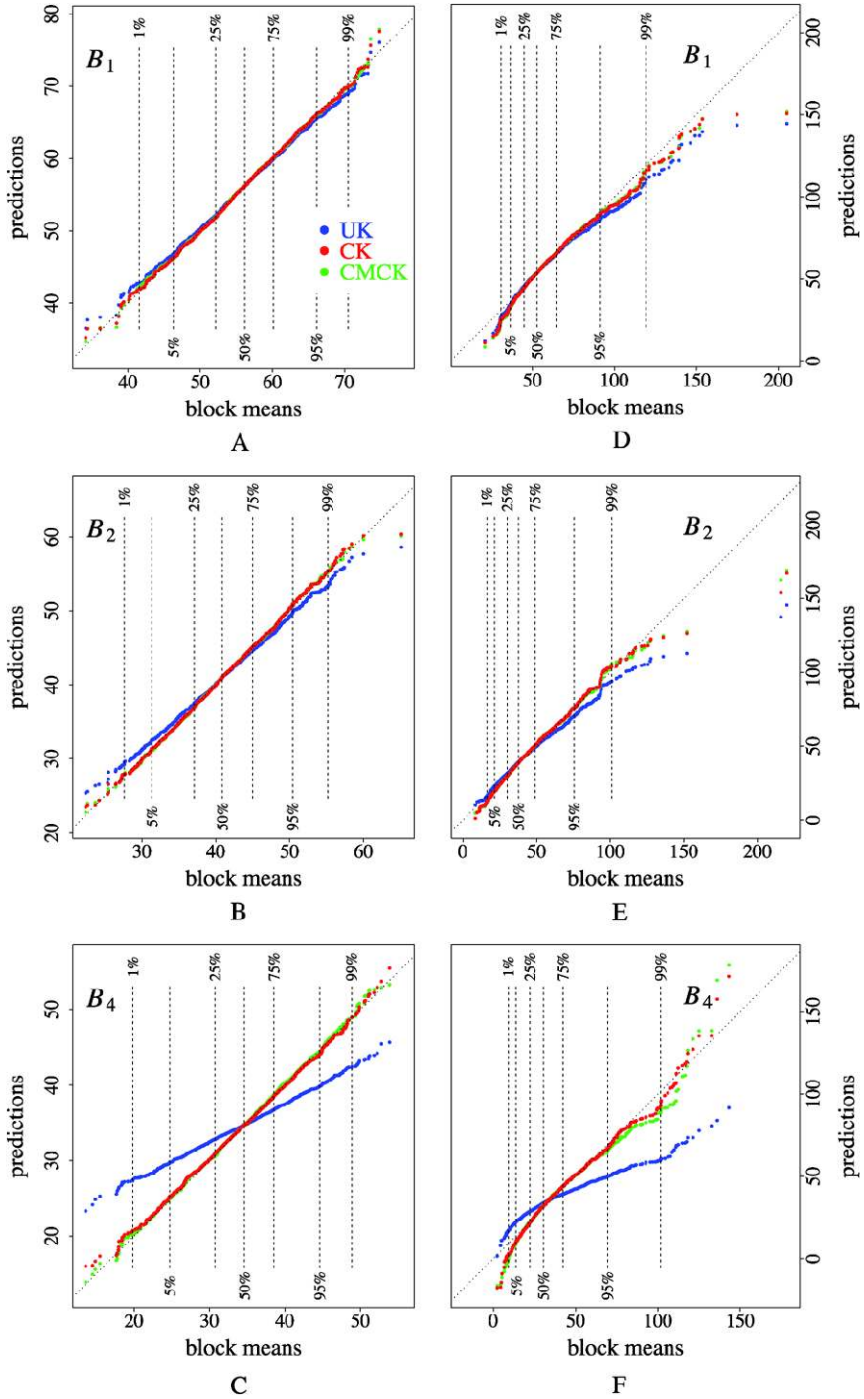

Fig. 6 Quantile–quantile plots of the predictions, $\widehat{Y}(B_i)$ and the simulated true block means $Y(B_i)$ for $\{\delta_G(\mathbf{s})\}$ (**A**–**C**) and $\{\delta_S(\mathbf{s})\}$ (**D**–**F**)

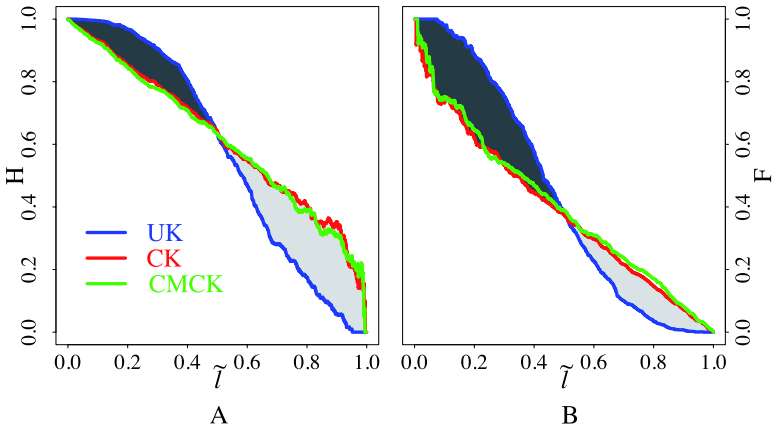

Fig. 7 (**A**) Hit rate H and (**B**) false alarm rate F for predicting threshold exceedance vs. the relative ranks $\tilde{l}$ of the thresholds $T_{[l]}$ (block B_4, $\{\delta_G(\mathbf{s})\}$)

contrast, the PSS of CMCK and CK were approximately constant for $\tilde{l}$ ranging from 0.2 to 0.9.

The advantage of CMCK and CK over UK and CS was observed for all the blocks and both signal processes, but it was small for B_1 and moderate for B_2 and B_3 (not shown). For $\{\delta_S(\mathbf{s})\}$, PSS of UK was no longer symmetrical to $\tilde{l} = 0.5$. The advantage of CMCK/CK over UK/CS was more pronounced for the small than the large thresholds. Finally, excepting block B_2 and $\{\delta_G(\mathbf{s})\}$ (Fig. 8B), PSS showed a slight but consistent advantage of CK over CMCK.

In view of the advantage of CMCK and CK over UK and, seemingly, also CS, one might rightly object that one could compute more informed predictions of threshold exceedance with the CS results. Instead of using the conditional medians, one could predict exceedance by the conditional quantiles, thereby specifically controlling the rate of false negatives. The other methods do not offer this possibility as they provide only the predictions and estimates of the MSPEs. It is therefore of interest to compare CMCK/CK and CS, taking into account the full information provided by CS.

Mason (1979) showed that given a probabilistic forecast of a binary event with outcomes yes and no, PSS is maximised by issuing a yes forecast if the modelled probability of the yes event exceeds the marginal probability of the yes events in the population. PSS of CS should therefore be maximised if instead of the conditional medians; the conditional $\tilde{l}$-quantiles were used to predict exceedance of threshold $T_{[l]}$. For the Gaussian signal process, this was indeed the case for all the blocks; predicting exceedance by the conditional quantiles clearly outperformed CK for $\tilde{l} < 0.3$ and $\tilde{l} > 0.7$ and was as good in between for B_4 (Fig. 9A). The hit and the false alarm rates were now approximately constant for the intermediate thresholds and increased (H) or decreased (F) for $\tilde{l} \to 0$ and $\tilde{l} \to 1$, respectively.

However, for $\{\delta_S(\mathbf{s})\}$, the tuned version of CS was not equally successful for block B_4. It was better than CK only for the largest thresholds, and for $\tilde{l} \in (0.5\text{–}0.7)$ it was even worse than UK (Fig. 9D). Compared with $\{\delta_G(\mathbf{s})\}$, H and F showed now a completely different pattern, which indicates that CS largely failed to model the

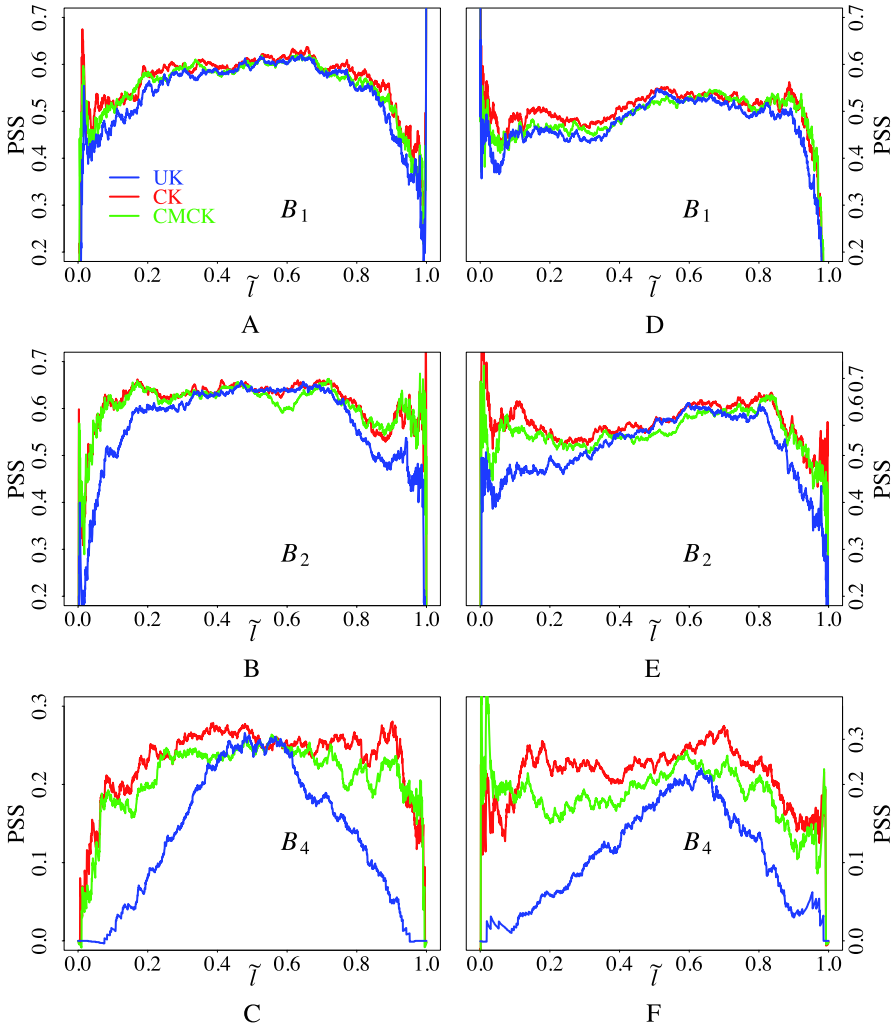

Fig. 8 Peirce skill score (PSS) for predicting threshold exceedance vs. the relative ranks $\tilde{l}$ of the thresholds $T_{[l]}$: (**A–C**) $\{\delta_G(\mathbf{s})\}$; (**D–F**) $\{\delta_S(\mathbf{s})\}$

conditional distributions of the block means for $\{\delta_S(\mathbf{s})\}$. This was confirmed by the coverage of one-sided prediction intervals computed from the conditional simulations for block B_4. The observed coverage differed considerably from the expected coverage of the intervals (Fig. 9E). For the Gaussian signal, this was not the case (Fig. 9B). CK outperformed the tuned version of CS also when predicting threshold exceedance for B_3 (not shown), albeit the differences between the methods were less pronounced. For blocks B_1 and B_2 the tuned version of CS showed a lesser conditional bias than CK for large $\tilde{l}$, but the bias score signalled substantial marginal bias (not shown). Irrespective of the signal process and for all the blocks, predicting threshold exceedance by upper tail conditional quantiles resulted in many false positives (Figs. 9C and 9F

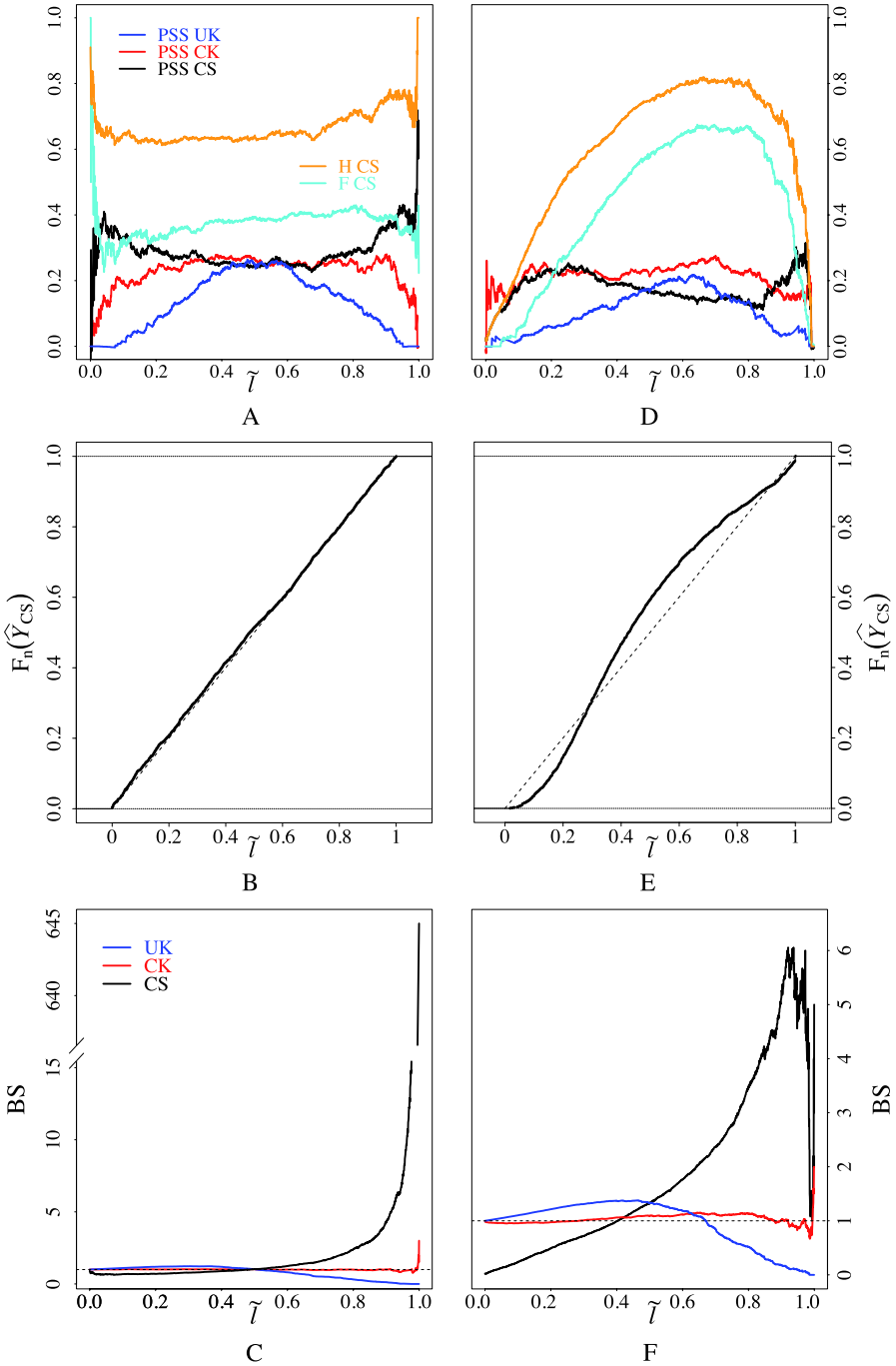

Fig. 9 Peirce skill score (PSS), (**A**, **D**), coverage of one sided prediction intervals (**B**, **E**) and bias score (BS), (**C**, **F**) for predicting threshold exceedance vs. the relative ranks $\tilde{l}$ of the thresholds: block B_4; (**A–C**) $\{\delta_G(\mathbf{s})\}$; (**D–F**) $\{\delta_S(\mathbf{s})\}$

for B_4). Thus, the price for lesser conditional bias was a very substantial increase in the marginal bias.

To summarise, we can state that UK failed to predict threshold exceedance, except in the case where local information was so dense that the smoothing bias of the method did not matter. CK consistently outperformed UK (clearly) and CMCK (slightly). CK was outdone only by CS for the Gaussian signal when threshold exceedance was predicted by the conditional quantiles. CS performed well in this instance because it captured prediction uncertainty successfully. However, CS failed in this respect for the positively skewed data and this impaired the predictions. CK (and CMCK) were quite resistant to distributional deviations and captured the upper tails of the skewed marginal distributions of the block means remarkably well.

5 Summary and Conclusions

Motivated by the task to survey soil contamination in the vicinity of stationary emitters of pollutants, our goal in this study was to find a simple, robust, and precise method for predicting linear and nonlinear functionals of the block means from data observed at points that show a spatial trend. So far, CS was the preferred tool for such prediction problems with local change of support. However, CS is highly parametric, is susceptible to model misspecification and, therefore, lacks the attractive, nonparametric flavor and the robustness of UK. Our simulations showed that CK offers a good compromise between robustness, acceptable precision for predicting the block means, and small bias when predicting threshold exceedance. Theoretically, the CK predictor might not exist, but we think that this is unlikely to happen in practise. In our simulations experiments, the CK predictor failed to exist only if the nugget-to-total-sill ratios were very large (e.g., >0.9 for block B_1) or if the blocks were far outside of the sampled domain and the spatial trend had to be extrapolated. Compared to UK, CK lost precision in predicting the block means only if the sampling was locally sparse. The variances of the UK predictions were then distinctly smaller than the variances of the block means. This discrepancy in the variances was the main cause for the bias of UK when predicting threshold exceedance. CK avoided this bias by inflating the variation of the GLS residuals, but this leads to a loss in precision when predicting the block means. By amplifying the moduli of the kriging weights, negative predictions are more likely to occur in CK. If this is a problem, CK predictions can be forced to nonnegative values by the method of Barnes and You (1992). For dense sampling, CK was hardly less precise than UK because no amplification of the GLS residuals was required as the variance of the UK predictions nearly matched the variances of the block means. Thus, irrespective of the sampling density and unlike UK, CK predicted threshold exceedance always well. CK was outperformed consistently in the Gaussian case when threshold exceedance was predicted by the conditional quantiles of CS. However, when sampling was sparse, CS was badly conditionally biased for the skewed data, and this demonstrates the vulnerability of the method to misspecification of the model. Furthermore, unlike CK, CS always showed a large marginal bias for large $\tilde{l}$, irrespective of the data distribution. CMCK did not show any advantage over CK in the simulations, neither for predicting linear nor nonlinear functionals of the block means. This contrasts Aldworth and

Cressie's findings and suggests that data with an apparent trend and/or local change of support prediction problems ask for a different optimal method.

In addition to the favourable statistical properties, CK is also simple to implement and does not requires more computing resources than UK. This contrasts with CS, which is still demanding with respect to computing time and storage capacity for medium to big real world problems. Based on these considerations, we recommend CK for spatial prediction of nonlinear quantities of block means from data observed at points.

Acknowledgements A part of this research was funded by a grant of the Swiss Federal Office for the Environment (FOEN). We gratefully acknowledge the support.

Appendix A: Taylor Series Approximation of a Nonlinear Predictor

Suppose $g(\cdot)$ is a smooth nonlinear scalar function of $\mathbf{Y}$ for which the first and second derivative exist and are continuous. Aldworth and Cressie (2003, p. 12) showed that $g(\mathbf{Y})$ can be approximated by a second-order Taylor series using the δ-method (Schott 1997, pp. 323–332) at $\boldsymbol{\mu} = \mathrm{E}[\mathbf{Y}]$

$$g(\mathbf{Y}) \approx g(\boldsymbol{\mu}) + \big(g'(\boldsymbol{\mu})\big)'(\mathbf{Y} - \boldsymbol{\mu}) + \frac{1}{2}(\mathbf{Y} - \boldsymbol{\mu})' g''(\boldsymbol{\mu})(\mathbf{Y} - \boldsymbol{\mu}), \tag{25}$$

where $g'(\boldsymbol{\mu}) = \frac{\partial g(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}$ is the $m \times 1$ vector with the first-order partial derivatives of $g(\boldsymbol{\mu})$; $g''(\boldsymbol{\mu}) = \frac{\partial^2 g(\boldsymbol{\mu})}{\partial \boldsymbol{\mu} \partial \boldsymbol{\mu}'}$ is the $m \times m$ Hessian matrix with the second-order partial derivatives of $g(\boldsymbol{\mu})$. Consequently,

$$E\big[g(\mathbf{Y})\big] \approx g(\boldsymbol{\mu}) + \frac{1}{2}\mathrm{tr}\big(g''(\boldsymbol{\mu})\mathrm{Cov}[\mathbf{Y}, \mathbf{Y}']\big) \tag{26}$$

and similarly

$$E\big[g(\widehat{\mathbf{Y}})\big] \approx g\big(\mathrm{E}[\widehat{\mathbf{Y}}]\big) + \frac{1}{2}\mathrm{tr}\big(g''\big(\mathrm{E}[\widehat{\mathbf{Y}}]\big)\mathrm{Cov}[\widehat{\mathbf{Y}}, \widehat{\mathbf{Y}}']\big), \tag{27}$$

where $\mathrm{tr}(\cdot)$ denotes the trace. The bias, $\mathrm{E}[g(\widehat{\mathbf{Y}}) - g(\mathbf{Y})]$, of $g(\widehat{\mathbf{Y}})$ vanishes approximately if the first and the second moments of $\widehat{\mathbf{Y}}$ and $\mathbf{Y}$ match. If in addition $\mathbf{Y}$ is Gaussian, then $\mathrm{E}[g(\widehat{\mathbf{Y}}) - g(\mathbf{Y})] = 0$. Hence, unlike $g(\widehat{\mathbf{Y}}_{\mathrm{CMCK}})$, $g(\widehat{\mathbf{Y}}_{\mathrm{UK}})$ is a biased predictor of $g(\mathbf{Y})$ because $\mathrm{Cov}[\widehat{\mathbf{Y}}_{\mathrm{UK}}, \widehat{\mathbf{Y}}'_{\mathrm{UK}}] \neq \mathrm{Cov}[\mathbf{Y}, \mathbf{Y}']$.

Appendix B: Expressing the Variance of Fitted Trend as a Mahalanobis Distance

We assume that the trend model includes an intercept. Hence, we can write $\mathbf{x}(\mathbf{s})' = (1, \underline{\mathbf{x}}(\mathbf{s})')$ and $\mathbf{X} = (\mathbf{1}, \underline{\mathbf{X}})$. Next, we compute the Cholesky decomposition of the covariance matrix of the data $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}'$ and orthogonalise the design matrix $\widetilde{\mathbf{X}} = \mathbf{L}^{-1}\mathbf{X}$ and its components $\widetilde{\mathbf{1}} = \mathbf{L}^{-1}\mathbf{1}$, $\underline{\widetilde{\mathbf{X}}} = \mathbf{L}^{-1}\underline{\mathbf{X}}$.

The covariance matrix of $\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}}$ can be expressed by

$$\begin{aligned}\mathrm{Var}[\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}}, \widehat{\boldsymbol{\beta}}'_{\mathrm{GLS}}] &= \left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1} = \begin{pmatrix} \mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1} & \mathbf{1}'\boldsymbol{\Sigma}^{-1}\underline{\mathbf{X}} \\ \underline{\mathbf{X}}'\boldsymbol{\Sigma}^{-1}\mathbf{1} & \underline{\mathbf{X}}'\boldsymbol{\Sigma}^{-1}\underline{\mathbf{X}} \end{pmatrix}^{-1} \\ &= (\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1} = \begin{pmatrix} \widetilde{\mathbf{1}}'\widetilde{\mathbf{1}} & \widetilde{\mathbf{1}}'\widetilde{\underline{\mathbf{X}}} \\ \widetilde{\underline{\mathbf{X}}}'\widetilde{\mathbf{1}} & \widetilde{\underline{\mathbf{X}}}'\widetilde{\underline{\mathbf{X}}} \end{pmatrix}^{-1}.\end{aligned}$$

Standard results from regression analysis provide

$$\mathrm{Var}[\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}}, \widehat{\boldsymbol{\beta}}'_{\mathrm{GLS}}] = \begin{pmatrix} 1/\widetilde{n} + \bar{\underline{\mathbf{x}}}'\mathbf{V}^{-1}\bar{\underline{\mathbf{x}}} & -\bar{\underline{\mathbf{x}}}'\mathbf{V}^{-1} \\ -\mathbf{V}^{-1}\bar{\underline{\mathbf{x}}} & \mathbf{V}^{-1} \end{pmatrix},$$

where

$$\bar{\underline{\mathbf{x}}} = \frac{1}{\widetilde{n}}\,\widetilde{\underline{\mathbf{X}}}'\widetilde{\mathbf{1}} = \frac{1}{(\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1})}\,\mathbf{X}'\,\boldsymbol{\Sigma}^{-1}\mathbf{1}$$

is the mean vector (centroid) of the columns of $\widetilde{\underline{\mathbf{X}}}$ and

$$\mathbf{V} = (\widetilde{\underline{\mathbf{X}}} - \widetilde{\mathbf{1}}\bar{\underline{\mathbf{x}}}')'(\widetilde{\underline{\mathbf{X}}} - \widetilde{\mathbf{1}}\bar{\underline{\mathbf{x}}}')$$

is the cross-product matrix of the centred, orthogonalised design matrix component $\underline{\mathbf{X}}$.

The variance of the fitted trend at location $\mathbf{s}$ can now be expressed

$$\begin{aligned}\mathrm{Var}\left[\mathbf{x}(\mathbf{s})'\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}}\right] &= \left(1, \underline{\mathbf{x}}(\mathbf{s})'\right)\mathrm{Var}[\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}}, \widehat{\boldsymbol{\beta}}'_{\mathrm{GLS}}]\begin{pmatrix} 1 \\ \underline{\mathbf{x}}(\mathbf{s}) \end{pmatrix} \\ &= \left(1, \bar{\underline{\mathbf{x}}}' + \left[\underline{\mathbf{x}}(\mathbf{s})' - \bar{\underline{\mathbf{x}}}'\right]\right)\mathrm{Var}[\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}}, \widehat{\boldsymbol{\beta}}'_{\mathrm{GLS}}]\begin{pmatrix} 1 \\ \bar{\underline{\mathbf{x}}} + [\underline{\mathbf{x}}(\mathbf{s}) - \bar{\underline{\mathbf{x}}}] \end{pmatrix} \\ &= 1/\widetilde{n} + \left(\underline{\mathbf{x}}(\mathbf{s}) - \bar{\underline{\mathbf{x}}}\right)'\mathbf{V}^{-1}\left(\underline{\mathbf{x}}(\mathbf{s}) - \bar{\underline{\mathbf{x}}}\right), \qquad (28)\end{aligned}$$

where again $1/\widetilde{n} = (\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1})^{-1}$.

Appendix C: An Alternate Expression for Q

The variance of the universal kriging predictor for location $\mathbf{s}_0$ can be expressed by Roth (1998, p. 1,001)

$$\begin{aligned}\mathrm{Var}\left[\widehat{Y}_{\mathrm{UK}}(\mathbf{s}_0)\right] &= \mathrm{Var}\left[\widehat{Y}_{\mathrm{SK}}(\mathbf{s}_0)\right] + \mathrm{Var}\left[\mathbf{x}(\mathbf{s}_0)'\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}}\right] \\ &\quad - \mathbf{c}(\mathbf{s}_{1\ldots n}, \mathbf{s}_0)'\boldsymbol{\Sigma}^{-1}\mathrm{Var}[\mathbf{X}\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}}]\boldsymbol{\Sigma}^{-1}\mathbf{c}(\mathbf{s}_{1\ldots n}, \mathbf{s}_0),\end{aligned}$$

where $\widehat{Y}_{\mathrm{SK}}(\mathbf{s}_0)$ is the simple kriging predictor of $Y(\mathbf{s}_0)$. $Q(\mathbf{s}_0)$ is therefore equal to

$$\begin{aligned}Q(\mathbf{s}_0) &= \mathbf{c}(\mathbf{s}_{1\ldots n}, \mathbf{s}_0)'\boldsymbol{\Sigma}^{-1}\mathbf{c}(\mathbf{s}_{1\ldots n}, \mathbf{s}_0) \\ &\quad - \mathbf{c}(\mathbf{s}_{1\ldots n}, \mathbf{s}_0)'\boldsymbol{\Sigma}^{-1}\mathbf{X}\left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{c}(\mathbf{s}_{1\ldots n}, \mathbf{s}_0)\end{aligned}$$

and depends on $\mathbf{s}_0$ only by $\mathbf{c}(\mathbf{s}_{1\ldots n}, \mathbf{s}_0)$ but not by $\mathbf{x}(\mathbf{s}_0)$.

References

Aldworth J, Cressie N (2003) Prediction of nonlinear spatial functionals. J Stat Plan Inference 112:3–41

Altfelder S, Beyer C, Duijnisveld WHM, Schneider J, Streck T (2002) Distribution of Cd in the vicinity of a metal smelter: interpolation of soil Cd concentrations with regard to regulative limits. Z Pflanzenernähr Bodenkd 165(6):697–705

Barnes RJ, You K (1992) Adding bounds to kriging. Math Geol 24(2):171–175

Van Brummelen T, Verweij R, Wedzinga S, van Gestel C (1996) Enrichment of polycyclic aromatic hydrocarbons in forest soils near a blast furnace plant. Chemosphere 32(2):293–314

Buxton BE, Wells DE, DiLuise G (1997) Comparison of three kriging methods for making soil remediation decisions. In: Baafi EY, Schofield NA (eds) Geostatistics Wollongong '96, vol 2. Kluwer Academic, Dordrecht, pp 984–995

Chan G, Wood ATA (1997) An algorithm for simulating stationary Gaussian random fields. Appl Stat-J R Stat C 46:171–181

Chilès JP, Delfiner P (1999) Geostatistics: modeling spatial uncertainty. Wiley, New York

Clifford D (2009) spatialCovariance: computation of spatial covariance matrices for data on rectangles. R package version 0.6-4

Cressie N (1993) Aggregation in geostatistical problems. In: Soares A (ed) Geostatistics Tróia '92, vol 1. Kluwer Academic, Dordrecht, pp 25–36

Cressie N, Johannesson G (2001) Kriging for cut-offs and other difficult problems. In: Monestiez P, Allard D, Froidevaux R (eds) geoENV III—Geostatistics for Environmental Applications. Kluwer Academic, Dordrecht, pp 299–310

Cressie N, Frey J, Harch B, Smith M (2006) Spatial prediction on a river network. J Agric Biol Environ Stat 11(2):127–150

Diggle PJ, Ribeiro PJ (2007) Model-based geostatistics. Springer, Berlin

Faraway J (2004) Linear models with R. Texts in statistical science. Chapman & Hall/CRC, Boca Raton

Frangi J, Richard D (1997) Heavy metal soil pollution cartography in northern France. Sci Total Environ 205(1):71–79

Goovaerts P (1997) Geostatistics for natural resources evaluation. Oxford University Press, New York

Goovaerts P, Trinh HT, Demond A, Franzblau A, Garabrant D, Gillespie BW, Lepkowski J, Adriaens P (2008) Geostatistical modeling of the spatial distribution of soil dioxin in the vicinity of an incinerator. 2. Verification and calibration study. Environ Sci Technol 42(10):3655–3661

Gotway CA, Young L (2002) Combining incompatible spatial data. J Am Stat Assoc 97(458):632–648

Harville DA (1977) Maximum likelihood approaches to variance component estimation and to related problems. J Am Stat Assoc 72:320–340

Harville DA (1997) Matrix algebra from a statistician's perspective. Springer, Berlin

Mason I (1979) On the reducing probability forecasts to yes/no forecasts. Mon Weather Rev 107(2):207–211

McGrath D, Zhang C, Owen TC (2004) Geostatistical analyses and hazard assessment on soil lead in silvermines area, Ireland. Environ Pollut 127:239–248

Van Meirvenne M, Vercauteren F, Hofman G, Ide G (1993) Geostatistical analysis of the cadmium pollution in North-West Limburg, Belgium. Biom Praxim 33(2):45–56

Papritz A, Herzig C, Borer F, Bono R (2005) Modelling the spatial distribution of copper in the soils around a metal smelter in northwestern Switzerland. In: Renard P, Demougeot-Renard H, Froidevaux R (eds) Geostatistics for environmental applications. Springer, Berlin, pp 343–354

Peirce CS (1884) The numerical measure of the success of predictions. Science 4(93):453–454

R Development Core Team (2009) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.r-project.org

Rawlins BG, Lark RM, Webster R, O'Donnell KE (2006) The use of soil survey data to determine the magnitude and extent of historic metal deposition related to atmospheric smelter emissions across Humberside, UK. Environ Pollut 143(3):416–426

Roth C (1998) Is lognormal kriging suitable for local estimation? Math Geol 30(8):999–1009

Saito H, Goovaerts P (2001) Accounting for source location and transport direction into geostatistical prediction of contaminants. Environ Sci Technol 35:4823–4829

Schlather M (2001) Simulation of stationary and isotropic random fields. R-News 1(2):18–20

Schott JR (1997) Matrix analysis for statistics. Wiley, New York

Starks TH, Sparks AR, Brown KW (1987) Geostatistical analysis of Palmerton soil survey data. Environ Monit Assess 9(3):239–261

Tercan AE (2004) Global recoverable reserve estimation by covariance-matching constrained kriging. Energy Sources 26:1177–1185

Wilks DS (2006) Statistical methods in the atmospheric sciences. International geophysics series, vol 91. Elsevier, Amsterdam

Xing X, Lu Y, Dawson R, Shi Y, Zhang H, Wang T, Liu W, Ren H (2005) A spatial temporal assessment of pollution from PCBs in China. Chemosphere 60(6):731–739